# MASTER'S (MS) IN INTERDISCIPLINARY DATA SCIENCE (MIDS)
## Executive Summary

This MS Degree in Interdisciplinary Data Science (MIDS) emerges from a particular vision of data science as an inherently interdisciplinary and team-based endeavor. It borrows technical tools from a variety of quantitative disciplines and connects them to data-driven inquiry in virtually all disciplinary domains. A data scientist of the kind we envision is someone whose technical depth in *some* quantitative area is complemented by broad awareness of existing and emerging analytical tools, and by an ability to leverage general technical and analytical proficiency to solve hard problems as part of a team. She is furthermore someone who can communicate effectively with those whose expertise lies in a domain area of the problem to be solved rather than in quantitative data techniques, and someone whose critical thinking skills can cross-disciplinary domains. The very nature of this brand of data science therefore requires interdisciplinary training within the quantitative sciences, exposure to data-driven problems in a variety of disciplines, and direct experience in interdisciplinary team-based science.

This intellectual vision of *interdisciplinary* and *team-based* data science leads us to propose not only a degree curriculum but also a complementary new model for advancing university-wide engagement in graduate education across departments and schools. Hosted jointly by the Social Science Research Institute (SSRI) and the Initiative in Information at Duke (iiD), it invites units across the Duke campus to contribute elective courses (which constitute half of the courses taken by students). A significant fraction of tuition revenue is then passed to contributing units in direct proportion to the extent to which each unit's courses are taken by MS students in the program. The core courses (which constitute the other half of courses taken by students) are also offered in collaboration with academic units – departments in A&S as well as a number of schools that have expressed interest and support, with substantial funds set aside to compensate collaborating partners as well as to invest in seeding faculty-led research projects that can form team-based capstone experiences for MS students, and often PhD students, postdocs, and undergraduates. The intent is to create a Data Science MS degree that is unique in its broad interdisciplinary character, combines technical training and project leadership with domain applications that extend into a variety of disciplines (including but not limited to the social sciences and humanities), provides strong incentives for many campus units to participate, and fosters new intellectual horizontal and vertical synergies across the university. The degree furthermore complements a simultaneously proposed Masters degree in Quantitative Management (Fuqua) with promising areas of course overlap as the programs develop, and its courses are likely to be of interest to Masters and PhD students in multiple units.

## Audience

The MIDS welcomes students from many disciplinary backgrounds who seek training in quantitative methods and exposure to how quantitative methods are applied to different domains and advanced through team-based projects. At the outset, we aim to take a relatively conservative approach in our recruitment, focusing on applicants with a baseline of quantitative and programming background. In the future, once appropriate support infrastructures and preparation pipelines are in place, we intend to recruit individuals of more diverse academic backgrounds. We expect this program will attract students with backgrounds in statistics and

computer science, as well as those from the social sciences, natural sciences, and humanities with some quantitative experience. We also expect that the program will have broad appeal for professionals working in non-profits, government, consulting, technology, media, business, and education. In addition to training in traditional and emerging quantitative methods, students will develop competencies in data security, data management, communication and team-based science. They will graduate with a highly practical skillset anchored in projects that emerge within Duke's highly interdisciplinary environment. Given the size of related MS degrees in peer institutions, we foresee no difficulty attracting an initial cohort of 20 students, with that number climbing to 30 students by year 3.

## Overview of the degree

*Courses*
·   30 credits over 3-4 semesters plus 4 credits of Graduate School training in Academic Integrity and Responsible Conduct of Research (AIRCR).
·   5 core courses:   *Data to Decision*, *Data Marshaling & Management*, *Data Modeling & Analysis*, *Data Visualization*, and a 2-semester (half credit) *Data Science Seminar*.
·   Capstone project/seminar
·   Data Science electives: Select graduate and advanced undergraduate courses, with the goal of balancing theory & methods with quantitative strategies in domain-specific applications.
·   Elective selections are subject to approval by the Director of Graduate Studies.

*Application*
·   Capstone project, typically integrated into faculty research programs, in the students' final semester(s): Hands-on leadership on a real-world data project (possibly but not necessarily connected to Bass Connections).

## Financial model

The Master's in Interdisciplinary Data Science (MIDS) offers an innovative model crafted to facilitate collaboration and inter-departmental (and inter-school) cooperation. Key features of the program will be centrally organized and housed in SSRI, including managing all administrative matters and absorbing the fixed costs of admitting students. Yet all departments and schools across Duke will be welcomed to propose and offer new courses (or modifications of existing courses) as the field of data science develops and changes, and the allocation of the program's resources will incentivize such innovation.  The program is structured to lower the barriers to entry for smaller units that would like to innovate but do not have the full suite of resources to offer an MS program that generates sufficient demand. As such, this program can play an important role in stemming a tide of many small and disconnected MA programs while providing the framework for units to collaborate to build a more integrated and connected program. Through the revenue-sharing model, a single course, or a set of courses, offered through the MIDS program (or in collaboration with it) can become a resource to the participating departments and simultaneously create intellectual connections across the university.
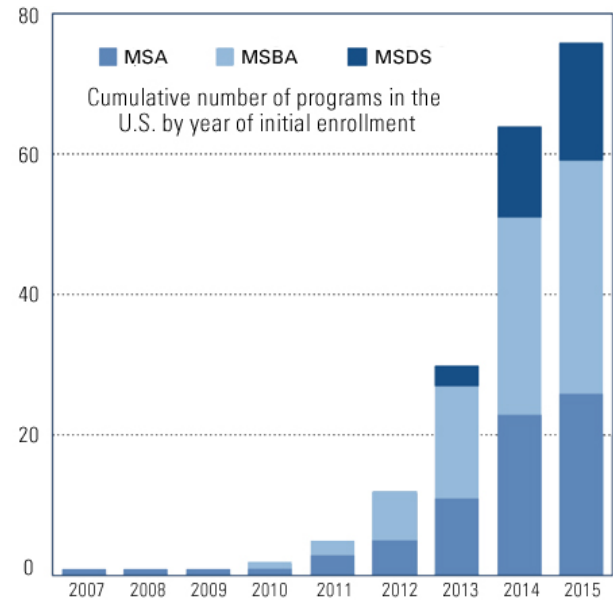
Put differently, the Interdisciplinary Data Science program seeks to make natural and immediate allies of departments, schools and institutes because of the potential for significant discretionary resources to emerge for all partners.  It is designed to diminish the perception that departments and institutes are competitors for scarce resources, particularly around faculty teaching. The alliances that we will create are furthermore fully in line with Duke's intellectual priorities: greater interdisciplinary collaboration on problem-based applications, with institutes and initiatives enabling coordination. As a result, while it may well be the case that new conversations might begin because of the potential for new revenues, success will depend on these conversations becoming intellectually deep – as it should – and in turn could have transformative (and positive) impacts on research and teaching outside the Interdisciplinary Data Science program (for postdocs, graduate students, and advanced undergraduates as well as initiatives like Bass Connections – and, indeed, for faculty).

In fact, the financial model envisions investing *directly* in intellectual cross-disciplinary collaborations that result in faculty-led research projects that in turn create team-based capstone experiences for MS students. Beyond the mere sharing of revenue to create a truly interdisciplinary data science degree, the model therefore aims directly to benefit research across all departments and schools. We envision that, in addition to enhancing the interdisciplinary research environment at Duke, this approach will furthermore make us significantly more competitive for NIH, NSF and DoD interdisciplinary training and research grants, and that our students will play important roles in the increasing number of cross-disciplinary teams (such as those that have emerged through the Bass program and Data+).

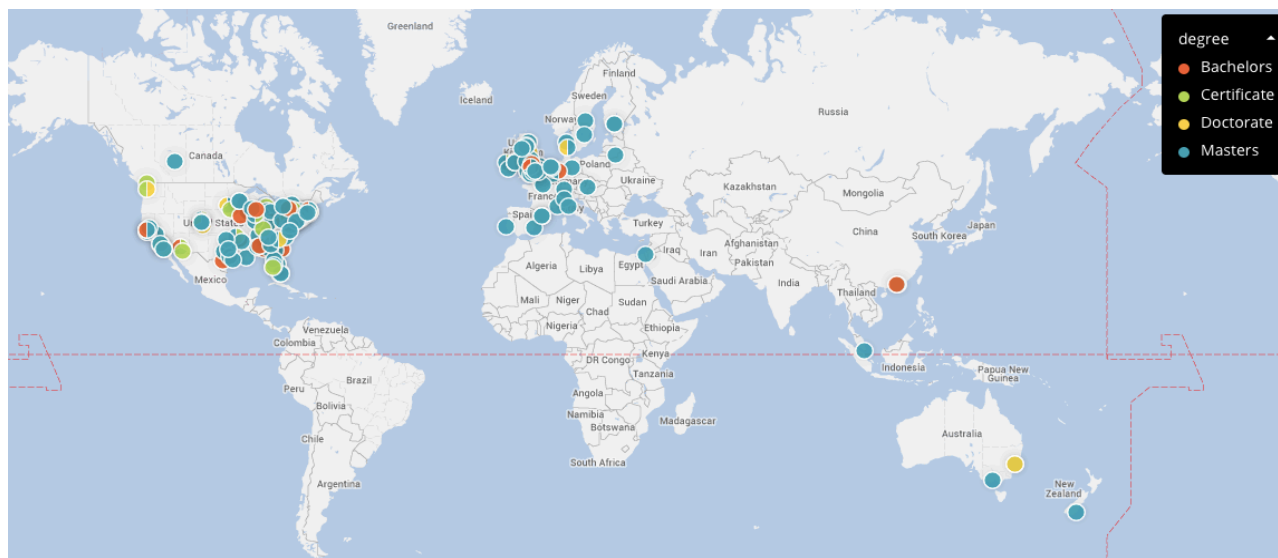# MASTER'S (MS) IN INTERDISCIPLINARY DATA SCIENCE: Proposal

## Section 1: Overview and Rationale

With the increasing availability of both big data and advanced computing technologies, the culture of big data is ascendant in nearly every area of inquiry. As a result, a surge in worldwide demand for data scientists outpaces the number of qualified applicants despite the rapid rise of new data science master's degree programs both within the U.S. (as illustrated in the accompanying graph to the right)[1] and Europe (as illustrated in the map at the bottom of the page).[2] While many of these new programs will train sophisticated data experts in particular technical subspecialties, few have the capacity to systematically and meaningfully connect a more holistic technical background with the variety of domains in need of data insights. Our contention is that the true power of data science in the culture of big data lies precisely in that connection, and that harnessing this requires both *interdisciplinary training* and genuine comfort with *team-based science*.

## 1.1 What is "data science"?

While the term "data science" can have different meanings in different contexts, the particular

---

[1] The graph depicts the growth in three types of programs: (1) Master of Science in Analytics (MSA), (2) Master of Science in Data Science (MSDS) and (3) Master of Science in Business Analytics (MSBA).
[2] The map is generated from data at http://data-science-university-programs.silk.co.

brand of data science we are interested in is fundamentally *interdisciplinary* and *team-based*. Its interdisciplinarity is found both *within the quantitative sciences* (e.g. mathematics, statistics, computer science) and in its engagement with applications *across multiple disciplinary domains outside the quantitative sciences* (e.g. the social sciences, the humanities, the medical sciences). It recognizes that this engagement is a two-way street: quantitative science does not just help other disciplines but is also inspired and motivated by unsolved problems in those disciplines. Further, given the complex and cross-disciplinary nature of data-driven problems in research, industry and government, data science as we envision it is almost by definition strongest in team settings where the data scientist is confident of her own analytic skills, able to leverage those with analytic depth in other team members, and ready to communicate with non-technical domain experts.

We have found this brand of data science resonating in our three-year experience with the Data+ summer program where teams of undergraduates, mentored by graduate students and post-docs, have intensively engaged with data around projects proposed by both academic and external "clients" across the natural sciences, social sciences and humanities. Within these projects we have identified key aspects of data science as practiced in teams, and these have in turn helped us define for ourselves "What is data science?" In the graphic below, each column represents one Data+ project from Summer 2016. A darkened square indicates that the tasks in the row were integral to that column's project. What emerges is an intellectual landscape that answers the question "What is data science?" for our purposes and suggests that a program aimed at facilitating this brand of data science should combine "diving deep" into some skills with a *broad awareness* of the other skills needed to solve data-related problems and the skills required to work in successful teams.



**What is data science?**

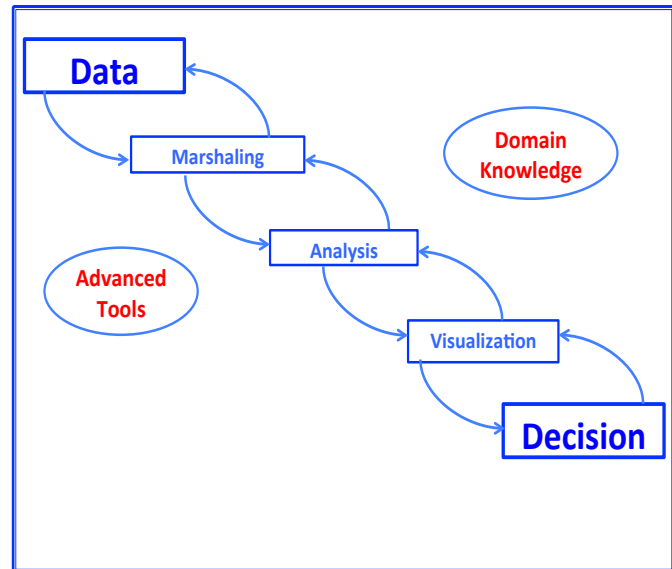Data+ projects reveal an intellectual landscape of interlocking pools

How to train Data Scientists? Deep dive into some pools, broad awareness of all

## 1.2 The "Data-to-Decision" Cycle

The intellectual landscape that emerges from our experience with the Data+ program illustrates that the specific skills identified in our team-based projects fall into three broad categories: *marshaling* data, *analyzing* data and *visualizing (and communicating) data*. We view these as the key steps in a "Data-to-Decision" cycle that allows entities to employ data to make and refine decisions, and it is this cycle that in turn forms the core of our proposed curriculum to train data scientists.

The "Data-to-Decision" cycle begins with the recognition that big data itself is not inherently meaningful for concrete decisions. Rather, the right data must be *marshaled* from big data, then *analyzed* by those with technical sophistication, and finally *visualized* in a way that speaks powerfully to the decisions at hand. Feedback loops back to the data then emerge as decisions are refined and new problems surface.



The tasks in each of the steps in the "Data-to-Decision" cycle stem from different quantitative disciplines (e.g. mathematics, statistics, computer science), implying that the type of data scientist we envision must be able to swim in multiple disciplinary pools. Yet domains outside these quantitative disciplines including industry, the non-profit world, and government, are eager to take advantage of the "Data-to-Decision" cycle to solve problems in their own subject matter. *Domain knowledge* in the area in which data is marshaled, analyzed and visualized will strengthen the data scientist's ability to do the "right" data science and to communicate it effectively, while mastering *advanced tools* in some of the mathematical disciplines will allow the emerging data scientist to define her own role while seeing how it connects to those of others with deeper expertise in other advanced tools.

## 1.3 Implications for an Interdisciplinary Data Science Curriculum

The holistic approach to interdisciplinary data science outlined above has several implications for how one would go about designing a curriculum aimed at training data scientists. These implications lead to discrete differences between the curriculum we propose and the type of curriculums often found in disciplinary programs that strive for technical depth in selective disciplines or that are sponsored by domain-specific schools (such as business schools, where many of the current data science programs reside).

1. *Broad awareness* of all the discrete parts of the data-to-decision cycle calls for the **core courses** (common to all students) of our proposed curriculum to be **centered on marshaling, analyzing** and *visualizing* **data**. This creates a common language shared by all students, and a common frame of reference they can bring to any data-driven project aimed at taking stakeholders from "data" to "decisions".

2. Since the specific skills within the data-to-decision cycle are rooted in multiple quantitative disciplines, these **core courses** must ***cross the boundaries of these disciplines***. Unlike the many technical data degrees that lie within a disciplinary home, the courses at the core of an interdisciplinary data science degree would therefore draw on expertise and faculty from different disciplines.

3. The need for students to *dive deep* into some technical aspects of the data-to-decision cycle and develop technical comparative advantage (that distinguishes them from others on teams) suggests the need for students to ***gain additional depth*** in some advanced technical topics not covered in sufficient depth in the core courses.

4. If the aim of the degree is to train data scientists who can address the diversity of topics we have seen emerge in our Data+ program, the curriculum must allow students flexibility in choosing electives from departments in different domain disciplines. Unlike advanced topics electives, these electives would ***deepen domain knowledge*** as data skills are employed within those domain areas.

5. Recognizing that interdisciplinary data science can function effectively only in teams where individual expertise is leveraged, the curriculum needs to find a central place for ***team-based data science*** – and a central place for students to practice the necessary skills that include being able to work in teams, lead sub-teams, engage in critical thinking across disciplines and communicate complex ideas to domain experts or external clients that lack technical sophistication.

6. Finally, the ambition inherent in these five implications requires a program that extends to at least three and typically four semesters – rather than the very compact data science degrees that are offered at a number of institutions.

## 1.4 Turning Broad Implications into an Interdisciplinary Data Science Curriculum

*Given these implications,* the Social Science Research Institute (SSRI) and the Information Initiative at Duke (iiD) therefore propose a new **Master's in Interdisciplinary Data Science (MIDS)***.* Rather than offering a more siloed MS degree focused on technical sophistication (as might be the case in a disciplinary department) or one particular domain application (as would be the case in business schools), we envision a degree program that extends across the university and invites broad collaboration to connect data science with real world problems as varied as using social media to track the spread of disease, visualizing thematic trends in literature across centuries, or harnessing the power of electronic medical records to improve population health. Duke's fertile cross-disciplinary culture makes Duke an ideal place for such a program, and makes it likely that Duke will quickly establish itself as a unique leader in *bringing big data to transformative real-world decisions.*

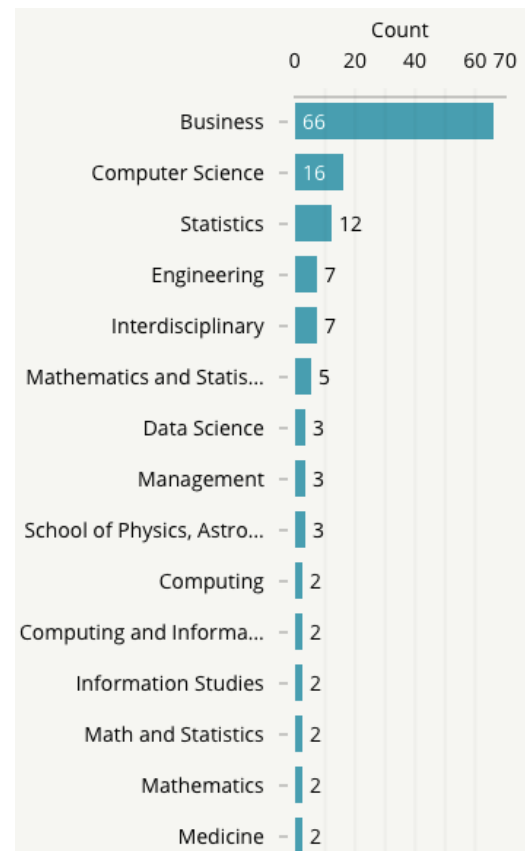The six implications outlined above map directly into the six key features of the program:

1. A **Data-to-Decisions** core course roots all students in the underlying approach to data science, and the three additional core courses ensure students are versed in the breadth of tasks involved in **Marshaling, Analyzing** and **Visualizing** data.

2. The core courses are developed by experts *across disciplines*, including some who hold interdisciplinary appointments in SSRI (with links to iiD).

3. Students are required to take 2-3 **electives to deepen technical expertise**.

4. Students are required to take 2-3 **electives to deepen domain knowledge** and its relevance to quantitative methods in at least one area of their choosing.

5. Throughout the program, students attend a **data science seminar** where they communicate about how data science impacts decisions and think critically across disciplines and industries.[3] The students' **capstone experience** involves playing a leadership role in a team-based data science project.

6. The program requires a minimum of three semesters, and likely four semesters for most students, with the potential for internships offering other experiences along the way.

## 1.5 Interdisciplinary Data Science within the Data Science Landscape

Our interdisciplinary and team-based approach to data science – and the curriculum this implies – distinguishes our program from most others that currently use the term "data science". By necessity, it requires engagement of expertise across the university, which in turn drives the financial model we propose – a model that relies heavily on partnerships to foster innovation.

As illustrated in the accompanying graph (to the right),[4] this engagement of expertise across a university is quite unique. Almost all current programs in data science reside within departments or professional schools, typically with a primarily technical focus or a dedication to business problems. Of the seven programs that bill themselves as "interdisciplinary", only 4 offer a Master's degree, and none of these envision the kinds of university-wide connections we propose to form.[5] Given the greater maturity of Duke's interdisciplinary infrastructure, it is not surprising that the enormous growth in data science programs elsewhere has been largely confined within the traditional university units that lack the collaborative reach we envision. Our opportunity to lead in the area of interdisciplinary data science emerges precisely because of Duke's long-



---

[3] The seminar is taken for credit in the first year of a student's program.

[4] The table is derived from data at http://data-science-university-programs.silk.co.

[5] The 4 programs are as follows: The University of Rochester's MS program offers optional concentrations in business or the biomedical or social sciences; NYU's MS program is focused on the development of new data science methods but offers electives in a number of domain specific areas; Georgia Tech's MS program primarily connects business, engineering and computer science; and Chapman University's MS program is heavily science focused.

standing commitment to innovate on top of these traditional units.

To better understand our place in the data science landscape, we have focused on four of our most prominent and proximate competitors, chosen because they have an interdisciplinary flavor and are not housed solely in business schools:

1) MS in Data Science, New York University,[6] a 2-year program for which 50% of the courses are required and 50% are elective.
2) MS in Data Science, Columbia University,[7] a flexible program (1 year when enrolled full-time) for which 70% of the courses are required and 30% are elective.
3) MS in Data Science, University of Virginia,[8] an 11-month program for which 80% of the courses are required and 20% are elective.
4) MS in Analytics, NC State University,[9] a 10-month program that "does not dwell on theory" for which 100% of the courses are required.

Of these, NYU offers the program most similar to ours, with broad sets of electives in diverse domain disciplines. Our approach, however, uses the data science seminar and capstone project to place significantly more emphasis on instruction in, and engagement with, practical, communication and team-based skills than NYU's programs. The other three programs are truly designed as 1-year programs, and therefore have curriculums that are significantly less broad and less deep than the one we propose. All of these competitors have enrollments that are significantly higher than those we envision for our program in the short to intermediate run.[10] While this is a strong indication that there is sufficient demand for a program like ours, our smaller size also allows us to pay much closer attention to features like capstone experiences that are most important.

## 1.6  Why SSRI and iiD?

Finally, we argue that SSRI and iiD are uniquely positioned to bring our proposed conception of interdisciplinary data science to the university's curriculum. SSRI is already deeply engaged in Duke's effort to become a global leader in the management of large-scale, interdisciplinary and data-intensive projects, particularly since its recent move to Gross Hall adjacent to iiD. It houses a growing repository of sensitive data in OIT supported environments that protect data privacy while simultaneously facilitating collaboration whenever possible. It furthermore takes a broad perspective of what constitutes "data", with new initiatives underway to explore, in collaboration with the interpretative social sciences and humanities, the very nature of how data is conceptualized in the broader social context. SSRI's faculty, fellows, and postdocs, combined with its technical capacity to securely store and analyze immense amounts of highly sensitive data, have therefore made the institute invaluable to the culture of computational research at Duke. The neighboring Information Initiative (iiD) is at the forefront of advancing data science in all its applications, with iiD researchers emerging from multiple departments and schools (Pratt, Math, Statistics, Computer Science) yet deeply engaged with researchers in the natural, medical and

---

[6] http://cds.nyu.edu/academics/ms-in-data-science/
[7] http://datascience.columbia.edu/master-of-science-in-data-science/
[8] http://dsi.virginia.edu/academics
[9] http://analytics.ncsu.edu/?page_id=1799
[10] Enrollments for NCSU, Columbia, NYU and UVA are 120, 90, 60 and 47, respectively.

social sciences as well as the humanities. Researchers in iiD have been collaborating with SSRI ever since their co-location in Gross Hall. The Data+ program just completed its most successful summer yet, and is a great example of this collaboration.

With SSRI on the second floor and iiD on the third floor, Gross Hall has therefore emerged as an ideal place from which an interdisciplinary Data Science MS degree along the "Data to Decision" vision can emerge. In order for it to be true to this vision, though, it cannot be siloed from the rest of Duke. Rather, it must be a cornerstone of an intellectually grounded university-wide, interdisciplinary effort that innovates the way quantitative methods and computational ideas are taught and applied in specific domains across the disciplines. Duke is one of the few places in the country that has the infrastructure and collaborative culture to create a program like this and thus distinguish Data Science from the many emerging programs that lack the ability to truly engage at the intersection of Data Science and its many and quickly changing application domains, or in the broader conversations about the social context in which data emerges and is conceptualized in today's world.

In fact, the Master's in Interdisciplinary Data Science (MIDS) emerges from SSRI's and iiD's shared commitment to this culture of interdisciplinarity. Duke has led the effort to organize interdisciplinary models and approaches, and the interdisciplinary Data Science program seeks to build on this foundation. Conventional quantitative methods are often siloed in the "hard" or social sciences, and subdivided further in the disciplines. Yet, the most innovative work in pressing social and applied problems isn't always possible within the epistemologies and approaches of a single discipline. The MIDS program acknowledges the need for technical proficiency and methodological training, but seeks to offer training that is both grounded in, and informed by the disciplines. The interdisciplinary environment at SSRI and iiD makes it a natural home for data science students whose interests do not fit squarely in a single discipline, and who seek careers that require quantitative, technical skills but not specialization in one field. The applications to natural and social sciences are obvious and exciting, but the program also seeks to build bridges with the humanities, where scholars are increasingly experimenting with data-driven methodologies and where, together with some of the interpretative social sciences, deep conversations about the very meaning we attach to data reside.

While the MIDS is designed to teach rigorous methods and techniques, we seek to develop graduates who are also well-rounded in critical thinking and communication skills. Our goal is to create "data literacy" in applied situations, which requires the ability to bring a critical eye, informed by multiple disciplinary perspectives, to data analysis. Our curriculum emphasizes communicating findings in clear, jargon-free language, an essential skill as graduates enter professions in which they must communicate findings to "clients" in business, the non-profit world, or government. In other words, we seek to combine skills in "hard" quantitative research with "soft" skills in critical thinking and communications. The ability to critically evaluate data becomes as important as the ability to persuasively report results, and yet, pedagogy in the former lags far behind pedagogy in the latter.

This commitment is most prominently translated into the proposed curriculum through the year-long Data Science Seminar and the team-based capstone project/seminar, which will be "learning and leading" opportunities for students to work on multi-disciplinary projects, including those located within Bass Connections. Two of the Bass themes are in fact housed within SSRI (Education and Human Development) and iiD (Information and Society), and SSRI and iiD are jointly collaborating with the Energy Initiative on an Energy Analytics Lab within Gross Hall (with the Energy Theme in Bass Connections representing the third theme housed in Gross Hall

within the Energy Initiative). Building upon Duke's pre-existing commitment to interdisciplinary research with a strong pedagogical focus, this final capstone project will mentor students in how data is used in a real-world situation while engaging them in project leadership, requiring critical acumen and team-building in addition to skills in quantitative methods. It will be enriched by the ongoing Data Science Seminar that makes space for critical engagement across many domains as well as connections to real world partners. Moreover, these experiences will force students to confront all of the challenges in marshaling, analyzing and visualizing data in an environment with multiple users, from multiple disciplines, and with multiple skillsets—precisely the kind of environment they will find after graduating, whether they go on to doctoral work or turn to professional life. Substantial funding for these projects is built into the budget.

We propose a 3-4 semester program, with possible pre-matriculation summer "boot camps" (initially organized with the math department) for those having critical skill gaps. Duke, including several faculty involved with the MIDS, already has highly successful Coursera courses whose content is available for such boot camps, and more are in the development and planning stages. SSRI also recently launched ModU, a platform for series of educational modules to be packaged in different types of "playlists." The first series of 140 (relatively) jargon-free modules focused on causal inference has already been developed, and has resulted in external partnerships with eBay to deepen some of the content. The proposed MIDS program would typically be taken over 1.5 to 2 years, but could in some instances be completed in 1 year if a full summer semester is included as part of the program. As data science is an enormously dynamic field, we have intentionally planned a great deal of flexibility into our curriculum and expect the changing requirements of the field to be reflected in our curriculum over time.


## Section 2: Relationship to existing programs

Data science shares some intellectual foundations with several quantitative Master's programs at Duke, but data science, which is highly applied and methodologically and topically agnostic, is emerging as a new field. As articulated in Section 1, we have taken a particular view of data science – interdisciplinary and team-based – that we believe distinguishes our program from those offered at peer institutions. We believe this interdisciplinary and team-based brand of data science also differentiates the MIDS from existing quantitative programs at Duke. Equally important, though, the MIDS creates opportunities for exciting synergies with existing quantitative programs at Duke. Below, we provide brief overviews of the Duke curricula most similar to the MIDS we propose, and then examine the possible overlaps and opportunities for collaboration between MIDS and these programs.

Before doing so, however, we offer some general observations. Most other quantitative Masters programs at Duke appropriately offer deep training in particular quantitative areas – training that in many ways mirrors PhD training in similar areas. This is one of their great strengths, but it is not what MIDS aims to achieve. Rather, the brand of data science we seek to advance requires less depth in any particular area while focusing on the connections between them. While our students will need to "dive deep" in some quantitative area, our students' awareness of the broader data landscape and its application to many domains will be what distinguishes them. Unlike students in other programs, they will have to take courses offered from faculty in multiple departments, and they will need to bring their skills to team-based and interdisciplinary environments much more so than would ever be expected in a more disciplinarily rooted

program. Our approach is furthermore one that will tend to move from application to method rather than from method to application and will place greater emphasis on emerging technologies driven by needs in applied projects. Given the team-based nature of our program, communication will play a more central role, as will engagement with students in domain electives.

When thinking about synergies, we should explore (1) how MIDS can benefit from quantitative programs elsewhere at Duke and (2) how other quantitative programs at Duke can benefit from MIDS.

(1) _Benefits for other programs from MIDS_: First, other programs at Duke will benefit from courses that would not otherwise be offered at Duke. These include the core courses in our proposal as well as elective courses that will emerge in response to our RFP outreach efforts as we seek collaborations with other departments. For instance, a program like the SOM's Biostatistics program may seek to develop a data science track that will utilize some of our core courses. Second, we have been careful to explicitly recognize that other departments and programs should financially benefit from collaborations, and we hope that these incentives will translate to new intellectual collaborations as well as spillovers between students that fall in different places within the larger data science landscape. Third, students in other programs will be able to participate in the broader interdisciplinary data analytics community we seek to facilitate, and to connect with projects and applications that emerge within that community. Fourth, we have set considerable funds aside for directly seeding faculty-led projects that engage our students, and these projects will no doubt provide opportunities for other students as well.

(2) _Benefits for MIDS from other quantitative Masters Programs_: First, with our program intentionally based on partnerships, many of the important electives for MIDS are already housed in departments of other programs across the university or will emerge within those programs. Second, MIDS would not be possible without the faculty in other programs whose research programs we hope to enhance but whose leadership we rely on for many of our team-based capstone projects. Third, given the aim of having our students understand their place – and the place of others – in teams focused on data projects, the deep disciplinary expertise of students in other programs will directly enhance the experience of our own students.

_Quantitative Masters Programs at Duke_

Master's in Statistical Science (MSS)
The MSS is a 2-year graduate degree program that provides a modern, comprehensive education in statistical theory, methods and computation and is similar to the first two years of the PhD program in Statistics. Courses and projects emphasize stochastic model-based approaches to problems of inference and prediction, highlighting Bayesian and machine learning methods. Students develop skills and expertise in modeling complex and potentially massive data sets and advanced computational methods while gaining exposure to a broad range of application contexts. The MSS is attractive to students interested in professional careers in industry, business and government, as well as those interested in pursuing PhD studies in statistics, the social sciences, or the natural sciences.

Master of Science in Computer Science (MSCS)
The MSCS is a 2-year program that provides theoretical and practical training in technologies and applications. There are 3 options for the degree: course only, with a required portfolio and oral exam, thesis, or project option with an oral defense. The program requires students to complete 4 or 6 computer science courses, and at least 4 approved electives in a field related to CS or the student's area of concentration. With this degree, graduates are able to move into a research career or pursue a PhD.

Master's in Statistical and Economic Modeling (MSEM)
The MSEM degree is a joint master's program of the Departments of Economics and Statistical Science, which prepares graduates in critical areas of statistics and computation linked to economics, finance and related applications, to be competitive in moving on into professional practice, R&D in commerce, industry and government – or to pursue PhD work. It is a relatively small program, and – in light of the fact that both economics and statistics now have large and mature Masters programs, there is an open question as to how active the joint program between the departments will be in the future.

Master's in Economics and Computation (MSEC)
The MSEC degree is a joint master's program of the Departments of Economics and Computer Science. The joint field of economics and computation has recently emerged from the shared intellectual needs of these two fields. Like MSEM, this is a relatively small program serving a specialized need at the intersection of the two departments.

Master of Management in Clinical Informatics (MMCi)
The MMCi is a 1-year graduate degree program developed through a partnership between the University's School of Medicine and The Fuqua School of Business to meet the demand for informaticians and IT-conversant health professionals. The program seeks to prepare graduates to operationalize, manage, and evaluate health IT in a variety of settings, including academia, government, and industry, and to succeed in managerial and executive positions in health care.

Master's of Engineering Management (MEM)
The MEM is a 1-year graduate degree program that can also be completed by current Duke undergraduates through Duke's 4+1 option (1 additional semester or year). This program provides a core management curriculum with courses in marketing, management, finance, intellectual property and business law. A workshop series helps students develop leadership, negotiation and communication skills. Technical electives allow students to choose a management concentration in technology, innovation, operation, entrepreneurship, financial engineering and other engineering disciplines. This degree is an alternative to an MBA for young engineers and provides a traditional campus based option for full and part time students as well as an online option for working professionals.

In addition to these existing programs, the Fuqua School of Business is launching a Master of Quantitative Management (MQM). While this degree is naturally aimed more narrowly at business students (as is true for many data science programs (see the graphic in Section 1.4)), it is also focused heavily on data science and analytics and emphasizes communication skills. Thus, even at this early stage, we have already begun discussions about potential electives that could

be shared, and we see MQM students as natural beneficiaries and participants in our Data Science Seminar.

Each of these Duke programs share commonalities with the Master's in Interdisciplinary Data Science we propose, but ours does not remotely duplicate any existing program. Two of the Master's degrees (MSEM and MSEC) are relatively small and topically focused within the discipline of Economics, while the MMCi focuses specifically and appropriately narrowly on health IT, and the MQM is aimed exclusively at business school students. Of the other degrees listed, the MSS and the MSCS have a basic science focus rooted in strong disciplines, whereas the core of the MIDS program is much more focused on drawing connections across the topics covered in great depth in these departments. MIDS will certainly benefit from the strong offerings in computer science and statistics, particularly for advanced tools electives, while MSS and MSCS will benefit from the application-focused Data Science Seminar where they can be exposed to, and draw inspiration from applications that require those skills. The MEM and MQM are the most professional of the degrees, preparing students exclusively for careers in business. While MIDS expects to form strong partnerships with industry, our program emphasizes the broader uses of data science, often in areas where data is only beginning to be recognized as important.

The data science program we propose embraces the need for deep training in the quantitative sciences, but is more applied and less focused on a particular discipline than is the MSS, the MQM or the MSCS. At the same time, The MIDS interdisciplinary nature makes it broader than the other degree programs. Its matriculants' interests will not typically fit in a single discipline, and they will seek careers or further advanced training across diverse fields that require a combination of quantitative skills and domain expertise. Unlike most of the other quantitative programs, ours is particularly focused on developing team-based skills – building on a view of interdisciplinarity that emphasizes strong disciplinary expertise coming together in diverse teams of scholars and practitioners. In this sense, it is tailor-made to leverage the Bass Connections programs around Duke, providing a layer of graduate engagement that emphasizes project management.

Despite these differences, our approach to data science necessitates that we cultivate relationships with programs that have different educational focuses, and that we continue to seek fruitful linkages with other programs on campus that have not yet incorporated quantitative approaches into their curricula. Other departments at Duke have demonstrated strong interest in the opportunity to design methodological courses with multidisciplinary appeal for the MIDS program, which have, historically, not attracted enough students to justify offering the course in a single department. For example, Mathematics and Statistics see value in joining forces on the Data Science Seminar for training first-year graduate students, and mathematics, as evidenced in their supporting letter, has expressed particularly strong interest in developing close ties with the Masters in Interdisciplinary Data Science (MIDS). We expect such collaborations to multiply, through our RFP outreach efforts to develop elective courses for the program.

We also believe that there is potential for MIDS and these other Master's degrees to form an interdisciplinary community of quantitative Master's-level students on campus to the benefit of all these programs' students and faculty. Indeed, we have already experimented with forums that might lend themselves to this, forums such as the ongoing iiD/SSRI co-sponsored Data Dialogue in which representatives from industry, government or the non-profit sector are paired with

faculty to explore the intersection between real world applications and frontier data science. This Dialogue meets regularly and already attracts students from many programs, and will form the basis for the Data Science Seminar that is one of our core requirements in the proposed Data Science MS degree.

In addition to other Masters degree programs, our program seeks to build relationships with both PhD and undergraduate students. At the undergraduate level, both undergraduates and MS students will benefit from having undergraduates involved in the faculty led projects that will form the basis for the MS capstone experiences. In many (perhaps most) cases, these projects are likely to be Bass projects in which the MS students take on the role of project managers. At the PhD level, a number of PhD programs in other disciplines will benefit directly from courses that are offered through our MS program, and our program may provide further infrastructure to increase Duke's success in attracting funding from agencies like NSF and NIH for doctoral training programs. In addition, discussions with social science departments have revealed that certain Masters courses, particularly in statistics, have become less open in recent years to PhD students as the demand from quantitative Masters students has grown. To insure that our MS program does not add to these pressures and helps relieve existing pressures, we propose in our budget to set aside resources for funding seats in Masters level statistics and computer science courses for social science PhD students.

## Section 3: Statement of resources needed for the program

As outlined below, most of what is required for the Interdisciplinary Data Science MS is already in place, with resources generated from tuition providing the rest and filling gaps.

### *a.  Business plan explaining program expenses and how they will be covered.*

For detailed anticipated program expenses and revenue, please see Appendix A: Five-Year Budget.

The program's largest budget category, on average, is financial aid. We anticipate an average support of 20% of tuition in the first, second, and third semesters. This amount increases to 75% in the fourth semester, so long as students are in good academic standing and are not externally funded. The rationale for the added financial aid in the fourth semester is that we believe most students will benefit from the extended time in the program, particularly as they become involved in team-based projects, and we would like to lower the financial barrier for this to happen. We envision that financial aid will, in some cases, be linked to mentorship of undergraduate research projects, Bass Connections or participation in related support activities such as those around the help desk in SSRI.

The next most significant program expense involves investments in courses for the program. We budgeted an initial amount of $150,000 for the fixed cost of offering the core courses, either through SSRI/iiD directly or in partnership with academic units. In addition, we include a generous budget to invest in departments and other academic units that contribute elective courses (jointly approved by MIDS and these units with pedagogical benefits to students in both partnering units in mind). Approximately $160,000 is set aside for this purpose in our initial cohort year, with this amount eventually climbing to almost $800,000 annually. For budgeting purposes, we derived these amounts based on an assumption of "pass-through payments" (to

partnering units) of $4,000 per MIDS student seat in an elective course. While these partnerships may in some cases literally take the form of such pass-through payments directly to departments based on MIDS enrollments, we are open to structuring these collaborations in other ways that work well for departments depending on their own circumstances. For instance, it may be that advance course developments funds are needed, or that TA support is essential. We envision annual RFPs to departments to propose collaborations through courses, with specific requests for departmental feedback on how resources might be most helpful in structuring the collaboration and selection criteria based on demonstrated mutual pedagogical benefits. We recognize that the $4,000 per student per course budget for such collaborations might be revisited in the future (particularly as other discussions in A&S are underway to structure similar mechanisms related to other Masters programs). We furthermore recognize that we may learn through this RFP process that there are a few standard models through which our partnerships will be structured, and that the "pass-through" mechanism used to derive budget numbers may not be the dominant actual mechanism that emerges. Whatever mechanisms emerge are meant to both compensate units for teaching students in the program *and, more importantly,* to invite units around campus to envision mutually beneficial partnerships that could attract our students together with others from other programs.

In response to concerns raised by colleagues in some social science departments, the budget also includes a provision for the pass-through payments to apply to social science PhD students who seek to enroll in heavily subscribed quantitative Masters courses. It has been the experience in some social science departments that, as quantitative Masters programs have grown, it has become more difficult – sometimes impossible – for PhD students to enroll in courses that some programs have come to rely on. The budget provision for PhD pass-through payments is our attempt to help resolve this issue in graduate education at Duke.

Beginning in year 2, we further envision providing seed funding for incubating the types of faculty-led research projects that can also serve as capstone projects for MS students. Often, we assume such funding can supplement funding from other sources, whether these are external or internal from programs like Bass Connections or those offered through initiatives like the Initiative in Innovation and Entrepreneurship. At the same time, the MS program itself will benefit these projects through the leadership provided by our graduate students in vertically integrated teams.

Instructional and administrative salaries for the management of the program are also budgeted. The program will employ a DGS, an Associate DGS, an Associate Director, an Assistant to the DGS, and a seasonal work-study. In addition, the budget sets aside funds for developing career services for the program's students.

Finally, operational costs include funds for student and staff conferences, software, telephone lines, website development, and advertising.

These costs will be covered entirely by tuition revenue. Surplus net revenue will be invested in infrastructure at SSRI and iiD – infrastructure that has direct spillover for the MS program. In light of evidence of significant demand for data science degrees at peer institutions, we do not believe there is significant risk that we would not meet enrollment targets. However, in the event that we do not, a significant portion of the budget will by definition not be required to the extent to which many of the costs are not fixed costs. For instance, while core courses would have to be offered, the "pass through payments" that fund partnerships would decrease in proportion to lower enrollments, as would financial aid and necessary funding for capstone projects. Since

these variable costs represent the bulk of the budget, the financial risk from unexpectedly low enrollments is muted. In the initial year, we expect to work with the Provost on a bridge loan to launch the program, providing some financial flexibility – and we are prepared to tap reserves in the event that there is a budget shortfall.

*b. Review of available resources*

Together, SSRI and iiD constitute a combination of infrastructure and people that form a "learning community" in which ideas can connect in novel and interesting ways. A number of resources facilitate this:

**Space**: The Social Science Research Institute (SSRI) occupies a significant portion of the second floor of Gross Hall. SSRI also offers Team Rooms for students, as does iiD. A number of data-intensive interdisciplinary projects already operate in these spaces, as does the Initiative in Innovation and Entrepreneurship with which we will seek collaborations.  In addition, data-related collaborations with medical school faculty are emerging within the space from our joint 2015-16 Social Science Challenge.

**Computing**: Computational resources are central to the success of the MIDS program.  SSRI offers extensive resources for large-scale data storage, methods, and processing, and has both expertise and connections to the larger Duke Computing Research infrastructure including OIT and ITSO. This is complemented by SSRI's Protected Research Data Network (PRDN), designed to accommodate the growing sensitive data needs at Duke. Together with technical expertise in iiD, there is sufficient infrastructure to accommodate projects requiring a high level of security and/or extensive computational power. Funds to cover expenses related to marginal additions to that infrastructure are included in the budget.

Many types of projects will involve sensitive data, and the existence of an infrastructure for such projects will be critical for forming industry and government partnerships that involve sensitive data transfers. Current configuration specifications for SSRI's PRDN infrastructure allow for this to be readily accommodated while also offering state-of-the art training in data privacy and IRB processes – both conceptually and in practice – to students in the program. In addition, SSRI's strategic plan is focused on advancing not only a state-of-the-art technical and legal sensitive data infrastructure but also to be a national leader on new issues related to the ethics of working with data on human subjects. This is particularly relevant to a program like MIDS where personally identifiable data will often be central to projects undertaken by our students – and where non-sensitive data is likely to become increasingly sensitive as it is merged across data sets.

**Methods Centers and emerging standing labs**: SSRI houses the Duke Network Analysis Center (DNAC), the Duke Initiative in Survey Methods (DISM), the Interdisciplinary Behavioral Research Center (with extensive experimental labs), a new initiative on qualitative methods, and PhD level staff with expertise in program evaluation, statistical methods, data management and data privacy. iiD is home to expertise in advanced statistical and machine learning methods as well as related areas involving "big data". Both iiD and SSRI have significant communities of post-docs and graduate students involved in these methodological areas, including many that staff the SSRI Connection Bar.

In addition, SSRI and iiD are in the process of establishing collaborations with other entities that will result in standing labs revolving around data in specific contexts. We are, for instance,

launching a "Durham Lab" in collaboration with a new non-profit (DataWorks) – a standing lab, including an ongoing undergraduate course – that facilitates the use of existing data, the collection of new data and the merging of datasets across many domain areas but focused on the City and County of Durham. We anticipate similar standing labs to emerge out of Data+ and Bass programs – enriching the landscape of interdisciplinary and vertically integrated teams at Duke. These labs will form one of the foundations on which MIDS team projects can build.

**Connection Bar**: SSRI's Connection Bar, open 10 a.m. to 6 p.m. on weekdays during the academic year, is a help desk staffed by advanced graduate students and postdoctoral scholars with expertise in using both quantitative and qualitative methods and tools. Students will receive assistance with research projects through in-person, one-on-one consults, and MS students will have the opportunity to assist others at the Bar with projects.

**Data+ and Data Expeditions**: iiD (in partnership with SSRI and other units) has already established two significant data-related programs that link graduate and undergraduate experiences at Duke. Data+ (a collaboration between the Math Department and SSRI, and more recently the Medical School) is an intensive 10-week summer program in which small teams of undergraduates use full-time fellowships to explore datasets that faculty are interested in incorporating into larger projects that link to external partners, or that may lead to future Bass Connections projects. Graduate students currently mentor the Data+ projects, and MIDS students could be integrated into these programs as part of their capstone experience, particularly in the Data+ programs that extend beyond the summer, including those that are resulting in more standing labs (as mentioned above). Data Expeditions is a program that encourages pairs of graduate students to curate a data set to be used in undergraduate courses.

**Bass Connections**: Three of the Bass Connections themes are housed in Gross Hall – and their "owning" institutes/initiatives are already collaborating on joint big data projects. We anticipate that these will be a significant source of capstone projects for MIDS students. We will also seek to connect to the other two Bass themes (that are not housed in Gross Hall) to develop similar collaborations.

**SSRI and iiD Education Cores**: Both SSRI and iiD have education cores built into their structures, and these cores can play significant support roles for the MIDS program. We are particularly interested in developing modular on-line tools that can support flipped classroom pedagogy during the academic year and serve as boot-camp modules during the summer prior to matriculation. ModU was recently launched as a virtual platform providing such modular infrastructure in SSRI, and an increasing number of Duke-offered Coursera courses are adding to the growing stock of virtual resources. We are in conversations with partners around the campus – in A&S, Fuqua, the SOM, the Nursing School and Pratt – to further populate these infrastructures that, of course, also have substantial spillover benefits across many undergraduate and graduate programs at Duke.

The faculty boards in SSRI and iiD will implement a joint oversight body for the degree, with the directors participating ex officio. Primary administrative support will come from staff at SSRI, including the following academic advising and administrative issues: admissions, financial aid, registration, graduation requirements, and career advising.

*c. Statement of additional resources needed.*

No significant additional resource needs are anticipated. The possible exceptions to this include a possible marginal increase in space needs for supporting staff for the MS program and a Provost-level bridge loan to start the program.

### *d. Potential for outside funding.*

We (i.e. SSRI and iiD) have begun to develop partnerships with government, non-profit, and industry groups whose data needs may be supported by the program while benefitting students in this (and related) programs. Data+ already has a strong record in this regard. As the program progresses, we expect such partnerships to thrive and new ones to form. For instance, we are considering possibilities of regularized student consulting clinics in longer-run partnerships with government, non-profit, and industry groups. Foundation funding, particularly in the humanities, may be available for projects that directly benefit the program.  In addition, grant support from federal agencies is available for longer-lasting projects incubated within the Interdisciplinary Data Science program, as well as for building greater capacity through cyber infrastructure grants (which we are already pursuing) and NIH, NSF and DoD training grants. SSRI's pre- and post-award grants infrastructure is prepared to assist in obtaining such outside funding and managing it when proposals are successful.

### *e. Financial aid, scholarship, or fellowship mechanisms to be employed.*

Financial aid will be offered by SSRI and iiD and administered in SSRI. This aid consists of partial tuition waivers in the first, second, and third semesters – anticipated to be 20% of tuition on average. Many students may also be eligible for a 75% tuition waiver in their fourth semester, so long as they are in good academic standing and are not externally funded. Financial aid will be awarded to those who apply and demonstrate a need. Particular attention will be paid to employing financial aid mechanisms to the end of insuring significant involvement of women and underrepresented minorities in the program.

### *f. Five-year student, faculty, and resource projections*

Please see Appendix A: Five-Year Budget.


## Section 4: Students

The Master's in Interdisciplinary Data Science (MIDS) will recruit students principally from two pools:  current and recent undergraduates seeking further training in order to attain their career goals, and existing professionals pursuing training to advance their careers or branch into new specialties related to their current expertise.  Although the curriculum we propose will serve these two populations in similar ways – that is, upon matriculation, students' typical experiences in the program will be largely indistinguishable, regardless of their starting point – our planned recruitment strategies do vary across the two populations, and we therefore discuss each in turn. For both subsets of potential students, the Associate Director will take responsibility for developing and implementing recruitment strategies, with advice from the faculty oversight board and assistance from select current students and alumni.

Before discussing specifics relevant for the main groups of potential students, we should be explicit about the pre-requisites and core competencies we expect of successful applicants. We believe it is prudent to set strong pre-requisites as the program is launched and experiment over

time and additional infrastructure to determine the extent to which intensive boot camp preparation can fill in gaps. Thus, our initial hard pre-requisites will be those we find in place in most other data science programs: single variable calculus, linear or matrix algebra, introductory statistics and introductory computer programming (in any language). As MIDS becomes more established, we will explore mechanisms to accommodate more diverse backgrounds that might enrich the program but would require pre-matriculation training.

While we believe it is important to maintain flexibility around the precise mix of students we would like to aim for, and learn from experience in the initial years, we should also be explicit about the underlying principles that should guide our learning. We do not intend for this program to exclusively serve recent Duke undergraduates; rather we aim to attract recent undergraduates who have relevant domain and/or technical backgrounds and are looking for a way to fill in gaps and integrate what they already know. We do not intend this program to exclusively serve students with work experience; rather, we believe the experience of recent undergraduates in our program will be strengthened by real world experience brought to the program; and those bringing such experience will be strengthened by the perspectives of more recently trained undergraduates. The mix of students in the program should therefore be sufficient to accomplish these aims. In addition, we do not envision a program exclusively tailored to either domestic or international students; rather, we believe the program is strengthened by diversity in background, with the important caveat that the team-based approach of the program requires English language fluency.

Finally, we currently aim for a steady state program size of 30 students but recognize that we may, within a short period, find ourselves in the position of having qualified applicant pools of much greater size. The temptation will exist to expand the program, and we may learn of ways to do so. However, the current target size is one that lends itself well to individualized attention to needs of students with different backgrounds and different goals. Particularly in light of the central feature of interdisciplinary, team-based engagement with problem solving, we believe that this is the best way to distinguish this program from others. Having said that, it is our experience that one learns a lot about what makes a program like this work as one gains experience. Over time, it becomes apparent what aspects of a program like this can be systematized and scaled, and what aspects continue to require a more personal approach. To the extent to which a number of features of the program can ultimately fall into the former category, scaling becomes easier. Data+, for instance, was initially a small program but, as experience was gained, has grown beyond its original target size. The same has been true for the highly successful MA program in economics. The limiting factor for MIDS will likely be maintaining quality of team-based projects involving MIDS students.

- *Recent undergraduates*

  We will target and actively recruit students from top colleges and universities, specifically, from universities similar to Duke in their attributes and quality (top-notch research faculty with strong undergraduate teaching and plentiful opportunities for undergraduates to develop as active members of research teams), and from selective liberal-arts colleges that provide students with similarly focused opportunities. For both universities and colleges, we will be attentive to concerns about diversity and seek out opportunities to connect with under-represented groups (and institutions), including nearby Historically Black Colleges (HBCUs) such as NCCU and NC A&T.

While we will welcome applicants from a wide variety of disciplinary backgrounds, we anticipate focusing our recruiting efforts on departments whose major requirements include quantitative training that is both broad and deep. These departments will certainly include most math, statistics, and computer science departments, and may also include a number of natural science and social science departments that teach quantitative approaches. This does not, of course, preclude our welcoming student applications from more interpretive or humanities-focused undergraduate backgrounds, particularly as the program matures. Within Data+, for instance, we have identified a cohort of humanities majors with serious coding skills, and they have made important contributions to Data+ teams.
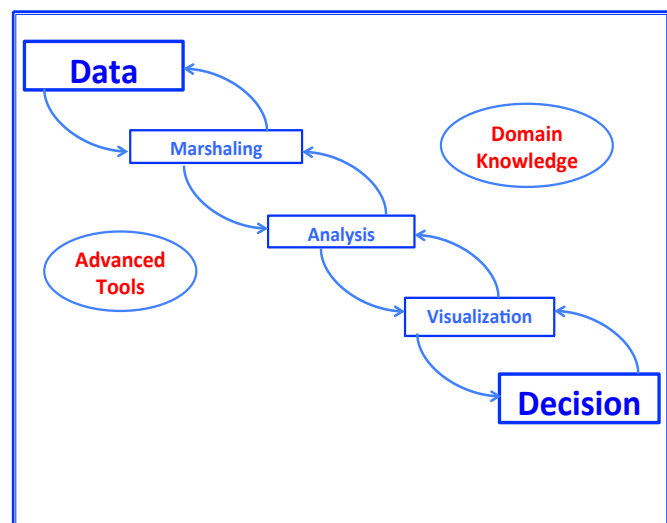
- *Existing professionals*

    Our immediate geographic area provides plentiful recruiting opportunities among professionals working at higher educational institutions, Research Triangle Park, and in non-profit, governmental, consulting, and technology institutions. Since our proposed program requires students to be physically present in classrooms during standard business hours, we anticipate that local professionals with scheduling flexibility and/or employer interest in supporting further training will be our primary targets. We plan to focus our recruitment efforts on this local population by reaching out to large employers in the Triangle area and their professional workers. We also anticipate that there will be select opportunities to attract professionals relocating to our region in connection with their spouse's or partner's decision to pursue educational or employment opportunities, and will seek to ensure that our faculty colleagues at Duke and other local universities are equipped to provide information about the MIDS when it may be of interest in such a context.

Of course, both for recent undergraduates and for existing professionals, we will track our recruitment yields and adapt accordingly. Extensive discussions with our colleagues in Statistics and Computer Science (that both have their own Masters degrees) have made us aware of the fact that the latter pool – existing professionals – will not always have technical backgrounds as strong as might be inferred from transcripts that are five to ten years old. This has made us considerably more attentive to planning for future summer boot camps and modular infrastructures leading up to matriculation in the program. Our plan is to start with modest support with a boot camp organized in collaboration with the math department and aimed at these students and then build on it over time.

## Section 5: Degree requirements

The curriculum of the proposed degree closely follows the "Data to Decision" cycle articulated as the guiding vision for the program (depicted again here). This cycle (depicted in blue in the diagram) forms the *core* of the program and gives rise to five

core courses. Advanced tools and domain knowledge (depicted in red in the diagram), on the other hand, are essential for enriching this cycle in service of data leading to "good" decisions. Students are given flexibility in which of these tools or sets of knowledge to emphasize through their choice of *electives* in the program.

This overview translates into the following degree requirements:

*Courses*
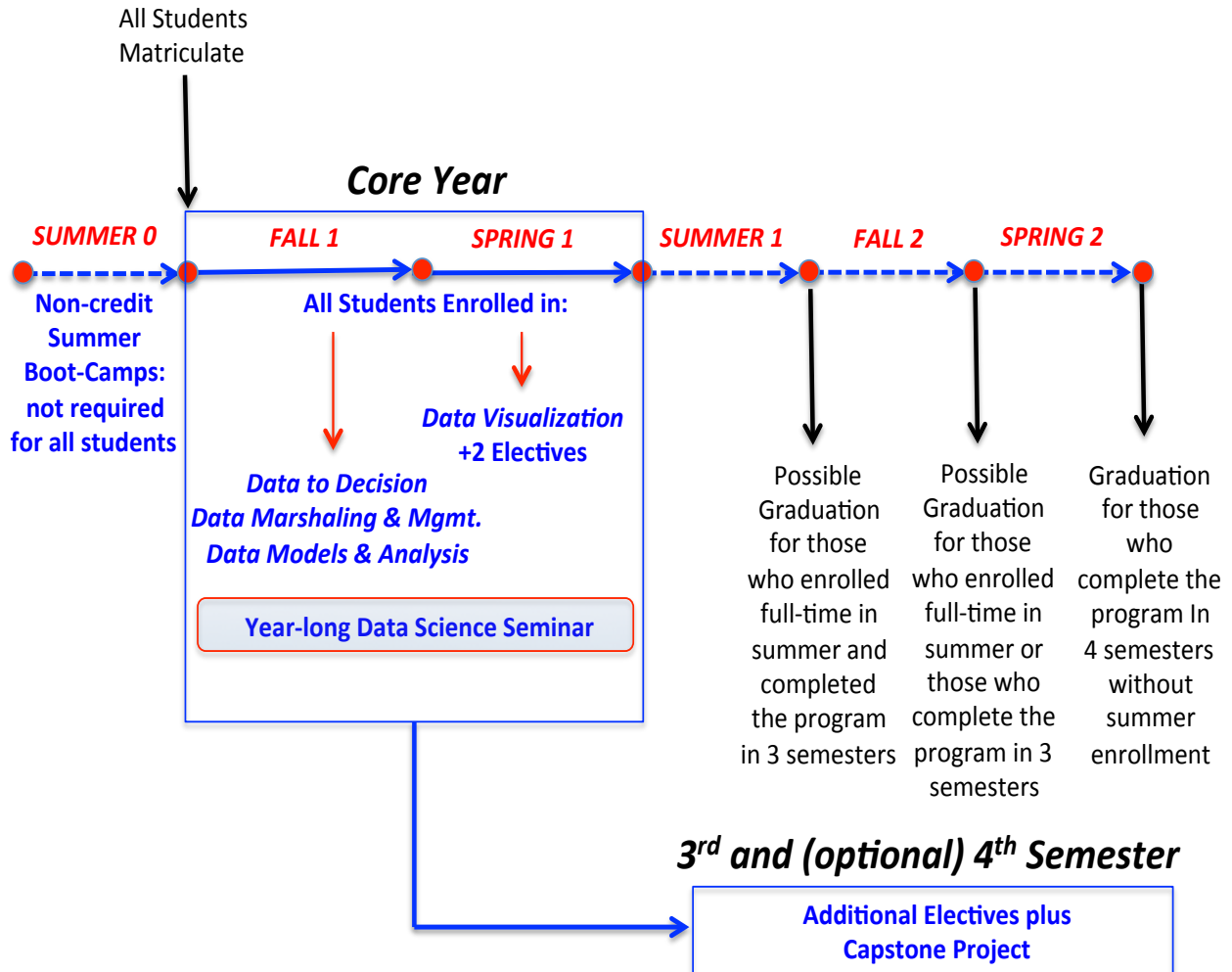·   30 credits over 3-4 semesters plus 4 AIRCR credits.
·   5 **core courses**:  *Data to Decision*, *Data Marshaling & Management*, *Data Modeling & Analysis*, *Data Visualization*, and a 2-semester (half-credit) *Data Science Seminar*.
·   Capstone project/seminar
·    5 *electives*: Select graduate-level and advanced undergraduate courses to build *advanced tools* and *domain knowledge* related to those tools (with at least 2 electives in advanced tools and at least 2 in domain knowledge).
·   Elective selections are subject to approval by the Director of Graduate Studies.

*Application*
·    Team-based capstone project in the students' final semester(s): Hands-on learning and leadership development on a real-world, typically team-based data project.

The diagram on the next page illustrates these requirements and the flexible paths along which they can be satisfied. We emphasize that, while a 12-month path is possible, we believe it will be applicable to only a very limited number of students. In fact, we are convinced that 18 months will be necessary in almost all cases, and we think a significant fraction of students will benefit from staying for an additional six months (and thus we provide 75% tuition relief for the fourth semester in the program). Students will have to petition for special exemption to graduate in 12 months, with the burden of proof for demonstrating that this is the optimal path resting with the student. From the perspective of the MIDS program, the primary argument for including a 12-month option is based on the possibility that some truly exceptional students whose background would markedly enrich the program are only able to devote 12 months to it. Such students would not only have to surpass the pre-requisites of the program, but we would furthermore have to be convinced that they are exceptional in ways that generate significant spillovers to other students. The 12-month option will therefore not be prominently featured in any material advertising the program, and students applying with the intent to graduate in 12 months will have to be approved not only by the normal admissions process but signed off by the governing committee of MIDS.

**"Summer 0"** is the summer prior to matriculation. In future years, summer boot camps will be offered, some within iiD and SSRI and others in partnership with collaborating units (as supported, for instance, in the letter of support from the math department). We will begin with a boot camp organized with the math department and then learn during our initial year what types of boot camps are most useful, and who the best partners for offering them might be. For the initial cohort we will, when necessary, advise students between their admission and matriculation on what modular preparation they might want to take prior to starting the program, but – in light of our initially conservative admissions criteria – we don't expect this to be necessary in most cases. These support structures are not required for all students, but may be required for some admitted students who need to fill in gaps in their background prior to starting the program.

All Students
Matriculate

**Core Year**

| SUMMER 0 | FALL 1 | SPRING 1 | SUMMER 1 | FALL 2 | SPRING 2 |

**Non-credit Summer Boot-Camps: not required for all students**

**All Students Enrolled in:**

**Data Visualization +2 Electives**

**Data to Decision**
**Data Marshaling & Mgmt.**
**Data Models & Analysis**

**Year-long Data Science Seminar**

Possible Graduation for those who enrolled full-time in summer and completed the program in 3 semesters

Possible Graduation for those who enrolled full-time in summer or those who complete the program in 3 semesters

Graduation for those who complete the program In 4 semesters without summer enrollment

**3rd and (optional) 4th Semester**

**Additional Electives plus**
**Capstone Project**

The **"Core Year"** consists of a fall and spring semester. Students start as one cohort in the fall, and all take the same 3 core courses and year-long (half-credit) Data Science Seminar. Spring semester of the core year has all students take the core Data Visualization course but also has students begin to select electives.

Following this, students will need to complete the program in a **"3rd and (optional) 4th Semester"** that consist(s) of the remainder of their electives and either a semester-long (or, if they take a 4th semester) or a 2-semester-long team-based capstone project. Some, depending on progress in the program and availability of necessary summer courses, may have the option of taking their third semester in the summer, thus graduating 1 year after matriculation. Others may enroll in their third semester during the summer, choose to stay for a fourth semester in the fall, and then graduate in December. Others may take the summer off (to participate in an internship, for example) and graduate in December after three semesters of enrollment, or extend their program to a fourth semester in the spring. A key feature of this flexibility in the speed with which students progress through the curriculum, of course, is the ability to serve students coming from different settings – e.g. students enrolling directly after completing their undergraduate learning versus those who may be taking a short leave from their established and ongoing professional roles, and with varying degrees of flexibility in their own schedules.

The following provides an initial list of potential electives from two "buckets" within A&S – advanced tools and domain knowledge related to those tools. These lists are not meant to be exclusive, with electives subject to approval in each case by the DGS, and with ongoing flexibility to add electives as new courses and opportunities for partnerships emerge. In addition, the list is not meant to suggest that only departments in A&S are potential partners; already, we have partnerships established with Pratt, have strong interest from the Nursing School, are discussing collaboration with Fuqua (and their new Masters program) and are in active dialogue with the SOM's Masters in Biostatistics. We furthermore envision providing seed funding for new elective courses that could serve both our program and others across the university. One example, for instance, is an emerging effort to catalyze the community of scholars who employ qualitative data in their research and to engage them directly in larger dialogues around how we conceptualize and use data. We hope that this effort will result in a broad interdisciplinary course around deeper non-technical issues that are nevertheless fundamental to the role that data science is assuming in today's world. Another example is our discussion with the Masters in Biostatistics (in the SOM) where we have identified broad areas of mutual interest.

| | |
|---|---|
| *Potential **advanced tools electives in A&S***: | STAT 521, 523, 561, 571;<br>COMPSCI 316, 516, 570, 571, 579;<br>ECON: 608, 612, 613, 703, 707, 883, 887<br>MATH: 361, 541, 545, 561, 563, 640<br>PSY: 767, 768, 770, 781S<br>POLSCI: 733, 748<br>SOC: 720, 726S, 728, 790 |
| *Potential **domain knowledge electives in A&S)*** (Courses in () not currently offered): | ARTHIST: 231.01, 508S.01, (541S.01)<br>CLST: 544L<br>ECON: 432S, 461, 471, 514, 553, 554, 673, 674, 881<br>MATH: 404, 573S, 581, 582<br>PSY: 353S, 673S, 681S, 685S<br>POLSCI: 230, 330, 508S, 550S, 551S, 632, 634<br>PUBPOL: 574, 597S, 605, 607, 633, 637<br>SOC: 333, 534, 643, 690<br>STA: 601, 640<br>VMS: 242.01, 567.01 |

(Course names for these course numbers are listed in Appendix E.)

At the conclusion of our program, a successful student should know how to:

- Think "data analytically" and do so within the chosen domain area
- Clean and prepare messy data sets using state-of-the-art methods
- Design and implement a strategy for applying methods in probability, statistics, linear algebra and optimization to diverse data sources in order to solve problems in any subject domain
- Exploit principles of visualization science to identify hidden patterns in data and communicate effective data stories
- Generalize principles of data analytics from case studies and presentations about how data were used to solve real problems in industry, government, nonprofits and academia

- Translate the results of quantitative analysis into succinct presentations for general audiences
- Contribute effectively to interdisciplinary data analysis teams and lead certain aspects of such teams
- Appreciate the growing ethical concerns around personally identifiable information on human subjects – even in cases where legal requirements for protecting such data do not yet exist.

**Section 6: New courses to be offered in connection with the program and identification of teaching faculty.**

We propose to collaborate with participating units to offer five new courses. These five courses form the core of the Data Science curriculum.

***Data to Decision*** – This course introduces students to the ways in which data can be used to shape, inform, or determine decisions and examines the data elements and analytic approaches that make meaningful decision-making possible.  The course explores issues related to inference and generalizability, and the extent to which data analysis can permit automation of decision-making, versus the extent to which a human element remains essential and at what stages of the process.  The course will cover probability, prediction, uncertainty, decision trees, sensitivity analysis, as well as an overview of Markov models and cost-effectiveness analysis. The course's focus centers on methods and problems relevant to big data, and incorporates an examination of the difference between complicated but predictable contexts, and complex and intrinsically unpredictable ones.

***Data Marshaling & Management*** – This course examines the different tools that are employed in data collection, including experiments, surveys, web scraping, mobile devices, and administrative records.  It considers the issues involved in decisions made during and about data collection as well as in those  made while combining data from multiple primary sources, over time, or across diverse contexts. It provides an overview of state-of-the-art standards, protocols, and software for naming, verifying, analyzing, citing, archiving, cataloging, distributing, translating, and replicating data used in research.     It also examines the opportunities and challenges involved in data sharing and replication in science, with an eye to promulgating norms that facilitate these practices.

***Data Modeling & Analysis*** – This course introduces key statistical and mathematical methods needed in the practice of data science. It covers basic methods in probability, statistics, linear algebra, and optimization. It provides an overview of linear least squares regression and its extensions to nonlinear regression, principal component analysis and dimension reduction, density estimation, and kernel methods. It also introduces basic methods in optimization, including linear programming and gradient descent, with some of their applications in machine learning. Computational aspects of all the above will be at the forefront, to ensure stability and scalability of the proposed techniques. Data analysis problems and small projects will be part of the course.

***Data Visualization*** – Visual representations of information can help us explore, understand and communicate the patterns hidden in data by displaying it in ways that augment human memory and cognition. To do this effectively, good design principles that capitalize on the strengths of the

human perceptual system must be applied. Computer algorithms from signal processing, machine learning and statistics may be used to transform data, leading to new understanding and actionable insights. This course will introduce the basic principles of data visualization, along with tools and techniques used to clean and process data for visual analysis. Course modules will be centered on analyzing real-world data sets, and the students' craft will be honed through group work and critiques. Practical software tools and languages, such as Tableau, OpenRefine and Python/Matlab, will be introduced.

*Data Science Seminar* – This year-long course is designed to expose and connect students to the many applications of data science skills and some of the deeper questions around data science that are often not confronted in more siloed programs. Each class features a different speaker from academia, industry, government, or non-profits, invited to share an ongoing data science project or related ideas. We are currently piloting this idea with considerable success in the iiD/SSRI sponsored Data Dialogue Seminar which pairs faculty with real-world partners in consecutive talks that explore the connection between applications and frontier data science. We are also exploring a collaboration between this seminar and weekly SSRI seminars held through the Duke Network Analysis Center (DNAC) directed and overseen by Jim Moody (Sociology).

*Course Staffing*

At this point, **Daniel Egger** (Pratt) is slated to teach the *Data to Decision* course, drawing on online modules that are already highly successful in a Coursera course, as well as bringing iiD and SSRI faculty into the course, both to introduce students to these faculty and to draw on their expertise. Given the modular content that has already been produced and new content that is in production, we envision much of this course taught as a "flipped classroom". Egger's co-instructor for the initial Coursera course, **Jana Schaich-Borg**, has recently joined SSRI as an Assistant Research Professor and may also be involved in the course. **Chris Bail** (Sociology) is slated to teach *Data Marshaling and Management*. He already teaches a successful PhD course that covers many of these topics and will likely generate much interest across departments with the MS version of that course. *Data Modeling & Analysis* will be offered in collaboration with the math department as that department is reimagining its linear algebra offerings. Faculty involved in this effort, and slated to teach in it, include **Paul Bendich**, **Tom Witelski** and **Ezra Miller**. Jana **Schaich-Borg** is slated to co-teach *Data Visualization* with **Eric Monson** (Computer Science), although we are also discussing the potential of involving faculty from the Media Arts and Science program as well as mathematics in the future. Finally, the *Data Science Seminar* will be run in conjunction with the Data Dialogue that is already successfully bringing together graduate students from across the campus. It will be run by **Robert Calderbank** (iiD), **Jerry Reiter** (Statistics), **Gale Boyd** (SSRi) and **Paul Bendich** (Mathematics). CVs for each of these faculty are included in Appendix B.

Any program that aims to form broad networks of partnerships is, of course, vulnerable to faculty turnover in partnering departments/schools or changing departmental/school needs trumping teaching needs in the program. This is less of a concern for the electives in our program as many of these already exist in some form, and new ones almost certainly will emerge from annual RFPs. The concern is more pronounced for the new core courses that we have proposed. We are fortunate, however, to have recently appointed two dynamic young faculty (and proven teachers) – Jana Schaich-Borg and Mariano Tepper – who could, in principle fill gaps in any of these courses. We furthermore expect to collaborate, as we did in the case of Chris Bail, with

departments in departmental hiring efforts where synergies emerge, and our strong alignment with the math department insures further stability. Data+ has quickly become very successful in attracting outside funding and has thus provided a nice infrastructure we can leverage, as have the SSRI efforts to build methods infrastructure in Gross Hall. Strong faculty support in SSRI and iiD offer additional insurance that inevitable faculty changes can be managed and the core of the program offered consistently.

*Program Leadership*

As indicated in the budget, the proposal envisions considerable faculty leadership as it seeks to develop its innovative structure and collaborations while also incubating partnerships with firms, government agencies and non-profits. In particular, we envision both a *Director of Graduate Studies (DGS)* and an *Associate Director of Graduate Studies (ADGS)*, with one having expertise linking to the social sciences and one to data science (not necessarily connected to social science). We furthermore envision *Co-Directors* that jointly chair an advisory board and develop collaborative initiatives to grow the program.

In the initial years of the program, we have identified **Gale Boyd** (SSRI) as DGS and **Jana Schaich-Borg** (SSRI and iiD) as ADGS. Gale Boyd directs the Research Data Center (RDC) within SSRI, a center that provides access to Census and other federal data through a secure facility within SSRI. Jana Schaich-Borg has recently joined the SSRI faculty and has a broad interdisciplinary research program that connects to both neuroscience and ethics. She is deeply interested in the intellectual program of Data Science as an interdisciplinary endeavor. We also propose that the Co-Directors of the MS Program will always be the Directors of SSRI and iiD, currently **Thomas Nechyba** (Economics and SSRI) and **Robert Calderbank** (Computer Science and iiD), or someone designated by them and approved by the advisory board. In the initial years of the program, the current directors of SSRI and iiD will serve in these roles.

## Section 7: Statement on Consequences

We anticipate no negative consequences for Duke graduate or undergraduate programs; rather, we expect that the expanded course offerings of the Master's in Interdisciplinary Data Science will enhance intellectual life for students in a range of programs from across the university and lead to new intellectual collaborations across disciplines. Particularly strong connections to Bass Connections are anticipated.

## Section 8. CV of faculty

See Appendix B.

## Section 9: Diversity Plan

The program recognizes and values diversity and the need to attract women and underrepresented minority groups to the field. For us, this does not merely reflect attention to social responsibility but is also central to the academic mission of the program. This is not a program that simply focuses on conveying technical skills, rather, its distinctiveness lies in its connection to important societal problems as well as insights central to the humanities. It is in

such engagements that diversity becomes intellectually central; without it, important perspectives are not adequately represented and the integrity of the program is threatened.

The program therefore intends to devote significant recruitment efforts to these ends, including establishing connections with HBCUs and women in STEM programs. As is indicated in the staffing budget, the program is well staffed, and this is in part to enable us to assign specific responsibilities to staff for recruiting a diverse pool into the program. The financial aid process will similarly be partly focused on the goal of diversity. These and other recommendations are outlined in the Office of Institutional Equity's "Diversifying Applicant Pools" paper. The program wishes to work closely with this office to help monitor the diversity of our applicants, and suggest ways to expand our recruitment efforts if we are not able to achieve a diverse applicant pool.

We think we are particularly well positioned to overcome the stubborn barriers to the engagement of women and underrepresented minorities that plague data science programs elsewhere, and we have preliminary evidence to support this. The Data+ program (coordinated in iiD in collaboration with Math and SSRI) launched (in its expanded form) in summer of 2015 with a large set of data intensive projects. The pool of undergraduates that spent an intensive summer engaging with data has been split roughly 50-50 between male and female undergraduates and included a demographically more diverse group than is typical in STEM fields. Our working hypothesis for this unanticipated outcome is that data-intensive engagements are significantly more attractive to non-traditional applicants if linked to important topics across the social, behavioral, medical and humanities disciplines – precisely the goal of the interdisciplinary and team-based MS in Interdisciplinary Data Science. In fact, we are exploring the possibility of collaborating with education professionals to develop a more fully defined strategy of recruiting underrepresented populations into STEM through linkages to projects that connect to these disciplines in compelling ways. In two-year intervals, we intend to review which strategies have been most successful and refine the overall strategy based on what we are learning.

## Section 10: Career Development Services

In the past academic year, we began discussions with other quantitative Masters programs as well as the Office of Career Services to explore a more coordinated effort at increasing paths into careers for students from these programs. We expect to intensify these efforts as we prepare to launch the Masters in Interdisciplinary Data Science. Our discussions have focused on coordinated career and internship fairs, and a more collective infrastructure to exploit returns to scale from such larger partnerships. In addition, we will explore extending these collaborations to Pratt and Fuqua given the quantitative programs that exist, or are emerging in these schools. Over time, we aim to create a strong alumni network that facilitates connections between new Interdisciplinary Data Science graduates and potential employers in industry, government, the nonprofit sector and academia, and we already have a head start given the growing network that is emerging from Data+. Program staff will devote significant resources to forming and growing relationships between the program and potential employers, and cultivating an active alumni network with engaged alums willing to mentor students. The program will make every effort to connect students with opportunities that match their interests, and seek out employers with specific interests in recruiting Duke Interdisciplinary Data Science students.

Indeed, we do not see these efforts as disconnected from academic goals for students while they are in the program. We envision the team-based capstone projects growing out of connections to industry, non-profits and governments who seek to engage with us around data, and we are already seeing "real world" participants in the Data Dialogue Seminar interested in developing pathways for students that participate. Connections to potential employers therefore emerge in part directly from efforts to engage real-world partners with data-science needs in the experiences students will have while in the program. Along these lines, we anticipate fruitful collaborations with university initiatives like Innovation and Entrepreneurship that actively foster connections to employers who rely on data scientists.

**Section 11: A student learning outcomes assessment process indicating how data on student learning outcomes is to be gathered, and how it will be used to improve the program.**

Data science is an emerging and changing field, a fact that necessitates a flexible and iterative approach to defining requirements for student learning outcomes. At the same time, of course, as with any pedagogical program, students deserve clear and explicit guidelines concerning course and degree expectations. To balance these two facts, we plan to work closely with program faculty to develop criteria for evaluating student success both at the course and at the degree level. Course requirements will incorporate metrics for evaluations, and data generated from these evaluations will be evaluated through the SSRI Program Evaluation Team, a newly formed group within SSRI. This team is already engaged with other units on campus to more deeply embed evaluation as part of program development. Analysis of these data, along with program-level evaluation tools (administered to students at the completion of the capstone course), and input from program alumni, which we will seek at one-, two- and five- year post-graduation points, will inform our understanding of student learning and offer ongoing opportunities to update the curriculum and improve instructional delivery methods. This should help ensure that as the field of data science develops we meet the highest standard for ensuring that program graduates master its knowledge base and skills.

See Appendix D for a Student Learning Outcomes Statement and Outcomes Assessment Plan.

In addition, we have given significant thought to the question of what "success" means for the MIDS program as a whole – and how it might be evaluated periodically. We believe there are three essential criteria that should form the basis of such an evaluation: (1) the number and quality of applicants; (2) the quality of post-MIDS placements; and (3) the amount and diversity of department and school engagements (across Duke) emerging from MIDS.

**Section 12: Letters of Support**

See Appendix C: Statements of Support.

## Appendix A: Five-Year Budget*

| COSTS | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
|---|---|---|---|---|---|
| **Administration** | | | | | |
| Faculty Co-Director | | | | | |
| DGS | $40,000 | $41,200 | $42,436 | $43,709 | $45,020 |
| Associate DGS | $30,000 | $30,900 | $31,827 | $32,782 | $33,765 |
| Associate Director | $75,000 | $77,250 | $79,568 | $81,955 | $84,413 |
| Assistant to DGS | $45,000 | $46,350 | $47,741 | $49,173 | $50,648 |
| Work Study | $5,000 | $5,150 | $5,305 | $5,464 | $5,628 |
| Fringe | $51,675 | 53,024 | 55,029 | 57,106 | 59,258 |
| **Instruction** | | | | | |
| 5 Core Classes | $150,000 | $154,500 | $159,135 | $163,909 | $168,826 |
| Pass Through Payments (including PhD Support)** | $220,000 | $620,000 | $750,000 | $840,000 | $840,000 |
| Boot camp Funding | $20,000 | $50,000 | $75,000 | $75,000 | $75,000 |
| Capstone Projects | | $200,000 | $250,000 | $300,000 | $350,000 |
| **Financial Aid** | $198,000 | $746,460 | $957,013 | $1,127,536 | $1,172,475 |
| **Career Services** | $30,000 | $50,000 | $50,000 | $50,000 | $50,000 |
| **Operation** | | | | | |
| Web Site | $30,000 | $10,000 | $10,000 | $10,000 | $10,000 |
| Supplies | $2,000 | $2,000 | $2,000 | $2,000 | $2,000 |
| Software (internal cost) | $1,000 | $1,250 | $1,500 | $1,500 | $1,500 |
| Data | $500 | $500 | $500 | $500 | $500 |
| Travel | $5,000 | $6,250 | $7,500 | $7,500 | $7,500 |
| Lunches | $2,000 | $2,000 | $2,000 | $2,000 | $2,000 |
| Staff Development | $6,000 | $6,000 | $6,000 | $6,000 | $6,000 |
| Phones | $1,243 | $1,243 | $1,243 | $1,243 | $1,243 |
| Miscellaneous | $10,000 | $10,000 | $10,000 | $10,000 | $10,000 |
| **Space & Computing** | | | | | |
| Connection Bar | $20,000 | $20,600 | $21,218 | $21,855 | $22,510 |
| SSRI Sensitive Data Services | $20,000 | $20,600 | $21,218 | $21,855 | $22,510 |
| Compute Services | $20,000 | $20,600 | $21,218 | $21,855 | $22,510 |
| **REVENUE** | $990,000 | $2,213,640.00 | $2,810,808 | $3,173,804 | $3,300,300 |
| **NET REVENUE** | $7,582 | $37,763 | $203,359 | $240,865 | $256,993 |

* Budget Assumptions on next page

** PhD support refers to pass through payments of $60,000 per year for social science PhD student seats in certain oversubscribed statistics and computers science Masters courses.

*Budget Assumptions

| | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
|---|---|---|---|---|---|
| Tuition Per Semester | $24,750 | $25,740 | $26,770 | $27,840 | $28,950 |
| Entering Class Size | 20 | 25 | 30 | 30 | 30 |
| Attrition Percentage | | 10 | 10 | 10 | 10 |
| Fringe | 26.5 | 26.4 | 26.6 | 26.8 | 27 |
| Merit Raise Percentage | 3 | 3 | 3 | 3 | 3 |
| Per Student Pass Through | $4,000 | $4,000 | $4,000 | $4,000 | $4,000 |
| Percent Aid in 1st-3rd Semesters | 20 | 20 | 20 | 20 | 20 |
| Percent Aid in 4th Semester | 75 | 75 | 75 | 75 | 75 |
| First year courses elsewhere | 2 | 2 | 2 | 2 | 2 |
| Second year courses elsewhere | | 5 | 5 | 5 | 5 |
| | | | | | |
| Stata licenses/1 per student | 1000 | 1250 | 1500 | 1500 | 1500 |
| $250/student for conference travel (reimbursement) | 5000 | 6250 | 7500 | 7500 | 7500 |
| Lunches/year | 4 | 4 | 4 | 4 | 4 |
| $1000/staff for conference registration & travel | $6,000 | $6,000 | $6,000 | $6,000 | $6,000 |
| N phones ($17.26/month) | 6 | 6 | 6 | 6 | 6 |

## Appendix B: Teaching Faculty CVs

CVs of faculty mentioned in the proposal are attached in alphabetic order:

Christopher Bail, Sociology
Paul Bendich, Math
Gale Boyd, SSRI and Economics
Robert Calderbank, Computer Science, Math and Electrical Engineering, iiD
Daniel Egger, Pratt
Ezra Miller, Math
Eric Monson, Duke Library
Thomas Nechyba, Economics, Public Policy, SSRI
Jerry Reiter, Statistics and iiD
Jana Schaich-Borg, SSRI and iiD
Mariano Tepper, SSRI and iiD
Tom Witelski, Math

# Christopher A. Bail

Duke University
Department of Sociology
254 Soc-Psych Hall
Durham, NC 27708

Phone: (919) 660-5643
Fax: (919) 660-5623
Email: christopher.bail@duke.edu
Homepage: http://chrisbail.net

## Employment

Assistant Professor of Sociology, Duke University, 2015-

Affiliate, Duke Network Analysis Center (DNAC), 2015-

Affiliate, Information Initative at Duke (iID), 2015-

Assistant Professor of Sociology, University of North Carolina at Chapel Hill, 2013-2015

Robert Wood Johnson Foundation Scholar, University of Michigan, 2011-2013

## Education

Ph.D. Sociology, Harvard University, 2003-2011

Visiting Scholar, London School of Economics and Political Science, 2008

Visiting Scholar, Fondation Nationale des Sciences Politiques (Paris), 2005

B.A. Government and French (Summa Cum Laude), Bowdoin College, 1998-2002

## Publications

### Book

Bail, Christopher A. 2015. *Terrified: How Anti-Muslim Fringe Organizations Became Mainstream.* Princeton University Press

   - *Outstanding Book Award, Association for Research on Non-Profit Organizations and Voluntary Action*

   - *Reviewed in Administrative Science Quarterly, Boston Review, Choice, Contemporary Sociology, Contexts, Salon, Sociology of Islam, and Political Science Quarterly*

   - *To be featured at an "Author Meets Critics" Panel at the Annual Meeting of the American Sociological Association in 2016.*

### Peer-Reviewed Journal Articles

Bail, Christopher A. Forthcoming, "Cultural Carrying Capacity: Organ Donation Advocacy, Discursive Framing, and Social Media Engagement." *Social Science and Medicine*

Bail, Christopher A. 2015. "Lost in a Random Forest: Using Big Data to Study Rare Events." *Big Data and Society* July: 1-3

Bail, Christopher A. 2015. "Taming Big Data: Using App Technology to Study Organizational Behavior on Social Media." *Sociological Methods and Research*

Bail, Christopher A. 2015. "The Public Life of Secrets: Deception, Disclosure, and Discursive Framing in the Policy Process." *Sociological Theory* 33(2): 97-124

Bail, Christopher A. 2014. "The Cultural Environment: Measuring Culture with Big Data." *Theory and Society* 43:3-4 465-482

Bail, Christopher A. 2012. "The Fringe Effect: Civil Society Organizations and the Evolution of Media Discourse about Islam since the September 11th Attacks." *American Sociological Review* 77:7 855-879

> *- Lead article*
> *-Distinguished Contribution to Scholarship Award, Political Sociology Section, American Sociological Association*
> *-Best Published Article Award, Communication and Information Technologies Section, American Sociological Association*
> *-Honorable Mention, Best Article Award, Society for the Scientific Study of Religion*

Bail, Christopher A. 2008. "The Configuration of Symbolic Boundaries against Immigrants in Europe." *American Sociological Review* 73:1 37-59

> *- Reprinted in Investigating the Social World: The Process and Practice of Research by Russell K. Schutt, New York: Sage*

Bail, Christopher A. (with Michele Lamont). 2005. "Sur les frontieres de la reconnaissance: Les categories internes et externes de l'identite collective (On the Boundaries of Recognition: Internal and External Categories of Collective Identity)." *Revue Europeene de Migrations Internationales* 21:2 61-90

### *Articles under Review*

Bail, Christopher A. "Emotional Feedback and the Viral Spread of Social Media Messages about Autism Spectrum Disorders" (Revise and Resubmit, *American Journal of Public Health*)

### *Articles in Progress*

Bail, Christopher A. and Taylor Whitten-Brown."Advocacy Organizations, Cultural Holes, and Public Conversation on Social Media."

Bail, Christopher A., Taylor Whitten-Brown, and Andreas Wimmer. "Using Google Trends Data to Map the Global Diffusion of Culture."

### *Book Reviews and Other Publications*

Bail, Christopher A. 2009. Review of "Europe Old and New: Transnationalism, Belonging, and Xenophobia." (by Ray Taras) *Contemporary Sociology*, 38:562

Bail, Christopher A. 2008. Review of "Two-Faced Racism: Whites in the Backstage and Frontstage." (by Leslie Houts Picca and Joe R. Feagin) *Ethnic and Racial Studies*, 31:6 1168-1183

Bail, Christopher A. (with Michele Lamont) 2007. "Bridging Boundaries: the Equalization Strategies of Stigmatized Ethno-racial Groups Compared." *Minda de Gunzburg Center for European Studies Working Paper Series*, 154

## Grants and Fellowships

National Science Foundation, "Civil Society Organizations, Social Media, and Public Attention" (Principal Investigator, $143,989)

Robert Wood Johnson Foundation, "Examining the role of culture in public understanding of Autism Spectrum Disorders," (Principal Investigator, $20,000)

Duke Bass Connections, "Scientific Social Network Analysis Project," (Co-Investigator, $25,000)

Amazon Education Research Grant, "Using Computational Methods to Study the Viral Spread of Social Media Messages," (Principal Investigator, $10,000)

Junior Faculty Development Grant, University of North Carolina, Chapel Hill ( $7,500)

Google Maps Engine Grant, Google 2013

Seed Grant, Odum Institute, University of North Carolina at Chapel Hill (with Kenneth Andrews, Neal Caren, Charles Kurzman, and Andrew Perrin), 2012-2013 ($13,000)

Dissertation Fellowship, Radcliffe Institute for Advanced Study (Declined), 2010 ($22,300)

Dissertation Improvement Grant, National Science Foundation, 2009 ($10,000)

Elliot Dissertation Fellowship for top dissertations in the social sciences at Harvard University,(Declined), 2009 ($25,000)

John L. Loeb Fellowship, Harvard Divinity School, 2009 ($10,000)

Krupp Fellowship, Minda de Gunzburg Center for European Studies, Harvard University (Declined), 2008 ($24,000)

Sinclair Kennedy Traveling Fellowship, Harvard University (Declined), 2008 ($22,000)

Merit Fellowship, Graduate School of Arts and Sciences, Harvard University (Declined), 2008 ($9,850)

Dissertation Research Fellowship, Center for American Political Studies, Harvard University, 2008 ($21,830)

Dissertation Fellowship, Horowitz Foundation for Social Policy, 2008 ($3,000)

Mid-Dissertation Fellowship, Weatherhead Center for International Affairs, Harvard University, 2008 ($3,000)

German Marshall Fund Fellowship, U.S. Council for European Studies, 2005 ($6,000)

European Network on Inequality Fellowship, Malcolm Weiner Center for Social Policy, Kennedy School of Government, Harvard University, 2005 ($6,500)

National Science Foundation Integrated Graduate Education and Research Training Dissertation Grant, 2004-2008 ($6,500)

## Distinctions

Max Weber Fellow, European University Institute (Declined), 2011

Finalist, Princeton Society of Fellows (Declined), 2011

Graduate Student Paper Prize, Society for the Study of Social Problems (for "The Public Life of Secrets: How Leaks Shaped the Evolution of Britain's Domestic Counter-Terrorism Policy"), 2010

Graduate Student Paper Prize, Society for the Scientific Study of Religion (for "The Fringe Effect: Civil Society Organizations and the Evolution of Media Discourse about Islam since the September 11th Attacks"), 2010

Honorable Mention, Suzanne Langer Prize, Sociology of Culture Section, American Sociological Association (for "The Configuration of Symbolic Boundaries against Immigrants in Europe"), 2010

Graduate Student Paper Prize, Political Sociology Section, American Sociological Association (for "The Configuration of Symbolic Boundaries against Immigrants in Europe"), 2009.

Finalist, Charlotte W. Necombe Fellowship, Woodrow Wilson Foundation, 2009

Aage B. Sorenson Award for Best Graduate Student Paper, Department of Sociology, Harvard University (for "The Configuration of Symbolic Boundaries against Immigrants in Europe") 2008

Distinction on Oral Examination, Department of Sociology, Harvard University, 2006

Certificate of Excellence in Teaching, Derek Bok Center for Teaching, Harvard University, 2006

Distinction on PhD Qualifying Paper, Department of Sociology, Harvard University, 2005

Certificate of Excellence in Teaching, Derek Bok Center for Teaching, Harvard University, 2005

Philo Sherman Bennet Award for Best Senior Thesis, Department of Government, Bowdoin College, 2002

Phi Beta Kappa, Bowdoin College, 2002

## Invited Lectures

Boston College, "Anti-Muslim Organizations and Social Media," 2015

Notre Dame University, "Measuring Cultural Emergence," 2015

European University Institute, "How Fringe Groups Shape the Integration of Muslims in the West," 2015

Princeton University, "Filling Cultural Holes: How Autism Advocacy Organizations Mobilize New Audiences on Social Media," 2015

City University of New York, "The Rhythm of Communicative Styles: Advocacy Organizations, Social Media, and Mass Persuasion" 2015

Duke University, "The Rhythm of Communicative Styles: Advocacy Organizations, Social Media, and Mass Persuasion" 2014

Ohio State University, "The Rhythm of Communicative Styles: Advocacy Organizations, Social Media, and Mass Persuasion" 2014

University of Michigan, "The Rhythm of Communicative Styles: Advocacy Organizations, Social Media, and Mass Persuasion" 2014

University of Chicago, "The Rhythm of Communicative Styles: Advocacy Organizations, Social Media, and Mass Persuasion" 2014

Brown University, "The Rhythm of Communicative Styles: Advocacy Organizations, Social Media, and Mass Persuasion" 2014

Princeton University, "Studying the Global Diffusion of Culture with Big Data," 2014

Facebook Data Science, "Winning Hearts and Minds: Advocacy Organizations, Field Rhythms, and Bystanders," 2014

University of Michigan (Blalock Lecture), "Measuring Culture with Big Data," 2014

University of California, Santa Barbara, "Measuring Culture with Big Data," 2014

North Carolina State University, "Big Data and Social Science" 2014

University of Michigan, Interdisciplinary Committee on Organizational Studies, "The Fringe Effect: Civil Society Organizations and the Evolution of Public Discourse about Islam," 2013

Notre Dame University, Department of Sociology, "The Fringe Effect: Civil Society Organizations and the Evolution of Public Discourse about Islam," 2013

University of British Columbia, Measuring Culture Conference, "Finding, Mapping, and Analyzing Discursive Fields using Big Data," 2013

University of North Carolina, Chapel Hill, Department of Sociology "Terrified: Emotions, Social Movements, and Islam, 2001-2008," 2010

Princeton University, Department of Sociology, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," 2010

University of Michigan, Department of Sociology, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," 2010

Duke University, Department of Sociology, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," 2010

Stanford University, Department of Sociology, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," 2010

University of California, Los Angeles, Department of Sociology, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," 2010

Vanderbilt University, Department of Sociology, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," 2010

New York University, Department of Sociology, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," 2010

Boston College, Department of Sociology, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," 2010

University of Georgia, Department of Sociology, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," 2010

Emory University, Department of Sociology, "Terrified: Emotions, Social Movements, and Islam, 2001-2008," 2010

Yale University, Department of Sociology "Making Terrorists Racist: The Culture of Secrecy in Britain's Domestic Counter-Terrorism Policy, 20001-2008," 2009

University of Georgia, Department of Sociology, "Making Terrorists Racist: The Boundaries of Multicultural and Counter-Terrorism Policy in Britain, 2001-2008," 2009

Harvard University, Weatherhead Center for International Affairs, "Inter-Ethnic Relations in Quebec: A Roundtable with Charles Taylor and Gerard Bouchard," 2009

Valencia, Spain, City-to-City Conference of the European Union, "Diverse Diversities: The Configuration of Symbolic Boundaries against Immigrants in Twenty-One European Countries," 2007

Yale University, Department of Sociology, "Bridging Boundaries: The Everyday De-stigmatization Strategies of Marginalized Groups Compared" (with Michele Lamont), 2005

## Conference Presentations

American Sociological Association Annual Meeting, San Francisco, CA "NSF Sponsored Panel on Big Data and Social Science Research," 2014.

American Sociological Association Annual Meeting, New York, NY "Winning Minds Through Hearts: Civil Society Organizations, Emotional Feedback, and Social Media," 2013.

American Sociological Association Annual Meeting, Denver, CO "The Fringe Effect: Civil Society Organizations and the Evolution of Media Discourse about Islam, 2001-2008," 2012.

American Sociological Association Annual Meeting, San Francisco, CA "How Terrorists became Racists," 2009

Harvard/Manchester Conference on Social Change, Manchester, England "Homeland Insecurity: Structural Holes in the Muslim Civil Rights Movement," 2008

American Sociological Association Annual Meeting, New York, NY, "The Configuration of Symbolic Boundaries toward Immigrants in Twenty-One European Countries," 2007

11th Annual Aage Sorensen Memorial Conference, Oxford University, England, "The New Black Brahmins? Elite African Immigrants and the Stigma of Africanite" 2007

National Association for Ethnic Studies Annual Meeting, New Paltz, NY "Diverse Diversities: The Configuration of Symbolic Boundaries against Immigrants in Twenty-One European Countries," 2007

Eastern Sociological Society, Philadelphia, PA."Diverse Diversities: The Configuration of Symbolic Boundaries against Immigrants in Twenty-One European Countries," 2007

American Sociological Association Annual Meeting Minorities, Montreal, Quebec "Three Worlds of Xenophobia: a Typology of Anti-Immigrant Boundaries in Europe," 2006

International Conference on Comparative Social Sciences, Sophia University, Tokyo, Japan "Three Worlds of Xenophobia: a Typology of Anti-Immigrant Boundaries in Europe," 2006

10th Annual Aage Sorensen Memorial Conference, Stockholm University, Sweden, "Three Worlds of Xenophobia: Towards a Typology of Anti-Immigrant Prejudice in Europe," 2006

Council for European Studies Annual Meeting, Section on Ethnic Boundaries and Regionalism, Chicago, IL "Three Worlds of Xenophobia: Mapping the Configuration of Ethnic Boundaries in Twenty European Countries," 2006

Multidisciplinary Program on Inequality and Social Policy, John F. Kennedy School of Government, Harvard

University, Cambridge, MA "Toward a Typology of Ethnic Boundaries: A Cognitive Analysis of Racial, Religious, Linguistic, and Cultural Boundaries in Twenty-One European Countries," 2005

Summer Institute on International Migration, Ethnic Diversity, and Cities, International School for the Humanities and Social Sciences, Universiteit van Amsterdam, Netherlands, "On the Ethnic Boundaries of Immigration: Economic, Cultural, and Racial Parameters of Xenophobia," 2005

9th Annual Aage Sorensen Memorial Conference: Cambridge, MA, "Black or African American? The Socio-Economic Attainment of Africans in the United States," 2005

# Teaching and Advising

Graduate Seminar on Computational Sociology, Duke University, 2015

Introduction to Sociology (Sociological Inquiry), Duke University, 2015

Graduate Seminar on Cultural Sociology, University of North Carolina Chapel Hill, 2015

Introduction to Sociology (Sociological Perspectives), University of North Carolina Chapel Hill, 2015

Introduction to Sociology (Sociological Perspectives), University of North Carolina Chapel Hill, 2013

Social Studies 10, Harvard Universtiy, 2007

Sociological Theory, Harvard University, 2006

PhD Thesis Committee Member for M. Ali Kadivar (UNC-Chapel Hill), Taylor Whitten-Brown (UNC-Chapel HIll), Haj Yazdiha (UNC-Chapel Hill), Brandon Gorman (UNC-Chapel Hill), and Charles Seguin (UNC-Chapel Hill).

Master's Thesis Committe Member for Haj Yazdiha (UNC-Chapel Hill), Brandon Gorman (UNC-Chapel Hill), and David Rigby (UNC-Chapel Hill)

# Professional Service

Book Review Editor, Social Forces, 2014-2015

Editorial Board, Sociological Theory, 2010-

Editorial Board, Social Forces, 2011-

Editorial Board, American Journal of Cultural Sociology, 2011-

Council Member, Section on Communication and Information Technologies, American Sociological Association, 2013-2014

Nominations Committee, Sections on Organizations, Occupations, and Work, American Sociological Assocation, 2015-2016.

Program Committee, International Conference on Computational Social Science, Helsinki Finland, 2015

Student Paper Award Committee, Communication and Information Technology Section of the American Sociological Association, 2015

Nominations Committee, Section on Political Sociology, American Sociological Association, 2013-2014

Nominations Committee, Section on the Sociology of Emotions, American Sociological Association, 2013-2014

Occassional Reviewer, Proceedings of the National Academy of Science, American Sociological Review, American Journal of Sociology, Oxford University Press, Social Forces, Sociological Theory, Ethnic and Racial Studies, Sociological Methods and Research, Social Problems, American Journal of Cultural Sociology, Sociological Forum, 2007-

Organizer, International Conference on Computational Social Sciencey (Helsinki, 2015)

Organizer, Computational Sociology Summit (Chicago, 2015)

Organizer, Computational Social Science and the Study of Social Behavior, Social Psychology Section, American Sociological Association Annual Meetings, 2014.

Presider, Section on Cultural Sociology, American Sociological Association Annual Meetings, 2013.

Presider, Section on Ethnic and Racial Minorities Roundtable, American Sociological Association Annual Meetings, 2006.

# Departmental Service

Executive Committee, Duke University, 2015-

Colloquium Series Organizer, Duke University, 2015-

Publications Committee, UNC-Chapel Hill, 2014-

Graduate Student Placement Coordinator, UNC-Chapel Hill, 2013-2014.

Committee on Undergraduate Studies, UNC-Chapel Hill, 2013-2014

# Languages

French, Mandarin Chinese (conversational), R, Python, Stata, LaTeX

Last updated: January 19, 2016

# Paul Louis Bendich

Curriculum Vitae

| Department of Mathematics | bendich@math.duke.edu |
| Physics Building 210 | www.paulbendich.com |
| Duke University | *Phone*: (919) 660-2811 |
| Durham, NC 27708-0320 | *Fax*: (919) 660-2821 |

**Research Interests**: Computational Topology and Geometry, Algebraic Topology, Stratified Spaces, Machine Learning and Statistics

## Education

**Duke University**                                   Durham, NC
*Ph.D. in Mathematics*                               8/2003–8/2008

**Duke University**                                   Durham, NC
*M.A. in Mathematics*                                8/2003–2/2005

**Grinnell College**                                  Grinnell, IA
*B.A. in Physics*                                    8/1997–5/2001

## Employment

- Assistant Research Professor (regular-rank), Department of Mathematics, Duke University, 4/2014- present

- Associate Director for Curricular Engagment, the Information Initiative at Duke (iiD), 7/2014- present

- Senior Mathematician, Geometric Data Analytics, 7/2014- present

- Visiting Assistant Professor, Department of Mathematics, Duke University, 1/2011- 3/2014

- Postdoctoral Associate, IST Austria, 8/2009–12/2010.

- Instructor, Department of Mathematics, Pennsylvania State University, 8/2008–7/2009.

## Publications

# 1 Journal Publications

- *Topological and Statistical Behavior Classifiers for Tracking Applications*, Paul Bendich, Sang Chin, Jesse Clarke, John Harer, Elizabeth Munch, David Porter, David Rouse, Nate Strawn, and Adam Watkins. IEEE Trans. on Aerospace and Electronic Systems, to appear.

- *Persistent Homology Analysis of Brain Artery Trees*, Paul Bendich, J.S. Marron, Ezra Miller, Sean Skwerer, and Alex Pieloch. Annals of Applied Statistics, to appear.

- *Probabilistic Frechet Means and Statistics on Vineyards*, Liz Munch, Kate Turner, Paul Bendich, Sayan Mukherjee, Jonathan Mattingly, and John Harer. Electronic Journal of Statistics, Volume 9, pp. 1173-1204.

- *Homology and Robustness of Level and Interlevel Sets*, Paul Bendich, Herbert Edelsbrunner, Dmitriy Morozov, and Amit Patel. Homology, Homotopy and Applications, Vol. 15 (2013), No. 1, pp.51-72.

- *A Point Calculus for Interlevel set Homology*, Paul Bendich, Sergio Cabello and Herbert Edelsbrunner. Pattern Recognition Letters (2012), 1436-1444.

- *Improving Homology Estimates with Random Walks*, Paul Bendich, Taras Galkovskyi and John Harer. Inverse Problems 27 (2011) 124002.

- *Persistent Intersection Homology*, Paul Bendich and John Harer. Foundations of Computational Mathematics, 11 (2011), no. 3, 305-336.

- *Computing Robustness and Persistence for Images*, Paul Bendich, Herbert Edelsbrunner, and Michael Kerber. IEEE Trans. Visual. and Comput. Graphics, 2010, pp. 1251-1260.

# 2 Conference Proceedings Publications

- *Cover Song Identification with Timbral Shape Sequences*, Christopher J. Tralie and Paul Bendich. Proceedings of the 2015 International Symposium on Music Information Retrieval, to appear.

- *Multi-scale Local Shape Analysis and Feature Selection for Machine Learning Applications*, with Ellen Gasparovic, John Harer, Rauf Ismailov, and Linda Ness. Proceedings of the 2015 International Joint Conference on Neural Networks, to appear.

- *Local Homology Transfer and Stratification Learning* with Bei Wang and Sayan Mukherjee. Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms Pages 1355-1370, 2012.

- *Persistent Homology under Non-Uniform Error* with Herbert Edelsbrunner, Michael Kerber, and Amit Patel. Proc. 35th Internat. Sympos. on Math. Found, of Comput, Science, 2010, pp. 12-23.

- *The Robustness of Level Sets* with Herbert Edelsbrunner, Dmitriy Morozov, and Amit Patel. Proc. 18th Europ. Sympos. Algorithms, 2010, pp. 1-10.

- *Inferring Local Homology from Sampled Stratified Spaces* with David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Dmitriy Morozov. Proc. 48th Sympos. on Found. of Comput. Science, 2007, pp. 536-546.

# 3 Dissertation

- *Analyzing Stratified Spaces Using Persistent Versions of Intersection and Local Homology*, Ph.D. Thesis, Duke University,2008.

## Undergraduates Mentored

- *Topology, Statistics, and Brain Data*, Data RTG, Summer 2014:
    - Carmen Cox (Duke)
    - Derrick Nowak (Duke)
    - Henry Farrell (Cornell)
    - Dong-Hwan Moon (Williams)
    - Alex Pieloch (Duke)

- *Multi-scale Topology for Signals and Images*, Data RTG, Summer 2013:
    - Bingxi Lin (Bryn Mawr)
    - Michael Ogez (Duke)
    - Benjamin Dreyzen (UNC)

– Joshua Martin (UNC-Greensboro)

- Marshall Ratliff, Duke PRUV 2015

    – Senior Thesis: *Introducing the Cover Tree to Music Information Retrieval*

- Bryan Jacobsen, Duke PRUV 2012

    – Senior Thesis: *A Fast Approximate Algorithm for Local Homology*

## Courses Taught

- *Combinatorics*, Duke Fall 2015.

- *Linear Algebra*, Duke, Fall 2013 and Spring 2014.

- *The Emerging Science of Complex Data (First-Year Seminar)*, Duke, Spring 2012 and Spring 2013.

- *Topology with Applications*, Duke, Fall 2012 and Fall 2014.

- *Topology*, Duke, Fall 2011.

- *Computational Topology*, IST Austria, Fall 2010.

- *Linear Algebra*, Penn State, Spring 2009.

- *Business Calculus II*, Penn State, Spring 2009.

- *Calculus I*, Penn State, Fall 2008.

- *Linear Algebra and Differential Equations*, Duke, Summers 2008 and 2007.

- *Laboratory Calculus II*, Duke, Summer 2006, Spring 2006.

- *Laboratory Calculus I*, Duke Fall 2005.

## Departmental and University Service

- Director, Data+ Program, The Information Initiative at Duke (iiD), 7/2014–Present.

- Director, Data Expeditions Program, The Information Initiative at Duke (iiD), 7/2014–Present.

- Coordinator, Research Training Grant (Structure in Complex Data), Duke, 1/2011–Present.

- Coordinator, Summer Undergraduate Research Program, Duke, 1/2011–Present.

- Organizer, Data Seminar, Duke, 8/2011–Present.

- Member and Founder, Graduate Student Calculus Curriculum Committee, Duke, Fall 2008.

## Workshops Organized

- Spring Topology and Dynamics Conference, Session on Applied Topology, University of Richmond, March 2014.

- LDHD: Topological Data Analysis, workshop at SAMSI, Feburary 2014.

- Computational Topology, workshop at Symposium on Computational Geometry, Chapel Hill, NC, June 2012.

- Computational Topology, workshop at SIAM Conference on Applied Algebraic Geometry, Raleigh, NC, October 2011.

## Seminar and Conference Talks

- *Persistent Homology Analysis of Brain Artery Trees*, DCAT, Copenhagen, 11/2014.

- *Topological Features for Machine Learning*, team-talk with Nate Strawn, Data Seminar, Duke University, 4/2014.

- *Persistent Local Homology: Theory, Applications, Computational Innovations*, Workshop on Topology and Statistics, SAMSI, 2/2014.

- *Persistent Homology: theory and computational innovations*, Algorithm Theory Seminar, North Carolina State University, 11/2013.

- *Stratifications and Persistent Homology*, SATANA Seminar, University of Illinois at Urbana-Champaign, 11/2013.

- *Brain-artery Trees and Persistent Homology*, iiD Seminar, Duke, 11/2013.

- *Probabilistic Frechet Means and Statistics on Vineyards*, Workshop on Applied Algebraic Topology, Bremen, 8/2013.

- *Stratification Learning via Local Homology Inference*, AMS Sectional, Boulder, CO, 4/2013.

- *Tracking with Persistence*, BANFF, 10/2012.

- Φ-*SoMap*, AMS-MAA Joint Meetings, Boston, MA, 1/2012.

- *Persistence Diagrams and the Information they Carry*, Data Seminar, Duke, 8/2011.

- *Stratification Learning via Local Homology Inference*, Invited Talk, INRIA-Saclay, 10/2010.

- *The 2-Point Formula*, CTIC, 10/2009.

- *Elevation on Stratified Spaces via Intersection Homology*, DARPA TDA meeting, 1/2009.

- *Teaching Without much Lecturing*, Education Seminar, Penn State, 1/2009.

- *Persistent Intersection Homology*, Algorithms Seminar, Duke, 3/2008.

- *Local Homology Vineyards*, DARPA TDA meeting, 1/2008.

- *Persistence*, Grad-Fac Seminar, Duke, 10/2007.

- *Persistent Local Homology*, DARPA TDA meeting, 5/2006.

## Professional Service

- Reviewer for Journal of Topology and Analysis

- Reviewer for SIAM Journal of Computing

- Reviewer for Symposium on Computational Geometry

- Reviewer for Symposium on Artificial Intelligence and Statistics

- Reviewer for Symposium on Discrete Algorithms

- Reviewer for Experimental Mathematics

- Reviewer for Foundations of Computational Mathematics

- Reviewer for Inverse Problems

- Reviewer for Discrete and Computational Geometry

- Reviewer for Revista Matematica Complutense

Duke University
DURHAM
NORTH CAROLINA
27708-0097

Social Science Research Institute (SSRI), 261 Gross Hall          Telephone (919) 660-1800
Box 97708                                                          Fax (919) 681-7984

# GALE A. BOYD

## Professional Experience:

| | |
|---|---|
| 2014 - Present | Assoc. Research Professor – SSRI & Dept. of Economics, Duke University |
| 2006 – Present | Director – Triangle Research Data Center (TRDC) |
| 2006 – 2014 | Senior Research Scholar – Department of Economics, Duke University |

Administration: Management and promotion of the TRDC, a social sciences research user facility operated in collaboration with U.S. Census Bureau dedicated to providing approved researchers access to non-public data for statistical purposes.   Established the Duke University / University of North Carolina / Research Triangle Institute partnership for the TRDC.   Assisting researchers with the proposal development process and representing the TRDC institutional partners with the Bureau.

Research:  Recent studies include using the non-public Census micro-data and other non-pubic data from industry and trade associations on energy, environmental, and productivity related issues for industrial energy efficiency and related energy/environmental policy research. Recent research includes preparing statistical benchmarks of energy performance in manufacturing plants, or *Energy Performance Indicators* (EPI), is supported by the EPA ENERGY STAR program and is used by industry for energy management and public recognition from ENERGY STAR.  Studies of the implications and causes of industry energy and total factor productivity distributions on management practices and environmental policy are also in progress.

Teaching:

| | |
|---|---|
| Econ 325s | - Economic Analysis of Energy Issues |
| Econ 431s | - Energy Markets and Environmental Impacts |
| MgrEcon 491 | - Environmental Economics |
| Energy 525 | - Bass Connections in Energy: Energy Efficiency in Industry |
| Econ 690 | - Selected Topics in Economics: Social and Economic Data |

| | |
|---|---|
| 2005 –2006 | Group Leader – Social Science, Policy, and Law (SSPL) |
| | Decision and Information Sciences Division (DIS) |
| | Argonne National Laboratory (ANL) |

Management of the SSPL section of DIS, encompassing 7 full time and 11 part time research staff, with disciplines ranging from law, emergency management, sociology, to economics.

| | |
|---|---|
| 2002 –2006 | Chairman - Chicago Regional Census Data Research Center |

The Chicago Census Research Data Center (RDC) is a social sciences research user facility operated in collaboration with The University of Chicago, Northwestern University, University of Illinois-Chicago, and Chicago Federal Reserve Bank. The board is responsible for coordination of funding for the RDC and project review. The chairman represents the RDC regarding funding and resources from the five member institutions, oversees the RDC executive director and Census administrator, writes external grant proposals, and promotes the expansion of research using the micro data.

2002 – 2004                    Senior Economist
                               BKE Associates, Santa Fe NM

Statistical models of baseline manufacturing plant energy use to provide normalized forecasts for third party verification for energy performance contracting between DaimlerChrysler and Detroit Energy Services.

1992 – 2001                    Senior Economist
                               Applied Economic Research, Inc., New York

Advised the president of the firm on econometrics and production function approaches to utility power plant modeling and risk assessment. Developed frontier models to benchmark plant-level and unit-level maintenance spending targets and outage costs and risks. Developed a risk assessment approach for PricewaterhouseCoopers that uses internal partner survey data and ordered probit modeling. Provided estimates for a bankruptcy prediction model for the life and health insurance sector. Developed parametric and nonparametric frontier production function models of the expenditures and reliability associated with maintaining fossil fuel power plants for private sector utility clients. Adapted a model of power plant dispatch to assess the impact of $NO_x$ regulatory compliance timetables on system reliability for the Utility Air Research Group.

1984 –2006                     Energy/Environmental Economist - ANL

Primary activities include microeconomic modeling of industrial energy demand, emissions, and productivity; forecasting activity and energy demand in various economic sectors; and integrating the above analyses into models of the energy market. This work has generated more than 100 journal articles, book chapters, conference papers, and technical reports, as well as numerous oral presentations. Since 1995, I have been a principal investigator for the U.S. Department of Energy (DOE) Office of Policy, analyzing industrial voluntary action programs, and for the U.S. Environmental Protection Agency (EPA) Office of Atmospheric Programs, conducting research on policy actions to reduce greenhouse gas emissions.

     *Conduct research on industrial productivity, energy, and emissions.* Principal investigator for ANL's industrial research program at the Center for Economic Studies, U.S. Bureau of the Census, directing research on proprietary plant-level data. Principal investigator for the EPA in researching the relationships among industrial productivity, energy, and emissions in support of various policy studies and in developing a series of industry-specific energy performance indicators for ENERGY STAR™. Primary author of chapters in five major DOE reports: (1) *Scenarios for a Clean Energy Future,* (2) *Scenarios of U.S. Carbon Reductions: Potential Impacts of Energy Technologies by 2010 and Beyond*, (3) *Technical Guidelines for Implementing Section 1605(B) of EPAct*, (4) *Limiting Net Greenhouse Gas Emissions in the United States,* and (5) *The Interrelationship between Environmental Goals, Productivity Improvement, and Increased Energy Efficiency in Integrated Paper and Steel Plants*. First proposed and developed the now widely used *Divisia* Index as a method for decomposing aggregate energy intensity (the journal article was mentioned as the second most highly cited in a 10-year anniversary retrospective by *Energy Economics*). Was part of a team to develop a system of energy efficiency indicators for DOE.

     *Develop forecasts and models for policy assessments.* Coordinate energy/economic modeling support between ANL and a University of Chicago lead team studying nuclear power. Participated on the team that developed the Electric Market Complex Adaptive System (EMCAS) model and conducted

policy analyses for the Illinois Commerce Commission. Helped organize the Governor of Illinois's Energy Market Workshop hosted by ANL. Updated ANL's Long-Term Industrial Energy Forecasting model to a new base year. Conduct scenario analyses of greenhouse gas reductions for the EPA. Provided guidance on modeling sector energy and emission trends for the (1) Interlaboratory Working Group on Energy Efficient and Low-Carbon Technologies, (2) National Acid Precipitation Assessment Program (NAPAP), (3) 1990 National Energy Strategy, (4) EPA Economic Growth Assessment (EGAS) model, and (5) Grand Canyon Visibility Commission Integrated Assessment System. Developed the Industrial Regional Activity/Energy Demand model and its emission forecasting components to support global climate analysis, NAPAP, and the EGAS model.

*Integrated technology evaluation and emission modeling.* Work with the EPA ENERGY STAR team of investigators to coordinate energy technology opportunity assessment reports with results obtained from the EPI analysis. Integrate results from the activities described above into energy market analyses for several projects sponsored by DOE, EPA, and others. Contributed to every phase of development of the NAPAP Emissions and Cost Integrated Model Set. Provided technical support to the Environmental Working Groups for the 1985, 1987, and 1990 National Energy Strategies. Led development of the National Energy Strategy Environmental Analysis Model. Led development of the Regional Emission*s* Database and Evaluation System to support the programmatic environmental impact statement (EIS) for the DOE Clean Coal Program. Evaluated industrial energy technologies — particularly those in the steel, cement, wet corn refining, automobile assembly, brewing, and paper industries — with regard to their energy efficiency and pollutant emissions.

| 1983 – 1984 | Researcher II |
| | Coal Technology Lab (CTL) and Department of Economics |
| | Southern Illinois University, Carbondale |

**Education:**

| Ph.D. | Economics, Southern Illinois University, Carbondale, 1984 |
| B.A. | Mathematics, Southern Illinois University, Carbondale, 1980 |

**Thesis** *Factor Intensity and Site Geology as Determinants of Returns to Scale in Coal Mining*
Applied Micro / Energy Economics          Rolf Färe - Advisor

**Reviewer Activities:**

Member Editorial Board – *Energy Efficiency*
Reviewer – *The Energy Journal, Energy Policy, Energy Economics, Journal of Economic Literature, Journal of Cleaner Production, Journal of Environmental Economics and Management, Energy, Energies, Annals of Operations Research, Resources and Energy Economics*

**Awards and Honors:**

CO-PI  NSF Grant Award HSD-0433990, Argonne National Lab/Duke University (2005 – 2009)
CO-PI  NSF Grant Award ITR-0427889, Chicago Census Data Research Center (2004 – 2008)
CO-PI  NSF Grant Award 0004335, Chicago Census Data Research Center (2001 – 2003)
ANL/University of Chicago Collaborative Grant Award (1997, 1998, 2005)
ANL Pacesetter Award (1986)
Doctoral Fellowship, Southern Illinois University - Carbondale, 1983
Masters Fellowship, Southern Illinois University - Carbondale, 1982

**Funding History**

| Project Title | Year | Thousand $ | Agency | Duration |
|---|---|---|---|---|
| Industrial energy and productivity benchmarks: Phase IV | 2014-2016 (projected) | $816 | EPA | 3 years |
| Auto Industry Energy and Pollution Benchmarking | 2012-2014 | $130 | Toyota, Nissan, & GM | 3 years |
| Industrial energy and productivity benchmarks: Phase IV | 2012-2013 | $544 | EPA | 2 years |
| Industrial energy and productivity benchmarks: Phase III | 2006-2011 | $1,712 | EPA | 5 years |
| Information Technology Challenges for Secure Access to Confidential Social Science Data | 2004-2007 | $460 | NSF | 3 years |
| Hedonic Models of Location Decisions with Applications to Geospatial Micro-data | 2004-2007 | $208 | NSF | 3 years |
| Industrial energy and productivity benchmarks: Phase II | 2002-2006 | $1,245 | EPA | 4 years |
| Industrial Actions to Reduce CO2 | 1995-2001 | $1,925 | EPA | 7 year |
| Industrial energy and productivity benchmarks: Phase I | 2001- 2003 | $320 | EPA | 2 years |
| Baseline issues for the Clean Development Mechanism | 2000 | $40 | DOE | 1 year |
| Impact of Early Action Carbon Credits on Voluntary Programs | 1999 | $50 | DOE | 1 year |
| Joint Modeling of Environmental Performance | 2000-1998 | $183 | DOE | 2 years |
| Regional Impacts of the 5-Lab Study | 1998 | $30 | EPA | 1 year |
| Interlab Study on Energy Efficiency and Carbon Mitigation: Industrial | 1997 | $50 | DOE | 1 year |
| Voluntary Agreements & Emission Trading | 1997 | $50 | DOE | 1 year |
| Glass Industry Plant Energy Benchmarking | 1996 | $25 | DOE | 1 year |
| MECS  Analysis at  Census | 1995 | $100 | DOE | 1 year |
| New Abatement Cost Estimates | 1995 | $30 | DOE | 1 year |
| Regional Industrial Activity Forecasting for Ozone Modeling | 1993 | $90 | EPA | 1 year |
| Evaluation of Global Climate Policy and Measures | 1992 | $50 | DOE | 1 year |
| Evaluation of Global Climate Policy and Measures | 1991 | $100 | DOE | 1 year |
| Plant Level Industrial Energy Analysis | 1989-1992 | $250 | DOE | 3 years |
| **Total** | | **$8,408** | | |

## Publications

**Recent working papers and papers under review:**

Gale Boyd, Jay Golden and Charles McClure, "Waste, Water, and VOC Benchmarks for U.S. Automobile Manufacturing," (July 2015), submitted to the *Journal of Cleaner Production*

Gale Boyd, *"Comparing the Statistical Distributions of Energy Efficiency in U.S. Manufacturing: Meta - Analysis of 24 Industry Case Studies"* Nicholas Institute for Environmental Policy Solutions: Duke Environmental Economics Working Paper Series: EE 15-03 (July 2015) Revise and resubmit to *Energy Efficiency*

Gale Boyd and Jay S. Golden, "Enhancing Energy Efficiency Reporting: Using Index Numbers to Report Corporate-Level Measures of Sustainability", Nicholas Institute for Environmental Policy Solutions: Duke Environmental Economics Working Paper Series: EE 15-04 (July 2015) submitted to the *Journal of Industrial Ecology*

Gale Boyd and Chung-Ying Lee, "Automotive Alliance: Motor Vehicle Assembly Plant Extended Benchmarking" working paper, May 16th, 2014

Gale Boyd, Tatyana Kuzmenko, Béla Személy, & Gang Zhang, "*Preliminary Analysis of the Distributions of Carbon and Energy Intensity for 27 Energy Intensive Trade Exposed Industrial Sectors,*" Nicholas Institute for Environmental Policy Solutions: Duke Environmental Economics Working Paper Series: EE 11-03 (April 2011)

**Publications: Journal Articles and Book Contributions**

Boyd, G. A. and E. M. Curtis (2014). "Evidence of an 'Energy-Management Gap' in U.S. manufacturing: Spillovers from firm management practices to energy efficiency." *Journal of Environmental Economics and Management* 68(3): 463-479.

Gale A. Boyd and Yifang Guo, "An Energy Performance Indicator for integrated paper and paperboard mills: A new statistical model helps mills set energy efficiency targets," *Paper360*, March/April 2014, pages 26-28

Gale Boyd, "Estimating the Changes in the Distribution of Energy Efficiency in the U.S. Automobile Assembly Industry," *Energy Economics*, Volume 42, March 2014, Pages 81–87

Boyd, G. and G. Zhang (2012). "Measuring improvement in energy efficiency of the US cement industry with the ENERGY STAR Energy Performance Indicator." *Energy Efficiency*: 1-12.

Elizabeth Dutrow, Gale Boyd, Ernst Worrell, Lew Dodendorf, "Engaging the U.S. Cement Industry to Improve Energy Performance," *Cement World,* (December 2009)

Gale Boyd, "Benchmarking Weather Sensitive Plant Level Manufacturing Energy Use" *Int. J. Global Energy Issues*, Vol. 32, Nos. 1/2, 2009, pp 4-18

Gale A. Boyd, invited book review "Emergence and Survival of New Businesses," *Journal of Economic Literature,* September 2008

Gale A. Boyd, "Estimating Plant Level Manufacturing Energy Efficiency with Stochastic Frontier Regression", *The Energy Journal*, Vol 29, No. 2, pp 23-44, (2008)

Boyd, G., E. Dutrow and W. Tunnesen, "The Evolution of the ENERGY STAR Industrial Energy Performance Indicator for Benchmarking Plant Level Manufacturing Energy Use." *Journal of Cleaner Production,* Invited paper for the special issue *Pollution Prevention and Cleaner Production in the United States of America* Volume 16, Issue 6, pp 709-715, April 2008

Gale Boyd, coeditor with Lorna Greening and Joe Roop, "Modeling Industrial Energy Consumption," Special Issue of *Energy Economics* Vol. 29, Issue 4, July 2007

Boyd, G.A., "A Statistical Model for Measuring the Efficiency Gap between Average and Best Practice Energy Use: The ENERGY STAR™ Industrial Energy Performance Indicator," *Journal of Industrial Ecology*, Vol. 9 (3): pp 51-56, (2005)

Worrell, E., S. Ramesohl, and G. Boyd, "Advances in Energy Forecasting Models based on Engineering-Economics" *Annual Review of Energy and Resources,* Vol. 29: 345-382, (2004)

Boyd, G.A., and J.M. Roop, "A Note on the Fisher Ideal Index Decomposition for Structural Change in Energy Intensity," *The Energy Journal* 25(1):87-101 (2004)

Boyd, G., G. Tolley, and J. Pang, "Plant Level Productivity, Efficiency, and Environmental Performance: An Example from the Glass Industry," *Environmental and Resource Economics*. 23(1):29-43 (Sept. 2002).

Boyd, G.A., J.C. Molburg, and J.D. Cavallo, "Estimates of Learning by Doing in Gas Turbine Electric Power Systems," *Energy Studies Review* 10(2):85-99 (2002).

Sanstad, A.H., S.J. DeCanio, G.A. Boyd, and J. Koomey, "Estimating Bounds on the Economy-Wide Effects of the CEF Policy Scenarios," *Energy Policy* 29:1299-1311 (2001).

Boyd, G.A., and J.A. Laitner, "Recent Trends in the U.S. Energy Intensity: An Index Number Analysis," *International Association for Energy Economics Newsletter,* pp. 4-9 (2nd quarter, 2001).

Boyd, G., and J. Pang, "Estimating the Linkage between Energy Efficiency and Productivity," *Energy Policy* 28:289-96 (2000).

Storey, M., G. Boyd, and J. Dowd, "Voluntary Agreements with Industry," pp. 209-228 in *Voluntary Approaches in Environmental Policy* (C. Carraro and F. Lévêque, editors), Kluwer Academic Publishers, Dordrecht, Netherlands (1999).

Boyd, G., and J. McClelland, "The Impact of Environmental Constraints on Productivity Improvement and Energy Efficiency in Integrated Paper Plants," *The Journal of Economics and Environmental Management* 38:121-146 (1999).

Storey, M., G. Boyd, and J. Dowd, "Voluntary Agreements with Industry," in *The Economics and Law of Voluntary Approaches*, special issue of *Fondazione Eri Enrico Mathei,* Series 26.97 (Feb. 1997).

Bock, M., G. Boyd, S. Karlson, and M. Ross, "Best Practice Electricity Use in Steel Minimills," *Iron and Steel Maker*, pp. 63-67 (May 1994).

Boyd, G., S. Karlson, M. Neifer, and M. Ross, "Energy Intensity Improvements in Steel Minimills," *Contemporary Policy Issues* XI (3):88-99 (July 1993).

Boyd, G., and S. Karlson, "The Impact of Energy Prices on Technology Choice in the United States Steel Industry," *Energy Journal* 14(2):47-56 (1993).

Boyd, G., "Models for Forecasting Industrial Energy Use," pp. 11-30 in *International Energy Modeling* (T. Sterner, editor), Chapman Hall, London, England (1992).

Kononov, Y.D., H.G. Huntington, E.A. Meduedeua, and G.A. Boyd, "The Effects of Changes in the Economic Structure on Energy Demand in the USSR and the US," pp. 47-64 in *International Energy Modelling* (T. Sterner, editor), Chapman and Hall, London, England (1992).

Streets D., C. Bloyd, G. Boyd, D. Santini, and T. Veselka, "Climate Change and U.S. Energy Policy," *Energy — The International Journal* 16(11-12):1437-1466 (1991).

Boyd, G., J. Fox, and D. Hanson, "Sets of Models," *Energy — The International Journal* 15(3-4) (Mar.-Apr. 1990).

Boyd, G., and M. Ross, "Structural Change in Manufacturing and Forecasting the Use of Electricity in the U.S. and Sweden," pp. 461-488 in *Efficient End Use and New Generation Technology and Their Implication for Electric Utility Planning and Policy*, Lund University Press (1989).

Ross, M., and G. Boyd, review of the book *The Changing Structure of American Industry and Energy Use Patterns,* invited by *Material Resources Forum* (1988).

Boyd, G.A., D.A. Hanson, and T. Sterner, "Decomposition of Changes in Energy Intensity: A Comparison of the Divisia Index and Other Methods," pp. 309-312 in *Energy Economics* (Oct. 1988).

Boyd, G., J.F. McDonald, M. Ross, and D.A. Hanson, "Separating the Changing Composition of U.S. Manufacturing Production from Energy Efficiency Improvements: A Divisia Index Approach," *The Energy Journal* 8(2):77-96 (1987).

Boyd, G., "Factor Intensity and Site Geology as Determinants of Returns to Scale in Coal Mining," *The Review of Economics and Statistics* 69(1):18-23 (Feb. 1987).

Boyd, G., and R. Färe, "Measuring the Efficiency of Decision Making Units: A Comment," *European Journal of Operations Research* 15:331-332 (1984).

**Publications: Conference Papers** *(underline denotes presenter)*

<u>Gale Boyd</u> and Walt Tunnessen, Plant Energy Benchmarking: A Ten Year Retrospective of the ENERGY STAR Energy Performance Indicators (ES-EPI) *Proc. ACEEE Summer Study on Energy Efficiency in Industry,* July 2013, Niagara Falls, NY

<u>Gale Boyd</u>, "A Statistical Approach to Plant-Level Energy Benchmarks and Baselines: The ENERGY STAR Manufacturing-Plant Energy Performance Indicator" Carbon Management Technology Conference, CMTC 150574 Orlando, Florida, USA, 7–9 February 2012.

<u>Gale Boyd</u>, Elizabeth Dutrow, Walt Tunnessen, Ernst Worrell, "Enabling the Energy Performance Improvement of Entire Industries and Measuring Success", *ACEEE Summer Study on Energy Efficiency in*

*Industry, "Energy Productivity in Industry: Partners and Opportunities"* July 26, 2011 - July 29, Niagara Falls, NY

Bloom Nicholas, <u>Gale Boyd</u>, Mark Curtis, Christos Genakos, Ralf Martin, and Raffaella Sadun, *Are Modern Management Practices a Source of Environmental Efficiency (or just Hot Air)? Examination of the US Experience* Southern Economic Association, San Antonio TX, Nov. 20th 2009

Tunnessen, W., <u>G. A. Boyd</u>, et al. (2007). *ENERGY STAR Industrial Focus for Pulp and Paper Mills - PPIC 2007 PANEL*. Pulp and Paper Industry Technical Conference, Williamsburg, VA

Gale A. Boyd and <u>Walt Tunnessen</u>, "Motivating Industrial Energy Efficiency through Performance-based Indicators" in *Proceedings - 2007 ACEEE Summer Study on Energy Efficiency in Industry IMPROVING INDUSTRIAL COMPETITIVENESS: ADAPTING TO VOLATILE ENERGY MARKETS, GLOBALIZATION, AND ENVIRONMENTAL* 2007, White Plains, New York

<u>Gale A. Boyd</u>, "Motivating Auto Manufacturing Energy Efficiency through Performance-based Indicators" in *Proceedings – Society of Automotive Engineering 2006 World Congress* April 3-6, 2006, Detroit, Michigan

Gale A. Boyd and <u>Elizabeth Dutrow</u>, "Motivating Industrial Energy Efficiency through Performance-based Indicators" in *Proceedings - 2005 ACEEE Summer Study on Energy Efficiency in Industry CUTTING THE HIGH COST OF ENERGY* July 19th - 22nd, 2005, West Point, New York

<u>Gale A. Boyd</u>, "Measuring Plant Level Energy Efficiency When Production Activities Are Not Homogeneous: The ENERGY STAR Energy Performance Indicator" in *Proceedings - 2005 ACEEE Summer Study on Energy Efficiency in Industry CUTTING THE HIGH COST OF ENERGY* July 19th - 22nd, 2005, West Point, New York

Gale A. Boyd, Marianne Mintz, <u>Ki Young Park</u>, and Anant Vyas, "Modal Choice in Product Shipments: Analysis of Nonpublic Census Microdata*"* in *TRANSPORTATION RESEARCH CIRCULAR* E-C088, January 2006, Transportation Research Board, Washington, DC 20001, Commodity Flow Survey (CFS) Conference, Boston, Massachusetts, July 8-9, 2005

<u>Guenter Conzelmann</u>, Gale Boyd, Vladimir Koritarov, and Tom Veselka, "Multi-Agent Power Market Simulation using EMCAS." Invited session on Economic Agents for Power Market Simulation, *2005 IEEE Power Engineering Society General Meeting,* June 12–16, 2005, San Francisco, California

<u>Gale A. Boyd</u>, "Stochastic Frontier Regressions for Measuring the Efficiency Gap between Average and Best Practice Energy Use."
       Illinois Economics Assoc., Oct. 15-16, 2004, Chicago IL
       and
       Southern Economic Assoc. 74th Annual Meeting, Nov. 21-23, 2004 New Orleans, LA

Conzelmann, Guenter, Gale Boyd, Richard Cirillo, Vladimir Koritarov, <u>Charles Macal</u>, Michael North, Prakash Thimmapuram, and Thomas Veselka. 2004. "Analyzing the Potential for Market Power Using an Agent-Based Modeling Approach: Results of a Detailed U.S. Power Market Simulation." Pp. 109-114 in *Proc. International Conference on Computing, Communication and Control Technologies: CCCT '04, Vol. VI*, edited by M.J. Savoie, H.-W. Chu, J. Michael, and P. Pace. Austin, Texas, USA: The University of Texas at Austin and IIIS: The International Institute of Informatics and Systemics.

Gale Boyd, Richard Cirillo, Guenter Conzelmann, Vladimir Koritarov, Charles Macal, Prakash Thimmapuram, and Tom Veselka, "Modeling the Restructured Illinois Electricity market as a Complex Adaptive System" *Energy, Environment, and Economics in a New Era*, 24th North American Conference of the USAEE/IAEE, Washington DC USA (July 8-10, 2004)

Worrell, E., S. Ramesohl, and G. Boyd, "Towards Increased Policy Relevance in Energy Modeling," *2003 ACEEE Summer Study on Energy Efficiency in Industry: Sustainability and Industry: Increasing Energy Efficiency and Reducing Emissions* (July 29-Aug. 1, 2003).

Boyd, G., "Parametric Approaches for Measuring the Efficiency Gap between Average and Best Practice Energy Use," *2003 ACEEE Summer Study on Energy Efficiency in Industry: Sustainability and Industry: Increasing Energy Efficiency and Reducing Emissions* (July 29-Aug. 1, 2003).

Boyd, G., "Measuring the Efficiency Gap between Average and Best Practice Energy Use," *2003 International Society for Industrial Ecology Second International Conference*, University of Michigan, Ann Arbor, MI (June 29-July 2, 2003).

Schoepfle, B., T. Veselka, D. Cirillo, G. Boyd, G. Conzelmann, V. Koritarov, C. Macal, M. North, and P. Thimmapuram "Modeling Electricity Market Behavior with Agent-Based Simulation," *99th Annual Meetings of the Association of American Geographers, Sessions on Geographical Perspectives on Complexity Theory and Complex Systems*, New Orleans, LA (Mar. 5-8, 2003).

Boyd, G., "Frontier Modeling with $CO_2$ as an Undesirable Output: Implications for Global Climate Policy," *72nd Annual Conference of the Southern Economics Association*, New Orleans, LA (Nov. 24-26, 2002).

Worrell, E., S. Ramesohl, and G. Boyd, "Towards Improved Realism and Policy Relevance in Engineering-Economic Analysis," *INEDIS/EETA Workshop on Policy Modeling in Industrial Energy Analysis*, Seoul, Korea (Nov. 7-8, 2002).

Veselka, T., G. Boyd, et al. "Simulating the Behavior of Electricity Markets with an Agent-Based Methodology: The Electricity Market Complex Adaptive System (EMCAS) Model," *22nd USAEE/IAEE North American Conference — Energy Markets in Turmoil: Making Sense of It All,* Vancouver, British Columbia, Canada (Oct. 6-8, 2002).

Boyd, G., "Industrial Sector Energy Intensity Trends and Projections," *e-vision 2002: Shaping our Future by Reducing Energy Intensity in the U.S. Economy*, Proceedings Vols. I and II, Rand Science and Technology Policy Institute, Washington, DC (May 14-16, 2002).

North, M., V. Koritarov, G. Boyd, T.D. Veselka, C.M. Macal, G.C. Conzelmann, and P.R. Thimmapuram, "E Laboratories: Agent-Based Modeling of Electricity Markets," *American Power Conference,* electronic proceedings, PennWell Corporation, Chicago, IL (April 2002).

North, M., C. Macal, V. Koritarov, T. Veselka, amd G. Boyd, "Fighting the Darkness: Illuminating Vulnerabilities in California's Electricity Pricing Policy," *The First Lake Arrowhead Conference on Computational Social Science Agent-Based Modeling in the Social Sciences*; Lake Arrowhead, CA (May 9-12, 2002).

Roop, J.M., D.B. Belzer, G.A. Boyd, and H. Huntington, "Indicators of Energy Intensity in the Economy: A Proposed Methodology," *Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy Expert Workshop on Indicators of Energy Intensity for the United States* (Jan. 11, 2001).

Boyd, G.A., "A Probit Model of Energy Efficiency Technology Decision Making," *Increasing Productivity through Energy Efficiency, 2001 ACEEE Summer Study on Energy Efficiency in Industry*, Tarrytown, NY (July 24-27, 2001).

Dowd, J., G.A. Boyd, and K. Friedman, "How Well Do Voluntary Agreements and Programs Perform at Improving Industrial Energy Efficiency?" *Increasing Productivity through Energy Efficiency, 2001 ACEEE Summer Study on Energy Efficiency in Industry*, Tarrytown, NY (July 24-27, 2001).

Boyd, G.A., and J.A. Laitner, "Recent Trends in the U.S. Energy Intensity: An Index Number Analysis," *2001: An Energy Odyssey, 24th Annual IAEE International Conference*, Houston, TX (Apr. 25-27, 2001).

Sanstad, A.H., S.J. DeCanio, and G.A. Boyd, "Combining Technological and General Equilibrium Perspectives on Carbon Abatement," *Western Economic Association*, Vancouver, British Columbia, Canada (July 2, 2000).

Sanstad, A.H., S.J. DeCanio, and G.A. Boyd, "Estimating the Macroeconomic Effects of the Clean Energy Future Policy Scenarios," *Proceedings of the Air and Waste Management Association 93rd Annual Conference and Exhibition,* Salt Lake City, UT (June 18-22, 2000).

R. Michael, S. Bernow, G. Boyd, R.N. Elliot, and J. Roop, "Analytical Approaches to Measuring the Potential for Carbon Emission Reductions in the Industrial Sectors of the United States and Canada," *International Workshop on Technologies to Reduce Greenhouse Gas Emissions: Engineering-Economic Analyses of Conserved Energy and Carbon,* Washington, DC (May 5-7, 1999).

Boyd, G.A., R. Färe, and S. Grosskopf, "International Productivity Growth with $CO_2$ as an Undesirable Output," Eastern Economic Association, Boston, MA (1998); also Southern Economic Association, New Orleans, LA (Nov. 1999).

Boyd, G., J. Roop, and M. Woodruff, "The Potential for Industrial Energy Efficiency Technology to Reduce U.S. Greenhouse Gas Emissions," *How Industry Will Procure Energy Efficiency Services in the 21st Century, 1997 ACEEE Summer Study on Energy Efficiency in Industry,* Saratoga Springs, NY (July 8-11, 1997).

Boyd, G., J. Roop, and M. Woodruff, "The Potential for Energy Efficient Technologies to Reduce Carbon Emissions in the United States: The Industrial Sector," *32nd Annual Intersociety Energy Conversion Engineering Conference,* Honolulu, HI (July 27-Aug. 1, 1997).

Boyd, G., J. Molburg, and R. Prince, "Alternative Methods of Marginal Abatement Cost Estimation: Non-parametric Distance Functions," *17th Annual North American Conference of the International Association for Energy Economics, (De)Regulation of Energy: Intersecting Business, Economics, and Policy,* Boston, MA (Oct. 27-30, 1996); also *8th Annual Conference of the European Association of Environmental and Resource Economics,* Tilburg, Netherlands (June 26-28, 1997).

Corio, M., G. Boyd, J. Belluci, and K. Perl, "Generation Cost Frontier Analysis, Dynamic Market Adjustment, and Strategic Gaming," *59th Annual American Power Conference,* Chicago, IL (Apr. 1-3, 1997). (personal research)

Boyd, G., and J. McClelland, "The Impact of Environmental Constraints on Productivity Improvement and Energy Efficiency in Integrated Paper and Steel Plants," *17th Annual North American Conference of the International Association for Energy Economics, (De)Regulation of Energy: Intersecting Business, Economics, and Policy,* Boston, Mass. (Oct. 27-30, 1996).

Corio, M., G. Boyd, and J. Bellucci, "Potential Cost Savings in Electric Utility Non-Fuel Operating Costs under Deregulation," *17th Annual North American Conference of the International Association for Energy Economics, (De)Regulation of Energy: Intersecting Business, Economics, and Policy,* Boston, MA (Oct. 27-30, 1996).

Corio, M., and G. Boyd, "Applying the Competitive Business Equation to Power Generation Operations and Economics," *EPRI Fossil Plant Heat Rate Improvement Conference,* Dallas, TX (May 22-24, 1996).

Boyd, G., J. Dowd, J. Freidman, and J. Quinn, "Productivity, Energy Efficiency, and Environmental Compliance in Integrated Pulp and Paper and Steel Plants," *ACEEE Industrial Summer Study,* Grand Island, NY (Aug. 1-4, 1995).

Boyd, G., and M. Ross, "Industrial Energy Intensity: Modeling Decision-Making," *ACEEE Industrial Summer Study,* Grand Island, NY (Aug. 1-4, 1995).

Boyd, G., J. McClelland, and M. Ross, "The Impact of Environmental Constraints on Productivity Improvement and Energy Efficiency in Integrated Paper and Steel Plants," *Proceedings of the 18th IAEE International Conference, Into the 21st Century: Harmonizing Energy Policy, Environment, and Sustainable Economic Growth,* Washington, DC (July 5-8, 1995).

Corio, M., and G. Boyd, "Management of Bus Bar Costs and Spending Tradeoffs for the Transition to Competitive Markets in Electricity," *American Power Conference 57th Annual Meeting,* Chicago, IL (Apr. 18-20, 1995).

Boyd, G., and J. McClelland, "Strategies for Reconciling Environmental Goals, Productivity Improvement and Increased Energy Efficiency in the Industrial Sector: An Analytic Framework," *Proceedings, Energy Efficiency and the Global Environment: Industrial Competitiveness and Sustainability,* Newport Beach, CA, sponsored by Southern California Edison Company (Feb. 8-9, 1995).

Boyd, G., and M. Corio, "The Cost-Reliability Frontier: New Techniques for Evaluating Cost/Reliability Trade-offs and Targeting Plant Spending and Performance," *American Power Conference 56th Annual Meeting,* Chicago, IL (Apr. 25-27, 1994); also *1994 International Joint Power Generation Conference,* Phoenix, AZ (Oct. 1995); also *Power Gen International '94,* Cologne Germany (May17-19, 1994).

Boyd, G., "The Impact of Energy Prices on Industrial Energy Efficiency and Productivity," *Workshop on Partnerships for Industrial Productivity through Energy Efficiency,* sponsored by the American Council for an Energy Efficient Economy, Portland, OR (Sept. 21-22, 1993).

Boyd, G., and M. Ross, "The Role of Industrial Energy Efficiency and Conservation in Negating Environmental Impacts," *Workshop on Market-Based Approaches to Environmental Policy,* Chicago, IL (Mar. 17, 1993).

Boyd, G., S. Karlson, M. Neifer, and M. Ross, "Vintage Effects in the Steel Industry: The Potential for Energy Intensity Improvements," *Western Economics Association International 67th Annual Conference,* San Francisco, CA (July 9-13, 1992).

Hanson, D., et al., "Forecasting the Market for $SO_2$ Emission Allowances under Uncertainty," *Proceedings: Eighth Electric Utility Forecasting Symposium — The Foundations of Planning,* Electric Power Research Institute Report TR-100396, Baltimore, MD, held in Oct. 1991 (published in Apr. 1992).

Boyd, G., E. Kokkelenberg, and M. Ross, "Technology Trends in Econometric Energy Models: Ignorance or Information?", *Energy Disruptions: Proceedings of the International Association for Energy Economists 13th Annual North American Conference* (Nov. 18-20, 1991).

Fisher, R., W. Breed, and G. Boyd, "The Design, Results, and Future Development of the National Energy Strategy Environmental Analysis Model (NESEAM)," *Energy Disruptions: Proceedings of the International Association for Energy Economists 13th Annual North American Conference* (Nov. 18-20, 1991).

Hanson, D., et al., "The 1990 Clean Air Act Amendments and the Great Lakes Economy: Challenges and Opportunities," *Energy Disruptions: Proceedings of the International Association for Energy Economists 13th Annual North American Conference* (Nov. 18-20, 1991).

Karlson, S., and G. Boyd, "The Impact of Energy Prices on Technology Choice in the United States Steel Industry," *Energy Disruptions: Proceedings of the International Association for Energy Economists 13th Annual North American Conference* (Nov. 18-20, 1991).

Fox, J., D. Hanson, and G. Boyd, "An Integrated Energy Model and Acid Precipitation Emissions Policy Model," *16th Annual Convention of the Eastern Economic Association,* Cincinnati, OH (Mar.-Apr. 1990).

Boyd, G., and M. Ross, "Intersectoral Impacts of Energy Taxes as a Method of $CO_2$ Emission Control," *American Association for the Advancement of Science Annual Meeting,* New Orleans, LA. (Feb. 15-20, 1990).

Boyd, G., D. Hanson, and H.W. Hochheiser, "Pseudo-data Approach to Obtaining World Oil Market Price Elasticities and Relative Confidence of Conditional Forecasts," *World Energy Markets: Coping with Instability, Proceedings of the 9th International Conference of the International Association of Energy Economists* (J.G. Rowse, editor), Calgary, Alberta, Canada (July 1987).

Hanson, D.A., G. Boyd, and H.W. Hochheiser, "The Formation of Conditional Expectation in World Oil Markets," *1987 Eastern Economics Association Convention,* Washington, DC (Mar. 1987).

Hanson, D.A., G. Boyd, D. South, M. Browdy, and J. McDonald, "Generating Sector Specific Regional Forecasts for Nationwide Energy Assessments," *33rd North American Meeting of the Regional Science Association,* Columbus, OH (Nov. 1986).

Boyd, G., and M. Ross, "Energy Demand in Metals Manufacturing and Effects of Changing Composition of Production," *Proceedings of the 8th Annual North American Conference of the International Association of Energy Economists,* Cambridge, MA (Nov. 1986).

Boyd, G., H.W. Hochheiser, J.F. McDonald, and M. Ross, "Energy Intensity in Manufacturing: A Comparison of Historical Results with Forecasts," *Proceedings of the 8th Annual North American Conference of the International Association of Energy Economists,* Cambridge, MA (Nov. 1986).

Boyd, G., and J. Kelley, "An Integration Methodology Applied to the National Energy Policy Plan Environmental Assessment," *Proceedings of the 8th Annual North American Conference of the International Association of Energy Economists,* Cambridge, MA (Nov. 1986).

## Publications:  Technical Reports and Working Papers

Schneck, Joshua, and Gale Boyd. "*Distribution of Emissions Permits to the U.S. Pulp and Paper Sector under Alternative Output-Based Allocation Schemes.*" Nicholas Institute for Environmental Policy Solutions Report NI R 11-10, Duke University Durham, NC: (November, 2011)

Gale A. Boyd and Gang Zhang, "*Measuring Improvement in the Energy Performance of the U.S. Cement Industry*" Nicholas Institute for Environmental Policy Solutions: Duke Environmental Economics Working Paper Series: EE 11-05 (May 2011)

Gale Boyd, Tatyana Kuzmenko, Béla Személy, & Gang Zhang, "*Preliminary Analysis Of The Distributions Of Carbon And Energy Intensity For 27 Energy Intensive Trade Exposed Industrial Sectors,*" Nicholas Institute for Environmental Policy Solutions: Duke Environmental Economics Working Paper Series: EE 11-03 (April 2011)

Gale A. Boyd, "*Assessing Improvement in the Energy Efficiency of U.S. Auto Assembly Plants*" Nicholas Institute for Environmental Policy Solutions: Duke Environmental Economics Working Paper Series: EE 10-01 (June 2010)

Gale A. Boyd, "Development of a Performance-based Industrial Energy Efficiency Indicator for Pharmaceutical Manufacturing Plants," report to the Environmental Protection Agency (Sept 30, 2009)

Joshua Schneck, Brian Murray, Jan Mazurek, and Gale Boyd, "Protecting Energy-Intensive Trade-Exposed Industry," Nicholas Institute Discussion Memo on H.R. 2454 American Clean Energy and Security Act of 2009 (October 2009)

Gale A. Boyd, "Development of a Performance-based Industrial Energy Efficiency Indicator for Food Processing Plants," report to the Environmental Protection Agency (Sept 30, 2009)

Gale A. Boyd, "Development of a Performance-based Industrial Energy Efficiency Indicator for Glass Manufacturing Plants," report to the Environmental Protection Agency (Sept 30, 2009)

Gale A. Boyd, "Estimating Plant Level Manufacturing Energy Efficiency with Stochastic Frontier Regression", Center for Economic Studies working paper **CES**-WP-07-07

Gale A. Boyd, *Development of a Performance-based Industrial Energy Efficiency Indicator for Cement Manufacturing Plants*, Argonne National Laboratory, ANL/DIS-06-3 (July 2006)

Gale A. Boyd, *Development of a Performance-based Industrial Energy Efficiency Indicator for Corn Refining Plants*, Argonne National Laboratory, ANL/DIS-06-4 (July 2006)

Gale A. Boyd, *Development of a Performance-based Industrial Energy Efficiency Indicator for Automobile Assembly Plants*, Argonne National Laboratory, ANL/DIS-05-3 (May 2005)

K.L. Stamber, G.A. Boyd, S.J. Flaim, S. Folga, D.J. Gotham, M. McLamore, M.H. Novak, J.M. Roop, C.G. Rossmann, S. Shamsuddin, and L.M. Zeichner, *Economics in Criticality and Restoration of Energy Infrastructure*s, Sandia National Laboratories Albuquerque, New Mexico (March 2005)

Boyd, G., *Technical Description of the ENERGY STAR Industrial Energy Performance Indicator,* report to the Environmental Protection Agency (May, 2004)

Boyd, G., *Benefits to the United States Bureau of the Census from CES Project #81 Comparison of the Distributions of Production and Energy Efficiency in Manufacturing*, Final Status Report to the Center for Economic Studies, U.S. Bureau of the Census (Jan. 31, 2003).

Worrell, E., S. Ramesohl, and G. Boyd, "Towards Improved Realism and Policy Relevance in Engineering-Economic Analysis," report prepared for the Engineering-Economic Technology Analysis Network: EETA (Aug. 2002).

Boyd, G., *Baseline Methodology Issues for Industrial Clean Development Mechanism (CDM) Projects,* report prepared for Office of Policy, U.S. Department of Energy (Aug. 2000).

Boyd, G., and Jeffrey Dowd, *Domestic Early Action Crediting and the Relationship to Voluntary Programs for Greenhouse Gas Reduction,* report prepared for Office of Policy, U.S. Department of Energy (May 2000).

Interlaboratory Working Group, "Estimating Bounds on the Macroeconomic Effects of the CEF Policy Scenarios," Appendix E-4 in *Scenarios for a Clean Energy Future,* ORNL/CON-476 and LBNL-44029 (Nov. 2000).

Interlaboratory Working Group on Energy Efficient and Low-Carbon Technologies, "Industry," Chapter 5 in *Scenarios of U.S. Carbon Reductions: Potential Impacts of Energy Technologies by 2010 and Beyond,* ORNL/CON-444 and LBNL-40533 (Sept. 1997).

U.S. Department of Energy, *The Interrelationship between Environmental Goals, Productivity Improvement, and Increased Energy Efficiency in Integrated Paper and Steel Plants,* USDOE/PO-055 Technical Report 5 (June 1997).

Boyd, G., *Market Segmentation Analysis and Forecasts for the Application of 5-15 MW Advanced Turbine Systems (ATS) in Industrial Establishments*, final consultant report prepared for Energy Resources International, Inc. (Feb. 1997).

Boyd, G., *Impact of Voluntary Agreements on Greenhouse Gas Emission in US Industry,* final consultant report to the Environment Directorate, Organization for Economic Cooperation and Development (Feb. 1996).

Molburg, J.C., G.A. Boyd, and J.D. Cavallo*, Learning Functions for New Electric Power Generating Technologies*, report to the Department of Energy (Sept. 1995).

Bailey, K., G. Boyd, T. Elliott, R.E. Fisher, J. Formento, K. Guziel, E. Kohout, A. Loeb, J. Molburg, E. Portante, C. Saricks, D. South, D. Streets, P. Thimmapuram, and M. Wang, *Development of Emission Control and New Technology Options for the Grand Canyon Visibility Transport Region, Volume 1*, Argonne National Laboratory, Argonne, IL (Oct. 1995).

Molburg, J.C., et al., *Factors Affecting $SO_2$ Emission Projections: A Preliminary Assessment*, report to the Department of Energy (Mar. 1994).

Boyd, G., and A. Fish, "Industrial Sector," Chapter 4 in *Guidelines for Voluntary Reporting of Greenhouse Gas Emissions, Emission Reductions, and Carbon Sequestration Activities — Energy Policy Act, Section 1605(b)*, U.S. Department of Energy, Office of Policy, Planning and Program Evaluation (Feb. 22, 1994).

Boyd G., M. Bock, S. Karlson, and M. Ross, *Vintage-Level Energy and Environmental Performance in Manufacturing Establishments,* ANL/DIS/TM-15 (May 1994).

Bock, M.J., G. Boyd, D. Rosenbaum, and M. Ross, *Case Studies of the Potential Effects of Carbon Taxation on the Stone, Clay, and Glass Industries,* ANL/EAIS/TM-91 (Dec. 1992).

Tompkins, M., et al., *Retrospective Clean Air Act Analysis: Sectoral Impact on Emissions and Control Costs from 1975 to 1990*, report to the Department of Energy (Aug. 1992).

Boyd, G., M. Neifer, and M. Ross, *Modeling Plant Level Industrial Energy Demand with the Longitudinal Research Database and the Manufacturing Energy Consumption Survey Database*, ANL/EAIS/TM-96 (Jan. 1992).

Tompkins, M., et al., *Retrospective Clean Air Act Analysis: Phase I Scoping Report*, report to the Department of Energy (Dec. 1991).

Boyd, G.A., et al., "Industrial Technology," Chapter 7 in *Report to the Congress of the United States, Limiting Net Greenhouse Gas Emissions in the United States, Vol. I, Energy Technologies*, DOE/PE-0101 (R.A. Bradley et al., editors), U.S. Department of Energy (Sept. 1991).

Boyd, G.A., J.C. Molburg, and M.H. Ross, *Potential for Environmental Impact Mitigation and Energy Conservation in Industry: Case Studies of Steel and Paper Industries*, report to the Department of Energy (Nov. 1991).

Fisher, R., G. Boyd, J. Molburg, G. Pandola, and C. Saricks, *National Energy Strategy Environmental Analysis Model (NESEAM): Documentation and Findings*, report to the Department of Energy (June 1991).

Marciniak, T.J., R.L. Cole, and G.A. Boyd, *Gravel Bed Combustor System: Preliminary Technical, Economic, and Market Analysis,* report to the Department of Energy (May 1991).

Bloyd, C., et al., *Backcasting Performance of the NAPAP Emissions Model Set*, ANL/EAIS/TM-32 (Feb. 1990).

Rausch, J., and G. Boyd, *Regional Emissions Data Base and Evaluation System (REDES): Technical and System Documentation of Version 2.0,* ANL/EAIS/TM-43 (Mar. 1991).

Boyd, G., M. Ross, and E. Kokkelenberg, *Sectoral Electricity and Fossil Fuel Demand in U.S. Manufacturing*, ANL/EAIS/TM-35 (Mar. 1991).

David Streets et al, *National Acid Precipitation Assessment Program Chapter 26: Methods for Modeling Future Emissions and Control Costs,* Washington DC, (Dec 1990)

Boyd, G., et al., *Technology Characterizations and Policy Options to Reduce GHG Emissions: Industrial Sector*, report to the Department of Energy (June 1989).

South, D.W., J.F. McDonald, M.J. Bragen, and G.A. Boyd, *Industrial VOC Model: Regionalized Forecasts of Uncontrolled Emissions by Source Category*, ANL/EES/TM-306 (Nov. 1988).

Boyd, G., C. VanGroningen, and L. Abel, *REDES Volume I: Overview and Methodology of the Regional Emissions Database and Evaluation System,* ANL/EES/TM-351, Vol. 1 (June 1988).

Boyd, G., M. Davis, and T. Veselka, *REDES Volume II: Methodology Documentation for the Regional Emissions Evaluation Database (REED)*, ANL/EES/TM-351, Vol. 2 (Aug. 1988).

Poch, L., J. Gillette, G. Boyd, *REDES Volume III: Environmental Profiles of Selected Clean Coal Technologies,* ANL/EES/TM-351, Vol. 3 (June 1988).

Mead, D., and G. Boyd, "Data Use by the INFORUM Model," in *Industrial Energy Demand,* Energy Modeling Forum Report No. 8 (1988).

Boyd, G., J.F. McDonald, M. Ross, and D.A. Hanson, "Effects of the Changing Composition of U.S. Manufacturing Production on Energy Demand," in *Industrial Energy Demand,* Energy Modeling Forum Report No. 8 (1988).

Boyd, G., R. Hemphill, D. Jankowski, and D. Hanson, "Fuel Choice Decision Making in Industrial Energy Models Examined by the Energy Modeling Forum 8," in *Industrial Energy Demand,* Energy Modeling Forum Report No. 8 (1988).

Boyd, G., and T. Kuczmowski, "Structural Shift of Energy Demand in the Macroeconomy," in *Industrial Energy Demand,* Energy Modeling Forum Report No. 8 (1988).

Boyd, G., J.F. McDonald, M. Ross, and D.A. Hanson, *Effects of the Changing Composition of U.S. Manufacturing Production on Energy Demand*, ANL/EES/TM-298 (Feb. 1986).

Boyd, G.A., M. Ross, C.M. Macal, and D.A. Hanson, *Disaggregation of Industrial Fossil Fuel Use in the 1985 National Energy Policy Plan: Methodology and Results*, ANL/EES/TM-303 (Jan. 1986).

D.A. Hanson, G.A. Boyd, M.J. Bragen, M. Browdy, and J.R. Peerenboom, *Advanced Utility Simulation Model (AUSM): State-Level Projections of Required Electrical Generation and Sources of Supply*, ANL/EES/TM-357 (Sept. 1985).

South D.W., M.J. Bragen, D.A. Hanson, and G.A. Boyd, *Advanced Utility Simulation Model (AUSM): Regionalized End-Use Electricity Demand Projections,* ANL/EES/TM-300 (June 1985).

P. Pieper, D.A. Hanson, D.W. South, and G.A. Boyd, *Review of the DRI Long-Term Macroeconomic Projections and Their Extension to the Year 2030*, ANL/EES-TM-358 (Apr. 1985).

Arey, D., D. Bauman, G. Boyd, B. Djinglewski, J. Crenshaw, J. Myers, and G. Parker, *Critical Points in the Coal Fuel Cycle: A Modeling Analysis*, prepared by Coal Extraction and Utilization Research Center, Southern Illinois University, Carbondale, IL, for U.S. Department of Energy (1984).

Myers, J., R. Färe, B. Yoon, G. Boyd, P. Byrnes, M. Mumuni, and K. Sung, *Productivity in Strip Mining of Coal*, USDOE/CMTC 6-22532, prepared by Coal Extraction and Utilization Research Center, Southern Illinois University, Carbondale, IL, for U.S. Department of Energy (1981).

**Oral Presentations (no paper or proceedings):**

Gale Boyd, "Statistical Energy Benchmarking for Manufacturing Plants: The ENERGY STAR Energy Performance Indicators (EPI)," invited presentation to the *Western Climate Initiative Symposium on Understanding the Value of Benchmarking,* Seattle WA, May 19, 2010

Nicholas Bloom, <u>Gale Boyd</u>, Mark Curtis, Christos Genakos, Ralf Martin, Raffaella Sadun, "Are Modern Management Practices a Source of Environmental Efficiency (or just Hot Air)? Examination of the US Experience", *Southern Economic Association*, San Antonio TX, Nov. 20th 2009

Gale Boyd, "Updates to the Paper and Paperboard Manufacturing Plant Energy Performance Indicator (EPI)" *ENERGY STAR Focus Meeting*, Washington DC, Sept 29-30, 2008

Gale Boyd, "Development and Testing of Glass Manufacturing Energy Performance Indicators (EPI)", *ENERGY STAR Focus Meeting*, Washington DC, Sept 29-30, 2008

Gale Boyd, "Development and Testing of Food Processing Energy Performance Indicators (EPI)", *ENERGY STAR Focus Meeting*, Washington DC, Sept 29-30, 2008

Gale Boyd, "Updates to the Petrochemicals Manufacturing Plant Energy Performance Indicator (EPI)" *ENERGY STAR Focus Meeting*, Washington DC, Sept 29-30, 2008

Gale Boyd, "Development and Testing the Paper and Paperboard Manufacturing Plant Energy Performance Indicator (EPI)," Presented to the *1st annual Pulp and Paper Industry ENERGY STAR Focus Meeting*, Washington DC, February 6, 2008

Gale Boyd, "Development and Testing the Petrochemical Manufacturing Plant Energy Performance Indicator (EPI)," Presented to the *1st annual Petrochemical Industry ENERGY STAR Focus Meeting*, Washington DC, December 13, 2007

Gale Boyd, Vince Gates, Mike McLoughlin, "The Art of Benchmarking Energy Use in a Manufacturing Plant" *World Energy Engineering Conference* Sept 7-12, 2006, Washington DC

Gale A. Boyd, "Estimating the Distribution of Plant Level Manufacturing Energy Efficiency with Stochastic Frontier Regression", research seminar presented to *Institute for Sustainable Energy, Environment, and Economy*, University of Calgary, November 17, 2005

Gale A. Boyd, "Searching for the 'Win-Win' Approach: The Potential for Improvements in Energy, Environmental and Economic Performance", public lecture invited by the *Institute for Sustainable Energy, Environment, and Economy*, University of Calgary, November 18, 2005

Gale A. Boyd (session chair) "Research Opportunities at the Chicago RDC", *35TH Annual Meeting of Illinois Economic Association*, October 21-22, 2005, University of Illinois at Chicago

Boyd, G., Betsy Dutrow, "Brainstorming the use of ENERGY STAR's rating approach as an incentive for continuous improvement in industry",*2005 ACEEE Summer Study on Energy Efficiency in Industry,* July 19th - 22nd, 2005, West Point, New York

Boyd, G., "Energy Performance Indicator for Cement Manufacturing," *2nd Annual ENERGY STAR Cement Industry Focus Meeting*, Tucson AZ, (May 5, 2005)

Boyd, G. (invited), "Energy Performance Indicators at Auto Assembly Plants: Statistical Tools for Assessing Plant Level Energy Performance," *12th Annual Great Lakes Region Waste Reduction & Energy Efficiency Workshop,* Livonia MI (October 26, 2004)

Boyd, G., "Stochastic Frontier Regressions for Measuring the Efficiency Gap Between Average and Best Practice Energy Use", *Illinois Economics Association*, Chicago Illinois (October 15th, 2004)

Boyd, G. (invited), "Methodology and Results of the Energy Performance Indicator for Cement Manufacturing," *Portland Cement Association 2004 Joint Fall Meeting,*, Chicago IL (September 27-29th, 2004)

Boyd, G., "Benchmarking Energy Use: The ENERGY STAR Corn Refining Plant Energy Performance Indicator," *2nd Annual ENERGY STAR corn refiners Focus Workshop,* Indianapolis, IN (September 9th, 2004)

Boyd, G., "Summary of Results from Revisions to the ENERGY STAR® Auto Assembly Plant Energy Performance Indicator", *3rd annual ENERGY STAR Auto Industry Focus Meeting*, Georgetown KY (August 11th 2004)

Boyd, G., "Results from of the Cement Manufacturing Energy Performance Indicator (EPI) Version 1.0," *ENERGY STAR® Cement Industry Focus Meeting*, Washington, DC (May 5th 2004)

Bhashkar Mazumder, Gale Boyd, and Lynn Riggs "Overview of the Midwests' Census Research Data Centers". *Midwest Economics Association Annual Meeting* Chicago IL (March 19-21, 2004)

Boyd, G., "Development and Testing Experience with the Energy Performance Indicator (EPI)," *ENERGY STAR Corn Refiners Industry Focus Meeting,* Washington, DC (December 11, 2003)

Boyd, G., "Current Status of the Auto Assembly Energy Performance Indicator," *2nd Annual ENERGY STAR® Motor Vehicle Industry Focus Meeting,* Detroit, MI (Aug. 15, 2003)

Boyd, G., B. Mazumder, and T.L. Riggs, *Research Opportunities at the Chicago Census Research Data Center,* series of invited workshops presented at Northwestern Institute of Policy Research (Feb. 10, 2003), University of Illinois – Chicago (Nov. 13, 2003), and The University of Chicago (Mar. 7, 2003).

Boyd, G., and T. Hicks, "Energy Performance Indicators," *ENERGY STAR® Motor Vehicle Industry Focus Meeting and ENERGY STAR® Brewery Industry Focus Meeting,* Washington, DC (June 2002).

Boyd, G., and T. Hicks, "Benchmarking Plant Productivity," *Earth Technologies Forum*, Washington, DC (Mar. 25-27, 2002)

Boyd, G., "Modeling the Impacts of Voluntary/Information Programs on US Industry's Energy Intensity," *National Energy Modeling System/Annual Energy Outlook Conference,* Arlington, VA (Mar. 12, 2002).

Boyd, G., "Baseline Issues for Determining GHG Emissions," Air and Waste Management Association 94th Annual Conference and Exhibition, Orlando, FL (June 26, 2001).

Boyd, G.A., "On the Frontier: Tools for Representing Technical Change and Efficiency," Energy Modeling Forum Workshop on Modeling Technological Change, Washington DC (June 5-7, 2001).

Boyd, G.A., "Recent Developments in U.S. Energy Markets," *Argonne National Laboratory/Department of Energy Community Leaders Roundtable* (May 9, 2001)

Boyd, G.A., K. Sang, and G. Tolley, "A Comparison of Environment-Adjusted Regional Productivity Growth in South Korea," *The University of Chicago Department of Economics Workshop in Economic Policy and Public Finance* (Feb. 14, 2001).

Boyd, G., "Global Warming: The Role of Technology in Limiting Greenhouse Gas Emissions," *Chicago Chapter of the United States Association of Energy Economics* (Jan. 29, 1998).

Boyd, G., "Industrial Energy Use after the Kyoto Agreement: Analytic Issues and Empirical Evidence," *28th Annual Meeting of the Illinois Economics Association* (Oct. 24, 1998).

Boyd, G., and J. McClelland, "Plant Level Productivity and Efficiency: The Integrated Pulp and Paper Industry," *First Industrial Energy Efficiency Symposium and Expo,* sponsored by the Office of Industrial Technologies, U.S. Department of Energy (May 1-3, 1995).

Boyd, G., "Frontier Analysis of Plant Level Costs: Application to Electric Power Reliability Costs," *Chicago Chapter of the International Association for Energy Economics* (Mar. 13, 1995).

Boyd, G., "End Use Technology Scenarios for the NEMS Industrial Module," *Energy Information Administration Workshop on 1995 Annual Energy Outlook Scenarios,* Washington, DC (June 6, 1994).

Boyd, G., "Energy Conservation as Pollution Prevention," *International Joint Commission 1993 Biennial Meeting,* Windsor, Ontario, Canada (Oct. 22-24, 1993).

Boyd, G., "A Production Function Approach to a Technology Adoption Model," *Illinois Economics Association,* Chicago, IL (Oct. 1-2, 1993).

Boyd, G. "Industrial Plant Vintage and Energy Intensity," *Center for Economic Studies* (Dec. 2, 1992).

Boyd, G., and M. Neifer, "Production Frontier Efficiency Measures in Energy Intensive Industries," *Midwest Economic Association,* Chicago, IL (Mar. 26-28, 1992).

Boyd, G., "Industrial Energy Use," *Illinois Economic Association,* Chicago, IL (Oct. 1990).

# ROBERT CALDERBANK

**1411 Woodland Drive**                          **Office: (919) 668 - 2728**
**Durham, NC 27701**                          **Email: robert.calderbank@duke.edu**


**Duke University: Founding Director, Information Initiative at Duke, 2013 to date**
**Professor of Computer Science, Mathematics and Electrical Engineering**
- Dean of Natural Sciences (2010-2013), responsible for research and teaching by 250 faculty in 8 departments
- Recruited 27 faculty including 18 women and under-represented minorities with the Chemistry Department tripling the number of women faculty
- Interim director of the Duke Initiative in Innovation and Entrepreneurship
- Initiated Duke in Silicon Valley and outreach to the Arts

**Princeton University: Professor of Mathematics and Electrical Engineering, and Director of the Program in Applied and Computational Mathematics, 2004 - 2010**
- Applied Mathematics ranked #1 in the 2010 NRC National Assessment
- Partnership with Mathematics doubled the number of mathematics majors

**AT&T Labs: VP Research, 2002 - 2003**
- Corporate Officer responsible for the first major Research Lab in the world focused on big data from business and network operations.
- Managed AT&T intellectual property with responsibility for licensing revenue.

**AT&T Labs: Information Sciences Research VP, 1997-2002**
- Pioneered big data analytics by creating and operating systems that capture the interaction of customers, networks and service.
- Invented wireless space-time codes found in billions of 4G cell phones and error correcting codes essential to quantum computing

**AT&T Labs / Bell Laboratories:  Research Department Head, 1989 – 1997**
                                      **Member of Technical Staff, 1980 – 1989**
- Developed magnetic recording technology that transformed AT&T Microelectronics into a commercially based supplier of electronic components
- Invented technology enabling a billion voiceband modems to communicate at close to the fundamental Shannon limit.

**Flarion Qualcomm Technologies: Technical Advisory Board, 2000 to date**
- OFDM pioneer acquired by Qualcomm in 2008 for $1B


**Awards**

Honorary Doctorate, Warwick University, 2013
National Academy of Engineering, 2005
IEEE Hamming Medal (2013), IEEE Millennium Medal (2000), IEEE Donald G. Fink
Prize Paper Award (2006), IEEE Information Theory Prize Paper Award (1995 and 1999)
AAAS Fellow (2013), AMS Fellow (2013) AT&T Fellow, 2000, and IEEE Fellow (1995)
Graduate Mentoring Award, Princeton University, 2008

**Education:**

**California Institute of Technology, 1977-1980:** PhD in Mathematics
**Oxford University, England, 1975-1976:** M.Sc. in Mathematics
**Warwick University, England, 1972-1975:** B.Sc. in Mathematics (first class honors)

## Professional Service

**AT&T:** Making connections between industry and universities:

- AT&T Fellowship Program: Migrated this Bell Laboratories program to AT&T and continued its extraordinary 35 year impact on the representation of under-represented minority and women students in science and engineering.
- Initiated collaborative research and student internship programs with Carnegie-Mellon, Rensselaer Polytechnic Institute, Georgia Tech, and the University of California at San Diego.

**Advisory Boards:** Developing strategic vision

- **USA:** Institute for Pure and Applied mathematics (Scientific Advisory Board – 2013 to date), American Institute of Mathematics (Scientific Advisory Board – 2006 to date), Rice University (Electrical Engineering Department and Engineering School – 2006 to date), University of Pennsylvania (Chair, Electrical Engineering External Review Committee – 2010), University of Michigan (Chair, Mathematics External Review Committee – 2009), Flarion Technologies (Technical Advisory Board – 2003 to 2008), Institute for Mathematics and Applications (Board of Governors – 1996 to 1999)

- **Europe:** Dresden University, Germany: cfaed - Center for Advancing Electronics Dresden (Scientific Advisory Board – 2012 to date)**,** University College Dublin, Institute for Complex and Adaptive Systems (Chair, International Advisory Board, 2007 to date)**,** RWTH Aachen University, Germany; UMIC Research Center - ultra high-speed mobile information and communication (Scientific Advisory Board -2007 to 2012),

- **Asia and Australia:** Ministry of Education, Singapore (Panel of Experts – Informatics and Mathematics, 2006 to date), Institute for Telecommunications Research, University of South Australia (Chair, External Review Committee – 2008 -development and presentation of a strategic plan that reinvented the organization, NICTA (National Information and Communications Technology, Australia) Sensor Network Program (Chair, External Review Committee – 2008)

**Professional Service:** Assessing and encouraging scholarship

- **National Academy of Engineering:** Membership Committee, Section 7, (2013 to date), German - American Frontiers of Engineering (Organizer, Wireless

Communications Session, Dresden, 1998, and Information Technology Session, Irvine, 1999)

- **IEEE:** IEEE Koji Kobayashi Computers and Communication Award Committee (2006 to 2009 - Chair in 2009), IEEE Richard W. Hamming Medal Committee (2006 to 2008), IEEE Prize Papers/Scholarship Awards Committee (2006 to 2009)

- **IEEE Information Theory Society:** Fellow Evaluation Committee (2011 to date – Chair since 2013), Board of Governors (1991-1996 and 2006-2008), IEEE International Symposium on Information Theory (Chair, Technical Program Committee – 2009), IEEE Transactions on Information Theory (Editor in Chief - 1995 to 1998, Guest Editor, Special Issue on Coding for Storage Devices – 1991, Associate Editor for Coding Techniques -1986 to1989).

- **IEEE Communications Society:** Fellow Evaluation Committee (2001- 2005 and Chair in 2005),  COMSOC Prize Committee (2005-2006)

- **National Science Foundations:** Science Foundation of Ireland – Mathematics Initiative aimed at revitalizing mathematics in Ireland and enabling industrial growth (Selection Committee, 2006), NSF Committee of Visitors, Division of Mathematical Sciences (2001)

- **American Mathematical Society:** Committee to Select the Gibbs Lecturer (2009 to 2012), Chair, Organizing Committee for the 2006 von Neumann Symposium – selected the theme of compressive sensing, Committee to Select Speakers at National Meetings (2006 to 2009), Committee to Select Speakers at NE Regional Meetings (2003 to 2006), AMS Centennial Fellowship Committee (1995 to1998 and Chair 1996 to1998)

Daniel Egger – Summary Bio

Executive in Residence, and Director of the Center for Quantitative Modeling (CQM), Pratt School of Engineering, Duke University (Since 2009)

Twenty years experience launching successful Data Analytics ventures, as technologist and CEO, mentor and member of Board of Directors of others' companies, and as Managing Partner of a seed capital fund.

Direct commercial experience with directed graph (hyperlink) analysis for search, insurance claim fraud detection, multi-factorial methods for cancer diagnostics, insurance risk pricing models (including developing novel insurance products with Lloyd's of London), GPS location-based marketing technologies, signal detection and back-testing for algorithmic trading, GPS location-based marketing technologies, etc.

Developed the first successful directed graph (hyperlink) analysis system for relevance ranking of Internet web pages – resulted in early patents (dating to 1993) subsequently licensed to all the major Internet search engines in the United States.

Currently teaching graduate level Engineering Management courses at Duke in
    Data Mining, Optimization, and Data Visualization &New Opportunities in Big Data

    Computational Finance & Advanced Financial Engineering.

Have also taught undergraduate and graduate Information Technology Entrepreneurship capstone courses as part-time Entrepreneur in Residence at Duke (2003-2008), and a Venture Law course at the University of Florida Law School. Served on the Online Learning committee of the North Carolina School of Science and Mathematics (Magnet public boarding high school).

Representative CQM consulting clients while at Duke include:
    Lazard Asset Management and Russell Investments – cross-sectional dispersion and its relation to active manager excess returns
    AdviseStream (Duke spinout acquired by Kaplan in 2014) – developed predictive analytics for medical school applicants to forecast their personal acceptance probabilities.
    InvestiQuant – currently testing novel models for algorithmic trading of equity futures.
    Collaboration with the Information Initiative at Duke (led by Robert Calderbank) to use automated eye-tracking data to diagnose developmental problems in young children.

Co-sponsor and mentor for the Annual Data-fest Competition at Duke (led by Mine Çetinkaya-Rundel who teaches the Coursera course "Data Analysis and Statistical Inference).

Founder, and Member of the Board of Directors of the non-profit Autoimmune Encephalitis Alliance – https://aealliance.org. My own son's experience with this illness is documented at https://lightofsasha.wordpress.com/.

Undergraduate and Law degrees from Yale University. One year graduate study in Mathematical Logic, Ludwig Maximilian University, Munich Germany.

# Curriculum Vitae
## Ezra Miller

| | | | |
|---|---|---|---|
| Office Address: | Duke University | | |
| | Mathematics Department | Office Phone: | (919) 660–2846 |
| | Box 90320 | Email Address: | ezra@math.duke.edu |
| | Durham, NC  27708–0320 | Homepage: | http://math.duke.edu/~ezra |
| Date of Birth: | 1974 (Maryland, USA) | Date of CV: | January 2016 |

### Education/Employment

| | | |
|---|---|---|
| 2015 – | | Professor of Statistical Science, Duke University |
| 2009 – | | Professor of Mathematics, Duke University |
| 2006 – 2008 | | Associate Professor of Mathematics, University of Minnesota |
| 2002 – 2006 | | Assistant Professor of Mathematics, University of Minnesota |
| 2002 – 2003 | | MSRI Postdoc, Mathematical Sciences Research Institute |
| 2000 – 2002 | | NSF Postdoc, Massachusetts Institute of Technology (mentor: Richard Stanley) |
| 2000 | Ph.D. | University of California, Berkeley, Mathematics (advisor: Bernd Sturmfels) |
| 1995 | Sc.B. | Brown University, Providence, RI, Mathematics (with Honors) |
| 1995 | A.B. | Brown University, Providence, RI, Music |

### Appointments / Visiting positions

| | |
|---|---|
| 2015 – 2016 | Faculty Fellow, SAMSI program on Computational Neuroscience |
| 2011 – 2014 | Associate Director, SAMSI (Statistical and Applied Mathematical Sciences Institute) |
| 2010 – 2011 | Faculty Fellow, SAMSI program on Analysis of Object Data |
| 2007 – 2008 | Member, Institute for Math and its Applications (IMA): Molecular and Cell Biology |
| 2006 – 2007 | Visiting Faculty, University of Michigan |

### Scientific/Academic honors, grants (see 'Organizing activities' & 'University service' for more grants)

| | |
|---|---|
| 2014 | Duke SciComm Fellow (Duke Science & Society program on science communication) |
| 2012 | Fellow of the American Mathematical Society |
| 2012 – 2017 | NSF Grant DMS-1127914 ($17,525,600 to fund SAMSI), co-PI: Richard Smith (UNC-CH Stat), Pierre Gremaud (NCSU Math), Ilse Ipsen (NCSU Math), Alan Karr (NISS) |
| 2010 – 2015 | NSF Grant DMS-1001437 (5 years, $415,804) |
| 2010 – | NSF & NSA conference grants (6 grants, $100,000 total) |
| 2007 – 2009 | University of Minnesota McKnight Presidential Fellow (comes with $45,000 grant) |
| 2007 | U of M Institute of Technology Guillermo E. Borja Award (comes with $3,500 grant) |
| 2005 – 2010 | NSF Faculty Early Career Development (CAREER) Award (5 years, $400,860 grant) |
| 2005 – 2007 | Univ. of Minnesota McKnight Land-Grant Professorship (comes with $90,000 grant) |
| 2003 – 2006 | National Science Foundation Grant (3 years, $126,385) |
| 2000 – 2002 | National Science Foundation Postdoctoral Research Fellowship |
| 1999 – 2000 | Alfred P. Sloan Doctoral Dissertation Fellowship |
| 1997 , 1999 | Julia B. Robinson Fellowship  (Berkeley Math Department) |
| 1997 | Charles B. Morrey, Jr. Award (Berkeley Math Department) |
| 1995 | David Howell Premium for Outstanding Achievement in Mathematics (Brown University math department, highest honor) |
| —— | Phi Beta Kappa  (nationwide honor awarded by Brown University chapter) |
| —— | Sigma Xi      (nationwide honor awarded by Brown University chapter) |
| 1994 | Manning Calculus Prize (Brown University) |
| 1993 | Intern for Research and Training, National Institutes of Health [NIH]* |
| 1991 – 1992 | Howard Hughes Medical Institute Scholar* |
| 1990 | Sobel Scholar* |

### Nonscientific academic awards

| | |
|---|---|
| 1995 | Arlan R. Coolidge Premium for Musical Excellence (Brown University Music Dept.) |
| 1994 | Buxtehude Premium for Excellence in Music (Brown University Music Dept.) |

### Research interests

Geometry, algebra, combinatorics, algorithms, probability, statistics, biology and other applications

---

*Laboratory biochemistry researcher chosen from nationwide search to work summers at the NIH in Bethesda, Maryland

## Publications

### Books and expository articles

accepted 10. *Fruit flies and moduli: interactions between biology and mathematics*, Notices of the American Mathematical Society, 1178–1184. doi:10.1090/noti1290  arXiv:q-bio.QM/1508.05381

2011  9. *Theory and applications of lattice point methods for binomial ideals*, in Combinatorial Aspects of Commutative Algebra and Algebraic Geometry, Proceedings of Abel Symposium held at Voss, Norway, 1–4 June 2009, Abel Symposia, vol. 6, Springer Berlin Heidelberg, 2011, pp. 99–154.

2009  8. (book edited with Viviana Ene) *Combinatorial aspects of commutative algebra*, Exploratory workshop on combinatorial commutative algebra and computer algebra, Mangalia, Romania, May 29–31, 2008, Contemporary Math, vol. 502, AMS, Providence, RI, 2009. (vii+184 pages)

2008  7. *What is . . . a toric variety?*  Notices of the American Mathematical Society **55** (2008), no. 5 (May), 586–587.

2007  6. (book with Srikanth Iyengar, Graham Leuschke, Anton Leykin, Claudia Miller, Anurag Singh, and Uli Walther) *Twenty-four hours of local cohomology*, Graduate Studies in Mathematics, vol. 87, American Mathematical Society, Providence, RI, 2007. (xvi+282 pages)

5. (book edited with Vic Reiner and Bernd Sturmfels) *Geometric Combinatorics*. Lectures from the Graduate Summer School of the Institute for Advanced Study/Park City Mathematics Institute, Park City, UT, July 11–31, 2004. IAS/Park City Math Series, vol. 13, American Math Society, Providence, RI; Institute for Advanced Study (IAS), Princeton, NJ, 2007. (xvi+691 pages)

4. (with Vic Reiner) *What is geometric combinatorics?* In Geometric combinatorics (Park City, UT, 2004), IAS/Park City Math. Series, vol. 13, American Math. Soc., Providence, RI, pp. 1–17.

2004  3. (book with Bernd Sturmfels) *Combinatorial Commutative Algebra*, Graduate Texts in Mathematics, vol. 227, Springer–Verlag, New York, 2004. (xiv+417 pages)

2. *Hilbert schemes of points in the plane*, Appendix to *Commutative algebra of N points in the plane*, by Mark Haiman, in Luchezar Avramov et al., (eds.), *Trends in Commutative Algebra*, MSRI Publications Vol. 51, Cambridge University Press, New York, 2004, pp. 153–180.

2001  1. (with David Perkinson) *Eight lectures on monomial ideals*, in Queen's Papers in Pure and Applied Mathematics, no. 120, 3–105. [Not in MathSciNet—I know not why; I've tried to fix it.]

### Peer-reviewed research articles

accepted 45. (with Thomas Kahle and Chris O'Neill) *Irreducible decomposition of binomial ideals*, Compositio Mathematica, 15 pages. arXiv:math.AC/1503.02607

44. (with Paul Bendich, Steve Marron, Alex Pieloch, and Sean Skwerer) *Persistent homology analysis of brain artery trees*, Annals of Applied Statistics, 18 pages. arXiv:stat.AP/1411.6652

2015 43. (with Stephan Huckemann, Jonathan C. Mattingly, and James Nolen) *Sticky central limit theorems at isolated hyperbolic planar singularities*, Electronic J. Probability **20** (2015), no. 78, 1–34. doi:10.1214/EJP.v20-3887  arXiv:math.PR/1410.6879

42. (with Christine Berkesch and Stephen Griffeth) *Systems of parameters and holonomicity of hypergeometric systems*, Pacific Journal of Mathematics, **276** (2015), no. 2, 281–286. doi:10.2140/pjm.2015.276.281  arXiv:math.AG/1302.0048

41. (with Megan Owen and Scott Provan) *Polyhedral computational geometry for averaging metric phylogenetic trees*, Advances in Applied Math. **15** (2015), 51–91. doi:10.1016/j.aam.2015.04.002  arXiv:math.MG/1211.7046

2014 40. (with Thomas Kahle) *Decompositions of commutative monoid congruences and binomial ideals*, Algebra & Number Theory **8** (2014), no. 6, 1297–1364. doi:10.2140/ant.2014.8-6  arXiv:math.AC/1107.4699v5

39. (with Manoj Gopalkrishnan and Anne Shiu) *A geometric approach to the Global Attractor Conjecture*, SIAM J. on Applied Dynamical Systems (SIADS) **13** (2014), no. 2, 758–797. doi: 10.1137/130928170  arXiv:math.DS/1305.5303

38. Sean Skwerer, Elizabeth Bullitt, Stephan Huckemann, Ezra Miller, Ipek Oguz, Megan Owen, Vic Patrangenaru, Scott Provan, and J.S. Marron, *Tree-oriented analysis of brain artery structure*, Journal of Mathematical Imaging and Vision, 18 pages. doi:10.1007/s10851-013-0473-0

2013 37. (with Thomas Hotz, Stephan Huckemann, Huiling Le, J. S. Marron, Jonathan C. Mattingly, James Nolen, Megan Owen, Vic Patrangenaru, and Sean Skwerer) *Sticky central limit theorems on open books*, Ann. of Applied Probability **23** (2013), no. 6, 2238–2258. doi:10.1214/12-AAP899 arXiv:math.PR/1202.4267

36. (with Manoj Gopalkrishnan and Anne Shiu) *A projection argument for differential inclusions, with applications to persistence of mass-action kinetics*, SIGMA [Symmetry, Integrability, and Geometry: Methods and Applications] **9** (2013), 025, 25 pages. doi:10.3842/SIGMA.2013.025 arXiv:math.DS/1208.0874

35. *Affine stratifications from finite misère quotients*, Journal of Algebraic Combinatorics **37** (2013), 1–9. doi:10.1007/s10801-012-0355-3  arXiv:math.CO/1009.2199

34. (with Alan Guo) *Algorithms for lattice games*, International Journal of Game Theory **42** (2013), no. 4, 777–788. doi:10.1007/s00182-012-0319-9  arXiv:math.CO/1105.5413

2012 (on parental leave)

2011 33. (with Isabella Novik and Ed Swartz) *Face rings of simplicial complexes with singularities*, Mathematische Annalen **351** (2011), 857–875. arXiv:math.AC/1001.2812

32. (with Alan Guo) *Lattice point methods for combinatorial games*, Advances in Applied Mathematics **46** (2011), 363–378. doi:10.1016/j.aam.2010.10.004  arXiv:math.CO/0908.3473; Corrigendum: ibid. **48** (2012), 269–271. arXiv:math.CO/1105.5420

31. (with Dave Anderson and Stephen Griffeth) *Positivity and Kleiman transversality in equivariant K-theory of homogeneous spaces*, J. European Math Society **13** (2011), 57–84. doi:10.4171/ JEMS/244  arXiv:math.AG/0808.2785

2010 30. (with Alicia Dickenstein and Laura Matusevich) *Binomial D-modules*, Duke Mathematical Journal **151**, no. 3 (2010), 385–429. arXiv:math.AG/0610353

29. (with Alicia Dickenstein and Laura Matusevich) *Combinatorics of binomial primary decomposition*, Mathematische Zeitschrift **264**, no. 4 (2010), 745–763. doi:10.1007/s00209-009-0487-x arXiv:math.AC/0803.3846

2009 28. *Topological Cohen–Macaulay criteria for monomial ideals*, in #8 listed above under Books and expository items, pp. 137–156. arXiv:math.AC/0809.1458

27. (with Allen Knutson and Alex Yong) *Gröbner geometry of vertex decompositions and of flagged tableaux*, J. für die reine und angewandte Mathematik **630** (2009), 1–31. arXiv:math.CO/0502144

2008 26. (with Allen Knutson and Alex Yong) *Tableau complexes*, Israel Journal of Mathematics **163** (2008), 317–343. arXiv:math.CO/0510487

25. (with Shin-Yao Jow) *Multiplier ideals of sums via cellular resolutions*, Mathematical Research Letters **15** (2008), no. 2, 359-373. arXiv:math.AG/0703299

24. (with Ning Jia) *Duality of antidiagonals and pipe dreams*, Séminaire Lotharingien de Combinatoire, B58e [the fifth paper in Issue 58] (2008), 6 pages. arXiv:math.CO/0706.3031

23. (with Igor Pak) *Metric combinatorics of convex polyhedra: cut loci and nonoverlapping unfoldings*, Discrete and Computational Geometry **39** (2008), no. 1–3, 339–388. doi:10.1007/s00454-006-1249-0  arXiv:math.MG/0312253

22. (with David Speyer) *A Kleiman–Bertini theorem for sheaf tensor products*, Journal of Algebraic Geometry **17** (2008), no. 2, 335–340. arXiv:math.AG/0601202

2007 (see Books and expository articles, above)

2006 21. (with Allen Knutson and Mark Shimozono) *Four positive formulae for type A quiver polynomials*, Inventiones Mathematicae **166** (2006), no. 2, 229–325. arXiv:math.AG/0308142

20. (with Laura Matusevich) *Combinatorics of rank jumps in simplicial hypergeometric systems*, Proceedings of the American Math Society **134** (2006), 1375–1381. arXiv:math.AC/0402071

19. (with Vic Reiner) *Stanley's simplicial poset conjecture, after M. Masuda*, Communications in Algebra **34** (2006), no. 3, 1049–1053.

2005 18. (with Laura Matusevich and Uli Walther) *Homological methods for hypergeometric families*, Journal of the American Math Society **18** (2005), no. 4, 919–941. arXiv:math.AG/0406383

17. *Alternating formulas for K-theoretic quiver polynomials*, Duke Math Journal **128** (2005), 1–17. arXiv:math.CO/0312250

16. (with Allen Knutson) *Gröbner geometry of Schubert polynomials*, Annals of Mathematics **161** (2005), 1245–1318. arXiv:math.AG/0110058

15. (with David Helm) *Algorithms for graded injective resolutions and local cohomology over semigroup rings*, Journal of Symbolic Computation **39** (2005), 373–395. arXiv:math.AC/0309256

14. (with Mikhail Kogan) *Toric degeneration of Schubert varieties and Gelfand–Tsetlin polytopes*, Advances in Mathematics **193** (2005), no. 1, 1–17. arXiv:math.AG/0303208

2004 13. (with Vic Reiner) *Reciprocal domains and Cohen–Macaulay d-complexes in $\mathbb{R}^d$*, The Electronic Journal of Combinatorics **11(2)** (2004), #N1 (9 pages). arXiv:math.CO/0408169

12. (with Allen Knutson) *Subword complexes in Coxeter groups*, Advances in Mathematics **184** (2004), 161–176. arXiv:math.CO/0309259

2003 11. (with David Helm) *Bass numbers of semigroup-graded local cohomology*, Pacific Journal of Mathematics **209** (2003), no. 1, 41–66. arXiv:math.AG/0010003

10. *Mitosis recursion for coefficients of Schubert polynomials*, Journal of Combinatorial Theory, Series A **103** (2003), 223–235. arXiv:math.CO/0212131

2002 9. *Cohen–Macaulay quotients of normal semigroup rings via irreducible resolutions*, Mathematical Research Letters **9** (2002), no. 1, 117–128. arXiv:math.AC/0110096

8. *Graded Greenlees–May duality and the Čech hull*, Local cohomology and its applications (Guanajuato, 1999), Lecture Notes in Pure and Appl. Math., vol. 226, Dekker, New York, 233–253.

7. *Planar graphs as minimal resolutions of trivariate monomial ideals*, Documenta Mathematica **7** (2002), 43–90. (electronically published: http://www.math.uiuc.edu/documenta/vol-07/03.html)

2000 6. *Resolutions and duality for monomial ideals*, Ph.D. thesis, University of California at Berkeley.

5. *The Alexander duality functors and local duality with monomial support*, Journal of Algebra **231** (2000), 180–234.

4. (with Bernd Sturmfels and Kohji Yanagawa) *Generic and cogeneric monomial ideals*, Journal of Symbolic Computation **29** (2000), 691–708.

3. *Icosahedra constructed from congruent triangles*, Discrete and Computational Geometry **24** (2000), no. 2–3, 437–451.

1999 2. (with Bernd Sturmfels) *Monomial ideals and planar graphs*, in Applied Algebra, Algebraic Algorithms and Error-Correcting Codes, [M. Fossorier, H. Imai, S. Lin and A. Poli, eds.], Proceedings of AAECC-13 (Honolulu, Nov. 1999), Springer Lecture Notes in Computer Science **1719**, 19–28.

1998 1. *Multiplicities of ideals in noetherian rings*. Beiträge zur Alg. und Geom. **39** (1998), no. 1, 47–51.

Conference abstracts (peer-reviewed and/or invited)

2014 7. (with Stephan Huckemann, Jonathan Mattingly, and James Nolen) *Topological definition of stickiness for means in arbitrary metric spaces*, in *Mini-Workshop: Asymptotic Statistics on Stratified Spaces*, abstracts from 29 September – 3 October 2014, organized by Aasa Feragen, Stephan Huckemann, J.S. Marron, and Ezra Miller, Oberwolfach rep. no. 44 (2014), p.2510. doi:10.4171/OWR/2014/44

2009 6. (with Alan Guo and Mike Weimerskirch) *Potential applications of commutative algebra to combinatorial game theory*, in *Kommutative Algebra*, abstracts from April 19–25, 2009 workshop, organized by W. Bruns, H. Flenner, and C. Huneke, Oberwolfach rep. **6**, no. 2 (2009), pp. 1180–1183.

2007 5. (with Alicia Dickenstein and Laura Matusevich) *Extended abstract: Binomial D-modules*, Proceedings MEGA (Effective Methods in Algebraic Geometry), Strobl, Austria, 2007, 13 pages. http://www.ricam.oeaw.ac.at/mega2007/electronic/electronic.html

4. (with Shin-Yao Jow) *Extended abstract: Cellular resolutions of multiplier ideals of sums*, in *Topological and geometric combinatorics*, abstracts from the Jan. 28–Feb. 3, 2007 workshop, organized by A. Björner, G. Kalai, and G. Ziegler, Oberwolfach rep. **4** (2007), no. 1, 195–271, pp.??–?? (3 pages).

2006 3. (with Laura Matusevich and Uli Walther) *Extended abstract: Homological methods for hypergeometric families*, in *Convex and algebraic geometry*, abstracts from the workshop held January 29–February 4, 2006, organized by K. Altmann, V. Batyrev, and B. Teissier, Oberwolfach reports **3** (2006), no. 1, 253–316, pp.??–?? (3 pages).

2003 2. (with David Helm) *Extended abstract: Algorithms for graded injective resolutions and local cohomology over semigroup rings*, Proceedings MEGA (Effective Methods in Algebraic Geometry), Kaiserslautern, Germany, 2003, 5 pages.

2002 1. (with Allen Knutson) *Extended abstract: Gröbner geometry of Schubert polynomials*, Proceedings FPSAC (Formal Power Series and Algebraic Combinatorics), Melbourne 2002, 10 pages.

Other publications

2009 1. *Alexander duality for monomial ideals and their resolutions*, Rejecta Mathematica **1** (2009), no. 1, 18–57. arXiv:math.AC/9812095

**Lecture series, lectures, and presentations (193 total)**

Lecture series

2010 Sep. İstanbul, Turkey: Decompositions of monoids and binomial ideals [5 hours of lectures]
—— Jan. Sevilla, Spain: Primary decomposition of binomial ideals [8 hours of lectures]
2009 June (canceled due to political unrest) IPM Tehran, Iran: Binomial ideals [4 hours of lectures]
2007 May CRM Montréal, Canada: Hypergeometric series and binomial ideals [5 hours of lectures]
2006 Sep. Constanţa, Romania: Multigraded commutative algebra [9 hours of lectures]
—— July Lincoln, Nebraska (IMMERSE program): Irreducible decomposition [3 hours of lectures]
2005 Sep. UNAM, Mexico City: Gröbner geometry of quiver polynomials [3 hours of lectures]
—— June Snowbird, Utah: Local cohomology and combinatorics [3 hours of lectures]
2004 Aug. Kyoto, Japan (RIMS workshop): Gröbner geometry of quiver polynomials [3 hours of lectures]
—— May Trieste, Italy (ICTP workshop): Combinatorial commutative algebra [4 hours of lectures]
1998 Aug. Berkeley, CA (MSRI Summer Program): Alexander duality [3 hours of lectures]

Invited (international audience)

2015 Sep. Oxford, UK: Biological apps of persistent homology: interactions with stat, geom, and alg
—— July Osaka, Japan: Multigraded moduli from fruit flies via stratified space statistics

—— June  U. Chicago: Persistent interactions between biology, topology, and statistics
2014 Oct.  Oberwolfach, Germany: Topological definition of sticky means in arbitrary metric spaces
—— May   Chicago, IL (Alg. Stat @ IIT): Applying persistent homology to brain artery and vein imaging
—— May   Durham, NC (SAMSI LDHD Transition): Topological analysis of geometrically stratified data
—— Mar.  Göttingen, Germany: Applying persistent homology to brain artery imaging
—— Mar.  Ulm, Germany: Asymptotics of sampling from topologically stratified spaces
—— Jan.   Halifax, NS (Comb.Alg/Alg.Comb): Binomial irreducible decomposition
2013 July  Luminy, France (CIRM): Binomial irreducible decomposition.
—— Apr.   Boston, MA: How do quivers and stratifications apply to biology?
2012 Dec  Berkeley (MSRI): Alexander duality and total positivity: a cluster/commutative connection
—— Oct.   Sandbjerg, Denmark: Geometry for samples of metric branched structures
—— May   Columbus, Ohio (MBI): Sticky central limit theorems at singularities
2011 Aug.  Prague, Czech Rep.: Sticky central limit theorems at singularities
—— Aug.   Miami Beach (JSM): Sticky central limit theorems at singularities
—— Jan.   Oslo, Norway: Statistics on data sampled from stratified spaces
2010 Nov.  Luminy, France: Decompositions of commutative monoid congruences and binomial ideals
—— June  Berkeley (AMS–SMM meeting): Face rings of simplicial complexes with singularities
2009 July  Halifax, Nova Scotia (Games-at-Dal): Lattice games
—— June  Voss, Norway (Abel Symposium): Applications of binomial commutative algebra
—— Apr.   Oberwolfach, Germany: Potential apps. of commutative algebra to combinatorial game theory
—— Mar.  MSRI, Berkeley: Equivariant transversality and $K$-theoretic positivity
—— Mar.  Thessaloniki, Greece: Geometry of flags and permutations
—— Mar.  Thessaloniki, Greece: Unfolding polyhedra
—— Mar.  Ioannina, Greece: Unfolding polyhedra
2008 May   Mangalia, Romania: Cellular resolutions of multiplier ideals of sums
2007 Dec  American Institute of Mathematics, Palo Alto: Nonoverlapping unfolding of polyhedra
—— July   Medellín, Colombia: Old hypergeometric mysteries and new toric algebra
—— July   Tianjin, China (FPSAC plenary speaker): Combinatorics of Horn hypergeometric series
—— May   CRM Montréal, Canada: Kleiman–Bertini theorems for sheaf tensor products
—— Feb.   Oberwolfach, Germany: Cellular resolutions of multiplier ideals of sums
2006 June  Snowbird, Utah: Binomial $D$-modules and lattice point geometry
—— May   Luminy, France: Lattice basis ideals and Horn systems
—— Feb.   Oberwolfach, Germany: Homological methods for hypergeometric families
2005 Dec.  Taichung, Taiwan (AMS–TMS): Positivity by degeneration
—— July   Seattle (Algebraic Geometry Warmup Week): Combinatorial positivity in algebraic geometry
—— July   Lisbon, Portugal (workshop on D-modules): Homological methods for hypergeometric families
2004 July  Park City, Utah (IAS/PCMI): Overview of geometric combinatorics
2003 June  Seville, Spain (AMS–RSME meeting): Positive combinatorial formulae for quiver polynomials
—— June  Kaiserslautern, Germany (MEGA): Computing injective resolutions over semigroup rings
—— May   Banff, Canada (BIRS workshop): Positive combinatorial formulas for quiver polynomials
—— Apr.   Oberwolfach, Germany (Topological Combinatorics workshop): Unfolding polyhedra
—— Apr.   Banff, Canada (BIRS workshop): Minors in products of matrices
2002 Sep.  MSRI Introductory workshop (Berkeley): Hilbert schemes of points in the plane
—— July   Melbourne, Australia (FPSAC one-hour talk): Gröbner geometry of Schubert polynomials
—— June  Trento, Italy (workshop): Gröbner geometry of Schubert polynomials via determinantal ideals
2001 May   Morelia, Mexico (AMS–SMM meeting): Embedding planar graphs in staircases
1999 Dec.  Guanajuato, Mexico (Local Cohomology workshop): The Čech hull
—— June  Essen, Germany (special seminar): Alexander duality & local duality with monomial support
1998 Apr.  Kyoto, Japan (RIMS): Alexander duality for arbitrary monomial ideals

Invited (domestic audience; not including seminars)

| | | |
|---|---|---|
| 2015 | Dec. | U. Minnesota (math colloquium): Interactions between topology, biology, and imaging |
| —— | Nov. | Texas A&M (Math Colloquium): Lessons in topology from statistics of brain arteries |
| —— | Sep. | UNC-Greensboro (Helen Barton Lecture Series): Persistent homology for biological images |
| —— | Mar. | NC State (Grad Math Wkshp): Applying persistent homology to brain artery and vein imaging |
| —— | Jan. | San Antonio (JMM): Persistent homology analysis of brain artery trees |
| 2014 | Nov. | U. Georgia (Stat & Math): Applying persistent homology to brain artery and vein imaging |
| —— | Jun. | Winthrop U.: Topology for statistical analysis of brain artery images |
| 2013 | Nov. | Columbia, SC (CommAlg in the SE): Binomial irreducible decomposition |
| —— | Oct. | Philadelphia, PA (AMS meeting): Binomial irreducible decomposition |
| —— | Mar. | U. Arizona (math colloquium): Geometric statistics on stratified spaces |
| —— | Feb. | UNC-CH (math colloquium): Biological applications of geometric statistics on stratified spaces |
| —— | Feb. | NC State (SUM Series): Statistics without addition or division |
| 2011 | Sep. | Wake Forest (AMS): Subword complexes and fibers of totally nonnegative parametrizations |
| —— | June | MSRI: How primary decomposition in commutative monoids is wrong |
| —— | May | SAMSI: Stratified statistics for evolutionary biology |
| —— | Apr. | Rice U.: Contractibility via subword complexes in the context of homogeneous spaces |
| —— | Mar. | Columbia, SC: How primary decomposition in commutative monoids is wrong |
| —— | Mar. | Statesboro, GA (AMS meeting, 1-hour): Sticky central limit theorems on polyhedral spaces |
| —— | Feb. | George Mason (math colloquium): Unfolding polyhedra |
| 2010 | July | NSA (math colloquium): Unfolding convex polyhedra |
| —— | May | U. Maryland (math colloquium): Unfolding convex polyhedra |
| —— | Apr. | Texas State (CombinaTexas): Lattice point methods for combinatorial games |
| —— | Mar. | George Mason (WCEIT): Lattice point methods for combinatorial games |
| 2009 | Oct. | Raleigh, NC (SE Lie Theory): Equivariant transversality and $K$-theoretic positivity |
| —— | Sep. | NC State (SUM Series): Unfolding polyhedra |
| —— | Apr. | Raleigh, NC (AMS meeting): Equivariant positivity via transversality in $K$-theory of $G/P$ |
| —— | Apr. | Raleigh, NC (AMS meeting): Topological Cohen–Macaulay criteria for monomial ideals |
| —— | Apr. | Raleigh, NC (AMS meeting): Combinatorial primary decomposition of binomial ideals |
| —— | Jan. | UNC Chapel Hill (math colloquium): Positivity from transversality and transitivity |
| 2008 | Dec. | U. Minnesota (math colloquium): Transversality in homogeneous spaces |
| —— | Feb. | UC San Diego: Metric geometry and unfoldings of polyhedra |
| —— | Feb. | UT Austin: Gröbner geometry of Schubert polynomials |
| —— | Jan. | Duke U.: Metric geometry and unfoldings of polyhedra |
| 2007 | Oct. | Rutgers U. (AMS meeting): Cellular resolutions of multiplier ideals of sums |
| —— | Apr. | U. Nebraska (math colloquium): Unfolding polyhedra |
| —— | Apr. | U. Kansas (math colloquium): Unfolding polyhedra |
| 2006 | Dec. | Indiana U. (math colloquium): Combinatorics in convexity, cohomology, and complex analysis |
| —— | Oct. | Cincinnati, Ohio (AMS meeting plenary speaker): Unfolding polyhedra |
| —— | Oct. | Ann Arbor, Michigan (Ohio State–UMichigan Alg. Geom. Workshop): Binomial schemes |
| —— | Mar. | ITCEP Family Fun Fair (interactive middle & high school presentation): Unfolding polyhedra |
| 2005 | Dec. | Reed College (math colloquium): Unfolding polyhedra |
| —— | Nov. | U. Washington (math colloquium): Unfolding polyhedra |
| —— | Nov. | Eugene, OR (AMS meeting): Duality of antidiagonals and pipe dreams |
| —— | Oct. | Lincoln, NE (AMS meeting): Encoding injective resolutions |
| —— | Mar. | U. Texas, Austin (math colloquium): Combinatorics in cohomology and convexity |
| —— | Jan. | Columbia U. (math colloquium): Combinatorics from geometry |
| 2004 | Sep. | Michigan State (math colloquium): Combinatorial positivity by geometric degeneration |

——— Apr.   Durham, NC (Duke Math Journal conf.): Combinatorial positivity by geometric degeneration
2003 Oct.   Binghamton, NY (AMS meeting): Unfolding convex polyhedral manifolds
——— Apr.   Western Alg. Geom. Seminar (at Stanford): Positivity of universal cohomological formulae
——— Mar.   Bay Area Discrete Math Day (UC Davis): Unfolding polyhedra in many dimenions
——— Feb.   UC Berkeley (math colloquium): Combinatorial positivity by geometric degeneration
2002 Nov.   Lubbock, TX (Red Raider Symposium): Combinatorial positivity by orbit degeneration
——— Oct.   Northeastern U. (AMS meeting): Positivity of quiver cycles via deformation
——— Oct.   Berkeley, CA (MSRI workshop): Positivity via Gröbner degeneration
——— Apr.   SUNY Albany (math colloquium): Determinants, permutations, and flags
——— Jan.   San Diego (AMS meeting): Determinantal ideals and combinatorics of Schubert polynomials
2001 Sep.   George Wash. (math colloq.): Geometry and combinatorics of flag manifolds via Gröbner bases
2000 Sep.   Toronto, Canada (AMS meeting): Gröbner geometry of formulae for Schubert polynomials
——— Jan.   Brown U. (math colloquium): Monomial ideals and duality
1999 Nov.   New Mexico State (math colloq.): Equivariant $K$-theory of flag manifolds and Gröbner bases
——— Sep.   U. Utah (AMS meeting): The Alexander duality functors
——— Sep.   U. Kansas (math colloquium): Resolutions of monomial ideals
——— Feb.   Reed College (math colloquium): Algebra and topology with alphabet blocks

Seminars and other specialized topics talks
2015 Dec.   U. Minnesota (Commutative Algebra): Irreducible decomposition of binomial ideals
——— Nov.   NC State (Algebra & Combinatorics): Multigraded modules from fruit flies
2015 Nov.   Texas A&M (Algebra & Combinatorics): Multigraded modules from fruit flies
——— Nov.   Clemson (AG/NT + Applied Discrete Math): Multigraded moduli from fruit flies
——— Jan.   Duke (MathBio Colloquium): Applying persistent homology for brain artery and vein imaging
2013 Mar.   U. Arizona (Probability): Sticky central limit theorems at singularities
2012 May   Columbus, OH (OSU Combinatorics): Combinatorics and topology of fibers in total positivity
——— Apr.   Duke U. Data Seminar: Sticky central limit theorems at singularities
2011 Mar.   Duke U. Recruitment: Geometric statistics: sampling from singular spaces
2010 Dec.   Duke U. MathBio: Statistics on spaces of phylogenetic trees
——— Dec.   Duke U. Brown Bag: Combinatorics of phylogenetic trees
——— May   MIT Combinatorics: Lattice point methods for combinatorial games
——— Feb.   Duke U. Algorithms: Unfolding convex polyhedra
2009 Dec.   Duke U. Algebraic Geometry: $K$-theory of toric varieties
——— Oct.   Duke U. (CTMS): Biological and chemical applications of polyhedral geometry and algebra
——— Sep.   Duke U. Algebraic Geometry: Combinatorics in cohomology of toric varieties
——— Sep.   Duke U. Algebraic Geometry: What is a toric variety?
2008 Oct.   U. Minnesota Schubert seminar: Cohomology rings of flag and toric varieties
——— Sep.   U. Minnesota Combinatorics: Topological Cohen–Macaulay criteria for monomial ideals
——— Apr.   U. Minnesota Comibnatorics: Combinatorics of Horn hypergeometric series
——— Feb.   U. Minnesota Junior Colloquium: Unfolding convex polyhedra
2007 Oct.   U. Minnesota Math Club: Unfolding polyhedra
——— Apr.   Kansas U. Combinatorics: Combinatorics of permutations from determinantal ideals
——— Apr.   U. Nebraska Algebra: Multiplier ideals of sums via cellular resolutions
——— Apr.   Michigan Geom. Rep. Thy: Kazhdan–Lusztig conjecture via intersection cohomology (overview)
2006 Dec.   Purdue Algebraic Geometry: Old mysteries about multivariate hypergeometric systems
——— Nov.   Northeastern GASC: Horn hypergeometric systems and binomial $D$-modules
——— Nov.   MIT Combinatorics: Hypergeometric series and binomial primary decomposition
——— Oct.   U. Michigan Algebraic Geometry: Toric primary decomposition and hypergeometric series
——— Feb.   U. Minnesota Combinatorics: $h$-vectors of Gorenstein polytopes

2005 Nov.  U. Washington Combinatorics: Simplicial complexes whose facets are Young tableaux
—— Nov.  MIT Combinatorics: Tableau complexes
—— Feb.  UC Berkeley Representation theory/geom./combinatorics: Families of hypergeometric systems
—— Feb.  U. Minnesota Combinatorics: Stanley's simplicial poset conjecture
2004 Sep.  U. Michigan Combinatorics: Unfolding polytopes
—— Mar.  U. Michigan Noncommutative algebra: Homological methods for hypergeometric families
—— Feb.  U. Minnesota Topology: Gelfand–Tsetlin patterns via toric degeneration of flag manifolds
2003 Apr.  UC Davis Combinatorics: Combinatorics of quiver polynomials
—— Mar.  Northeastern U. Algebra: Quiver polynomials and Schubert varieties
—— Mar.  MIT Combinatorics: Combinatorics of quiver polynomials
—— Jan.  U. Michigan Algebraic geometry: Positive formulae for quiver polynomials
2002 Sep.  U. Michigan Combinatorics: Gröbner geometry of Schubert polynomials
—— Mar.  Columbia U. Algebraic geometry: Gröbner geometry of flag varieties
—— Mar.  U. Michigan Combinatorics: Subword complexes in Coxeter groups
—— Feb.  U. Minnesota Combinatorics: Words and subwords in symmetric groups
2001 Oct.  Cornell U. Algebra: Determinantal ideals and the combinatorics of Schubert polynomials
—— Sep.  George Wash. U. Algebra: Embedding planar graphs in staircases to resolve monomial ideals
—— May  UC Berkeley special seminar: Planar graphs in 3-dimensional staircases
—— Apr.  Harvard–MIT Algebraic Geometry: An open problem on determinantal loci
—— Feb.  MIT Combinatorics: Subword complexes in Coxeter groups and Schubert varieties
2000 Nov.  UMass Amherst (Valley Geometry Seminar): Gröbner geometry of Schubert polynomials
—— Sep.  MIT Combinatorics: Gröbner geometry of formulae for Schubert polynomials
—— Aug.  UC Berkeley workshop: Haiman's proof of the $n!$ conjecture
—— May  UC Berkeley Algebra: Gröbner bases for determinantal ideals and K-theory of flag manifolds
1999 Nov.  New Mexico State U. Algebra: Injective resolutions
—— Sep.  UC Berkeley Algebra: Projective dimension versus support-regularity
—— Sep.  U. Kansas Algebra: The canonical Čech complex
—— Feb.  UC Berkeley Grad Student Seminar: Combinatorial commutative algebra
—— Feb.  UC Berkeley Algebra: Generic and cogeneric monomial ideals
1998 Sep.  U. Michigan Combinatorics: Cellular homology and duality in the study of monomial ideals
—— Sep.  MIT Combinatorics: Alexander duality for monomial ideals and their resolutions
—— Aug.  UC Berkeley Algebra: The cohull resolution
1997 May  UC Berkeley Algebra: Degrees on graded modules
—— Mar.  Rutgers U. Algebra: Characterization of multiplicity functions of ideals
1996 Nov.  UC Berkeley Abelian Varieties: Multiplying points on group schemes
—— Oct.  UC Berkeley Grad Student Seminar: Where is the geometry in schemes?
—— July  MSRI Summer Program: Gröbner bases over PIDs

## Outreach talks, panels, and other presentations

2015 Oct.  SAMSI (Undergrad Workshop): Topology for statistical analysis of brain artery images
2014 May  Green Hope High School (AP Stat): Topology for statistical analysis of brain artery images
—— Apr.  Duke U. (Noyce Learning Conference): Topology for statistical analysis of brain artery images
2012 Sep  Duke Postdoc Services / School of Medicine panel: NSF Funding from the Inside

**Organizing activities (conferences, meetings, annual programs, public lectures, etc.)**

| | | |
|---|---|---|
| 2016 | Aug. | Director (with Viviana Ene): Summer school on Applicable Combinatorial Commutative Algebra, Romania |
| 2016 | Jul. | Scientific Committee Member (with François Bergeron, Patricia Hersh, Gregor Kemper, Peter Symonds, Jerzy Weyman): Conference on algebraic combinatorics and group actions, Herstmonceux Castle, England |
| 2015 – 2016 | | Local Scientific Coordinator, SAMSI year-long program on Challenges in Computational Neuroscience |
| 2014 | Sep. | Organizer, Oberwolfach mini-workshop: Asymptotic statistics on stratified spaces (with Aasa Feragen, Stephan Huckemann, J. S.Marron) |
| 2014 | Sep. | Advisory Board, Meeting on Combinatorial Commutative Algebra [MOCCA 2014], Trentino, Italy (other board members: Alicia Dickenstein, Winfried Bruns; organizers: Alexandru Constantinescu, Thomas Kahle, Matteo Varbaro) |
| | | NSF DMS-1439356: $12,000 (sole PI) |
| 2014 | May | Organizer, Fields Institute workshop: Geometric topological and graphical model methods in statistics (with Peter Kim and Hélène Massam) |
| 2013 | Dec. | Organizer, SAMSI Public Lecture by Greg Fishel (WRAL-TV Chief Meteorologist): "The changing climate of weather prognostication: the irony of uncertainty leading to better forecasts" |
| 2013 – 2014 | | Directorate Organizer, SAMSI year-long program on Low-Dimensional Structure in High-Dimensional Systems (LDHD) |
| 2013 | June | Scientific Committee Member (with Viviana Ene and others): Experimental and theoretical methods in algebra, geometry, and topology, Mangalia, Romania. |
| 2013 | June | Directorate Organizer, SAMSI Summer Program on Neuroimaging Data Analysis (NDA) |
| 2012 – 2013 | | Organizer, MSRI year-long program on Commutative Algebra (with David Eisenbud, Srikanth Iyengar, Anurag Singh, and Karen Smith). |
| 2012 | May | Lead Organizer, MBI workshop on statistics, geometry, and combinatorics on stratified spaces arising from biological problems (co-organizers: Stephan Huckemann, Huiling Le, Megan Owen, Victor Patrangenaru) |
| 2012 | Feb. | Organizer, Triangle Lectures in Combinatorics (with Sonja Mapes, Christine Berkesch) |
| 2011 | June | Program Committee Member: Formal Power Series and Algebraic Combinatorics (FP-SAC), Reykjavik, Iceland. |
| 2010 – 2011 | | SAMSI Working Group leader: Analysis of data on stratified sample spaces (with Victor Patrangenaru) |
| 2010 – | | Steering Committee member and organizer (with Patricia Hersh, Nathan Reading, and J. Scott Provan): Triangle Lectures in Combinatorics [TLC], central North Carolina. 2014: Seth Sullivant added as co-PI and Gabor Pataki added to replace J. Scott Provan. |
| | | 2014–2016  NSA: $20,000 |
| | | 2014–2016  NSF DMS-1202691: $20,800 |
| | | 2012–2014  NSA: $28,000 |
| | | 2012–2014  NSF DMS-1202691: $8,000 |
| | | 2010–2012  NSF DMS-1000130: $10,600 |
| 2009 | Oct. | Working group session organizer (with William Graham): Combinatorial Lie theory and applications, first installment of Southeastern Lie Theory Workshop Series, Raleigh, NC. |
| 2009 | July | Program Committee Member: Formal Power Series and Algebraic Combinatorics (FP-SAC), Research Institute for Symbolic Computation (RISC), Linz, Austria. |
| 2008 | May | Scientific Committee Member, with Jürgen Herzog, Gerhard Pfister, and Dorin Popescu: Exploratory workshop on combinatorial commutative algebra and computer algebra, Mangalia, Romania. |

| | | |
|---|---|---|
| 2008 | Jan. | Organizer, with Michael Albert, Elwyn Berlekamp, Martin Mueller, Richard Nowakowski, and David Wolfe: Workshop on Combinatorial Game Theory, Banff International Research Station (BIRS), Banff, Canada. |
| 2007 | Dec. | Organizer, with Bob Connelly and Igor Pak: AIM Research Conference Center Workshop on Rigidity and Polyhedra, American Institute of Mathematics, Palo Alto, CA. |
| 2006 | Oct. | Invited organizer, with Igor Pak: Special session on Geometric Combinatorics at the regional American Mathematical Society meeting in Cincinnati, OH. |
| 2006 | Sep. | Organizer, with Mircea Becheanu, Viviana Ene, Cristodor Ionescu, Dorin Popescu, Mirela Ştefănescu: 15th Romanian National School of Algebra, Constanţa, Romania |
| 2005 | Aug. | Invited organizer, with William Fulton: "Algebraic Geometry and Combinatorics" session at the Summer Institute in Algebraic Geometry (Seattle, WA). This international meeting takes place once every ten years; our session along with the two others during the second week (of the three-week meeting) attracted the largest-ever gathering of algebraic geometers anywhere in the world, around 450 participants. |
| 2004 | Oct. | Organizer, with Frank Sottile: Special session on Modern Schubert Calculus at the regional American Mathematical Society meeting in Evanston, IL. Our special session had around 30 participants. |
| 2004 | July | Organizer and Steering Committee member, with Bernd Sturmfels and Victor Reiner: 2004 IAS/Park City Math Institute Summer Session on Geometric Combinatorics. 300 participants range from high school teachers through undergraduate students and faculty. Organizing duties included the Graduate Summer School (100 students) as well as the international Research Program (60 researchers). |
| 2003 | May | Organizer, with Serkan Hoşten: Special session on Combinatorial Commutative Algebra and Algebraic Geometry at the regional American Mathematical Society meeting in San Francisco, CA. Our special session had around 25 participants. |

## Editorial activities

| | |
|---|---|
| 2014 – | Editorial Board Member, Journal of Combinatorial Theory, Series A (JCTA) |
| 2011 – | Editorial Board Member, Discrete Mathematics |
| 2009 – | Editorial Board Member, Advances in Mathematics |
| 2009 – | Editorial Board Member, SIAM Journal on Discrete Mathematics |
| 2008 – | Editorial Board Member, Beiträge zur Algebra und Geometrie |
| 2007 –2010 | Associate Editor, Discrete Mathematics |

## Professional committees

| | |
|---|---|
| 2013 – 2015 | AMS Fellows Selection Committee |
| 2012 – 2015 | AMS Short Course Subcommittee |
| 2012 – 2014 | AMS Southeastern Section Program Committee |
| 2007 | AMS-IMS-SIAM Committee on Joint Summer Research Conferences in the Mathematical Sciences (committee was disbanded after 2007 due to lack of funding) |

## Other service

| | | |
|---|---|---|
| 2014 | Nov | External Review Committee: Discrete Mathematics Doctoral Programme (for Austrian Science Fund) |
| 2010 | Apr | External Review Committee: University of Southern California |

## Advisory boards

2013 – America's Amazing Teen Project http://prezi.com/-y4ofsndc3mg/americas-amazing-teens
Summary: brilliant teenagers across America will upload a short video presenting their
ground-breaking project, research, invention, or technological breakthrough. After an
extensive vetting process, the best and brightest teens will be placed into the competi-
tion. Members from our esteemed advisory board, as well as the online voting public,
will vote for the 12 monthly finalists.
Length of contest: 12 month online campaign
Who can enter: teenagers (13-19 years old) across the country

2013 – Duke U. Noyce Program http://web.duke.edu/MAT/fellowships-robert_noyce_fellows.html
Summary: the program recruits and prepares highly successful STEM graduates to be-
come highly effective middle school science and math teachers. Run by Duke University
Program in Education in collaboration with Durham Public Schools and the STEM de-
partments in Trinity College of Arts and Sciences. See 'University Service' on p. 16 for
NSF grant information.

2014 – Duke U. Middle Grades
Summary: NSF capacity-building grant (funded in summer 2014) designed to explore
the institutionalization of a middle-grades MAT strand for STEM subjects. Other
board members include the Superintendent of Durham Public Schools as well as some
school Principals and teachers, along with a few Arts & Sciences faculty and Program
in Education PIs.

## Other community outreach

2014 Fall Duke Science & Society SciComm Fellow. The program teaches faculty to develop skills
in communicating with the public, the media, and lawmakers by putting those skills
into action; Fall 2014 was the inaugural class of fellows for this competitive program

2014 May Guest lecturer to multiple Green Hope High School AP Statistics classes on "Topology
for statistical analysis of brain artery images". Audience consisted of a few dozen high
school students potentially entering STEM fields

2014 Apr Lecturer to Noyce Learning Conference on "Topology for statistical analysis of brain
artery images". Audience consisted of a dozen middle- and high-school science teachers

2014 Mar WGHP-TV nightly news, interviewed by Bob Buckley regarding the $1 billion Warren
Buffett offered for a perfect NCAA bracket ⟨http://myfox8.com/2014/03/18/odds-of-
winning-warren-buffetts-billion-dollar-ncaa-bracket-challenge⟩

2014 Jan Consultant for LA Times and CBS Evening News: regarding the $1 billion Warren Buf-
fett offered for a perfect NCAA bracket ⟨http://www.latimes.com/business/la-fi-buffett-
basketball-bet-20140122,0,7653962.story#axzz2r8uphpac⟩

2013 – Consultant for WRAL-TV weather: helping Greg Fishel and Nate Johnson with display
of data uncertainty for primetime on-air public consumption

2006 Mar ITCEP Family Fun Fair presenter: Led interactive session for middle school and high
school students on "Unfolding polyhedra". Students built models of polyhedra, learning
about connections with math research and applied problems like robot motion-planning

2004 Mar Judge for the Minnesota Academy of Sciences State Science Fair (high school and junior
high levels)

## Referee and review activities

Journals and AMS publications refereed

Acta Mathematica

Advances in Applied Mathematics

Advances in Mathematics

Algebra & Number Theory

Algorithms

American Journal of Mathematics

American Mathematical Monthly

AMS Graduate Texts in Mathematics

Annals of Combinatorics

Appl. Alg. in Engin. Commun. and Comput.

Asian Journal of Mathematics

Bulletin of the London Math Society

Canadian Mathematical Bulletin

Collectanea Mathematica

Compositio Mathematica

Contemporary Mathematics

Contributions to Discrete Mathematics

Crelle's Journal

Designs, Codes, and Cryptography

Discrete Applied Mathematics

Discrete and Computational Geometry

Discrete Mathematics

Duke Mathematical Journal

Electronic Journal of Combinatorics

Experimental Mathematics

Foundations of Computational Mathematics

Games of No Chance

International Mathematics Research Notices

Illinois Journal of Mathematics

International Journal of Game Theory

International Journal of Mathematics

Inventiones Mathematicae

Israel Journal of Mathematics

Journal of Algebra

Journal of Algebra and its Applications

Journal of Algebraic Combinatorics

Journal of the American Mathematical Society

Journal of Combinatorial Theory, Series A

Journal of Commutative Algebra

Journal of the European Math Society

Journal of Graph Theory

Journal of Mathematical Imaging and Vision

Journal of Mathematics

Journal of Pure and Applied Algebra

Journal of Symbolic Computation

Linear Algebra and its Applications

Math Proceedings of Cambridge Phil. Society

Mathematical Research Letters

Mathematics of Computation

Mathematische Annalen

Mathematische Zeitschrift

Memoirs of the American Mathematical Society

Michigan Mathematical Journal

Proceedings of the American Math Society

Proceedings A of the Royal Society of Edinburgh

PLOS ONE [Public Library of Science]

Quarterly Journal of Pure and Applied Math

Rocky Mountain Journal of Mathematics

Selecta Mathematica

SIAM Journal of Discrete Mathematics

Springer Lecture Notes in Mathematics

Transactions of the American Math Society

Transformation Groups

Conferences refereed

Applied Algebra, Algebraic Algorithms and Error Correcting Codes (AAECC)

Effective Methods in Algebraic Geometry (MEGA)

Formal Power Series and Algebraic Combinatorics (FPSAC)

IEEE Conference on Decision and Control

International Congress on Mathematical Software (ICMS)

International Symposium on Symbolic and Algebraic Computation (ISSAC)

Granting agencies refereed

Simons Foundation

National Science Foundation [NSF] (panels and individual review of conference proposals)

National Security Agency [NSA] Mathematical Sciences Program

Natural Sciences and Engineering Research Council [NSERC], Canada

Miscellaneous review activities

> External evaluator, Discrete Mathematics Doctoral Programme (for Austrian Science Fund, 2014)
> External evaluator, University of Southern California Math Department (2010)
> Referee for Isaac Newton Institute program development
> Referee for various countries' national level prizes or awards
> Consultation for JSAGE (publisher of free open source refereed math software implementations)
> External Ph.D. dissertation review (for Sangjib Kim, student of Roger Howe, Yale University, 2005)
> Math Reviews reviewer

## Teaching awards

2004  Feb.  Award from the "Thank a Teacher" program run by the University of Minnesota Center for Teaching and Learning Services, "In appreciation of [my] teaching style and dedication to helping students learn" Calculus I (Math 1271)

## Teaching activities

Course development at Duke

2011–2014  Math Everywhere (Math 181), with Ingrid Daubechies and Jonathan Mattingly: a course for undergraduate students who might not otherwise take any math courses at all but who want a broad overview of what modern mathematics is about

Courses taught at Duke

2015  Spring  Vector Calculus (Math 222)
2014  Fall  Linear Algebra (Math 221)
2014  Spring  Math Everywhere (Math 181)
2013  Spring  Math Everywhere (Math 181)
2012  Spring  Math Everywhere (Math 181)
2010  Fall  Algebraic geometry (Math 273)
2010  Spring  Commutative algebra (Math 252)
2009  Fall  Rings, modules, and Galois theory (Math 251)
2009  Spring  Current research in algebra: combinatorial commutative algebra (Math 358)

Courses taught at Minnesota

2008  Fall  Honors Abstract Algebra (Math 5285H)
2006  Spring  Calculus of one variable (Math 1271)
2006  Spring  Topics in Combinatorics: Polytopes (Math 8680)
2005  Spring  Graph theory and nonenumerative combinatorics (Math 5707)
2005  Spring  Calculus of one variable (Math 1271)
2004  Fall  Topics in Algebra: Combinatorics, algebra, and geometry of determinants (Math 8300)
2004  Spring  Graph theory and nonenumerative combinatorics (Math 5707)
2003  Fall  Calculus of one variable (Math 1271)
2003  Fall  Calculus of one variable (Math 1271)

Recitations led at MIT (as postdoctoral instructor)

2002  Spring  Differential Equations (18.03)

Recitations led at UC Berkeley (as graduate student instructor)

1996  Fall  Linear algebra (Math 53W: workshop section)
1996  Spring  Multivariable calculus (Math 35)
1995  Fall  Calculus of one variable (Math 1B)

Special tutorials

| | | |
|---|---|---|
| 2005 | July | Mentor, Algebraic Geometry Warmup Workshop (Seattle, WA) |
| 1999 | June | Computer tutorial leader, COCOA VI Summer School (Turin, Italy) |
| 1995 | | Study group facilitator for honors multivariable calculus at Brown University |
| 1995 | | Counselor, PROMYS summer program in number theory at Boston University. Lived in dorm for 6 weeks with high-school student participants, helping with math questions |
| 1994–1995 | | Tutor for Brown University Mathematics Department |

Seminars organized (at Duke unless otherwise specified)

| | | |
|---|---|---|
| 2012 | Spring | Mathematical Biology Colloquium |
| 2009 | Fall | Topics in algebraic geometry: toric varieties |
| 2007–2008 | | Combinatorics seminar (at University of Minnesota) |
| 2006 | Fall | Topics in algebraic geometry seminar (at University of Michigan) |

**Mentoring activities**

Undergraduate students advised (while at Duke)

| | |
|---|---|
| 2014 | Aaron Park, Biomed Engineering / Math, summer research: geometry of lung arteries |
| 2013 | Victoria Cheng, Mathematical biology |
| 2012– | Rowena (Jingxing) Gan, PRUV: Geometry of harmony in Impressionist music |
| 2010 | Ezgi Kantarcı (Bogaziçi University), summer research: Combinatorial Game Theory |
| 2009–2011 | Alan Guo, PRUV: Combinatorial Game Theory [currently MIT Computer Science grad student; was Duke Faculty Scholar, Goldwater honorable mention] |

Undergraduate students advised (while at Minnesota)

| | |
|---|---|
| 2008–2009 | Nathaniel Born (CS), UROP: Implementing a nonoverlapping unfolding algorithm |
| 2007–2008 | Matthew Coudron (honors program; Goldwater scholar): Visualizing flows on polyhedra |
| 2003–2007 | David Molitor (Math and Economics major with Physics minor; Goldwater scholar) |

Ph.D. students advised (at Duke unless othewise specified)

| | |
|---|---|
| 2015– | Ashleigh Thomas |
| 2014– | Dmitry Vagner |
| 2011–2015 | Kangkang Wang (joint w/Aspinwall): Determinant, wall monodromy and spherical functor |
| 2009–2014 | Christopher O'Neill: Monoid congruences, binomial ideals, and their decompositions |
| 2004–2012 | Patrick Byrnes (Minnesota, joint w/Vic Reiner): Structural aspects of differential posets |
| 2005–2015 | Robert Edman (Minnesota, Vic Reiner took over after my departure from UMN): Diameter and coherence of monotonoe path graphs in low corank |

Visiting Ph.D. students mentored

| | |
|---|---|
| 2009 grad | Dumitru Stamate (Bucharest, Romania; Fullbright scholar 2006–2008 at Minnesota) |
| 2008 grad | Susan Sierra (at Michigan) [currently Lecturer at University of Edinburgh] |
| 2008 grad | Raman Sanyal (TU Berlin; mentored at UMN) [currently Juniorprofessor, TU Berlin] |

Postdocs mentored

| | |
|---|---|
| 2015– | Adam Jaeger, SAMSI and Duke Stat |
| 2015– | Justin Allman, Duke Math |
| 2011–2013 | Christine Berkesch, Duke Math [currently Assistant Professor, U. Minnesota] |

2010–2011   Anne Shiu, NSF Postdoc, Duke Math [currently Assistant Professor, Texas A&M]
2010, Fall   Manoj Gopalkrishnan, visitor to Duke Math [currently Faculty member, TIFR-Mumbai]
2009–2011   Sonja Mapes, Duke Math [currently Assistant Professor of the Practice, Notre Dame]
2009–2010   Megan Owen, SAMSI/NCSU Math [currently Assistant Professor of CS, Lehman College]
2006–2009   Stephen Griffeth, Minnesota Math [currently Associate Professor, U. Talca, Chile]
2006–2009   Drew Armstrong, Minnesota Math [currently Associate Prof., U. Miami]
2006–2008   Milena Hering, Minnesota Math [currently Lecturer at U. Edinburgh]
2005–2008   Alex Yong, Minnesota Math [currently Associate Prof., U. Illinois, Urbana-Champaign]
2005–2008   Calin Chindris, Minnesota Math [currently Assistant Prof., U. Missouri]
2003–2005   Jeremy Martin, Minnesota Math [currently Associate Prof., U. Kansas]

## University service

2014 –        Master of Arts in Teaching (MAT) Program: Math Content Advisor
2014 – 2016   co-PI, Robert Noyce Teacher Fellowship Program, an $800,000 initial NSF grant +
              subsequent capacity-building grant (NSF DUE-1439799: $300,000; "Integrative Middle
              School STEM Teacher Preparation: A Collaborative Capacity Building Project at Duke
              University" with Jan Riggsbee, Jack Bookman, Kate Allman, Ronen Plesser), awarded
              through the Master of Arts in Teaching program at Duke; see 'Advisory boards' on
              p. 12 for additional program details

## Departmental service

### Departmental committees at Duke

(see online CV at https://fds.duke.edu/db/aas/math/faculty/ezra/cv.html)

### Departmental committees at Minnesota

Tenure (2007–2008)
Undergraduate curriculum (2005–2006)
Honors Program (2005–2006)
Postdoc hiring (2004–2005)
Instruction evaluation (2004–2005, 2007–2008)
Social activities (2003–2004)

### Departmental committees at Berkeley

Preliminary exam (1999)

### Dissertation committees

Sean Skwerer (2014, UNC-Chapel Hill, Stat/Operations Research)
Ruth Davidson (2014, NC State, Math)
Chris O'Neill (2014, Duke)
John Steenbergen (2013, Duke)
Elizabeth Munch (2013, Duke)
Kaisa Taipale (chair, 2010, Minnesota)
Meghan Allen (2009, Dalhousie University, Halifax, NS, Canada)
Andrew Berget (2009, Minnesota)
Jessica Striker (2008, Minnesota)
Molly Maxwell (chair, 2007, Minnesota)
Ning Jia (2007, Minnesota)

Oral prelim exam committees (Duke Math unless specified)

Brian Fitzpatrick (2013)
Sean Skwerer (2013, UNC Chapel Hill Stat/Operations Research)
Ruth Davidson (2013, NCSU Math)
Ben Gaines (2012)
Kangkang Wang (2012)
Chris O'Neill (2012)
John Steenbergen (2011)
Anil Venkatesh (2011)
Elizabeth Munch (2011)
Robert Edman (2008, Minnesota, Math)
Kaisa Taipale (2007, Minnesota, Math)
Brendon Rhoades (2007, Minnesota, Math)
Patrick Byrnes (2005, Minnesota, Math)
Molly Maxwell (2005, Minnesota, Math)
Jayashree Sadagopan (2005, Minnesota, Computer Science Department)
Sangwook Kim (2005, Minnesota, Math)
Ning Jia (2004, Minnesota, Math)
Dan Drake (2004, Minnesota, Math)

Oral qual committees (Duke Math)

Christopher O'Neill (2011)
Anil Venkatesh (2010)
Benjamin Gaines (2010)
Tatsunari Watanabe (2010)

# Eric E Monson

233B Perkins Library · Box 90175 · Durham, NC 27708-1075
emonson@duke.edu · 919.613.1833 · http://emonson.com

## Employment

**Data Visualization Analyst** (9/2015 – *present*), Duke University Libraries · Data and Visualization Services

Consult and conduct workshops to help students, faculty and staff with data visualization, literacy, organization and processing for their research.

**Research Scientist** (9/2007 – 8/2015), Duke Visualization & Interactive Systems · Computer Sci · Media Arts + Sciences

Scientific and information visualization research, consulting, teaching and software development. Creation of novel 2D and 3D web and desktop visualizations, interfaces and analysis methods for complex data sets. Integration of technology and quantitative methods into humanistic scholarship, including data gathering, cleaning, modeling and visualization.

**Research Scientist** (9/2006 – 8/2007) · **Research Associate** (9/2002 – 8/2006), Duke University Physics with Anna L Lin.

Experimental and numerical studies of the signaling between neurons and glial cells using fluorescence imaging and nonlinear spatio-temporal pattern analysis.

## Education

**Postdoctoral Researcher**, University of Michigan Chemistry (9/1999 – 8/2002) with Raoul Kopelman.

Managed a large *National Cancer Institute, Unconventional Innovations Program* contract for early cancer detection and therapy using targeted, nanoparticle-based MRI contrast and photodynamic therapy agents.

**Ph.D. in Applied Physics**, University of Michigan, Ann Arbor (February 1999).

*Thesis title:* Nano-scale spatial and temporal fluorescence fluctuations in near-field microscopy, photobleaching recovery, and non-classical elementary reaction kinetics.
*Thesis advisor:* Raoul Kopelman, Kasimir Fajans Collegiate Professor of Chemistry, Physics and Applied Physics.

**S.B. in Physics**, Massachusetts Institute of Technology (June 1990).

## Research Interests & Skills

*Research Interests – Current:* Scientific and information visualization, data and visual literacy, dimensionality reduction and dictionary learning techniques, relational data modeling, graphic design, image processing, natural language processing – *Past:* Live-cell fluorescence microscopy, intracellular ion sensors, lasers and optical instrumentation, numerical simulations of neuron, reaction-diffusion and immune systems

*Languages & Software – Proficient:* Python, Matlab, C++, JavaScript, D3.js, OpenRefine, RegEx, Tableau, Excel, Gephi, Palladio, MongoDB, SQL, HDF5, Elasticsearch, VTK/ParaView, MALLET, NLTK, Adobe Illustrator/Photoshop/InDesign, HTML/CSS, Markdown, Git – *Functional:* Bash, Avizo, SketchUp, Filemaker Pro, Apache Solr – *Familiar:* PHP, C#/WPF

## Teaching & Mentorship

DUKE UNIVERSITY

Directed the Visualization & Interactive Systems group (summer 2012 – *present*). This involved direct supervision of the technology and art consultant Sarah Goetz, organization of materials for public events and demonstrations, and running weekly group meetings with researchers from MA+S, Engineering, and Data and Visualization Services.

Advised and mentored Ronak Etemadpour, a visiting graduate student from Jacobs University (Bremen, Germany), during her year of research at Duke (2012 calendar year). Also served on her thesis committee (2013).

Taught and developed curriculum and examinations for an innovative, semester-long coding technology independent study, through which Katie Jentleson became the first Duke Art History graduate student to have computer programming (Python, SQL and Regular Expressions) accepted as one of her foreign language equivalency exams (fall 2012).

Administered a year-long independent study to Katherine de Vos Devine during which we developed databases and prototype web-based visualization platforms for exploring language usage in intellectual property case law (2011 – 2012).

Consulted with many Historical Art Markets undergraduate and graduate students on data gathering, cleaning, modeling (relational database construction) and visualization. Also developed tutorial materials for and presented in-class on these topics (2011 – *present*).

Developed and conducted training sessions for scientists and engineers (undergraduates, grad students and faculty) in many aspects of visualization and visual literacy; including workshops on poster design, illustration techniques, and network visualization software; and preparation of diagrams, charts and graphs (2008 – *present*).

### UNIVERSITY OF MICHIGAN

Taught introductory chemistry lab sections (fall 1995 – winter 1997).

Received Milton Tamres Outstanding Teaching Award (1997).

Received Rackham Graduate School Outstanding Graduate Student Instructor Award (1997).

Participated in and helped organize Summer Science for Girls (2 years, Physics department).

Conducted oral and video in-class teaching feedback sessions through the Center for Research on Learning and Teaching.

## Service

Co-organized and ran a weekly talk series, the Visualization Friday Forum (fall 2012 – *present*). It averages 50 attendees per week and brings together researchers and students from many departments across campus and other local universities.

Helped bring the *Spaces and Places: Mapping Science* exhibit of 100 maps over 3 campus venues to Duke (spring 2015).

Served on the MA+S Steering Committee (2008 – 2015), and the Visual Studies Executive Committee (2012 – 2013).

Organized and ran the weekly Media Arts + Sciences Rendezvous talk series where faculty and student have a chance to share their recent work to both educate and get feedback from the community (fall 2011 – spring 2015).

## Presentations (selected from more than 49 since 2000)

### RESEARCH RESULTS & METHODS

"Putting a Spit Shine on the Getty Provenance Index", *Media Arts + Sciences Rendezvous* (May 2015)

"Low-Dimensional Representations for Learning Fast Molecular Dynamics Simulations and New Visualizations", *Visualization Friday Forum* (w/Miles Crosskey, Feb 2014)

"Staying Broad and Shallow: Learning on the Fly" [*] *Doing Digital Humanities, Duke FHI* (Dec 2013)

"Enabling Art History Visualization" [*] *Center for Molecular and Biomolecular Imaging Symposium* (Dec 2012)

"Visualizing Text: Tools & Techniques" [*] *Text > Data Speaker Series, Duke Library* (Nov 2012)

"Data Visualization: Some Whys and Hows" [*] *Digital Research 1.0 Talk Series, Duke Library* (Mar 2012)

"Markov Modeling of Rat Hippocampal Cell Dynamics" [†] *Nonlinear Sciences Gordon Conference* (Jun 2007)

"Spatio-Temporal Dynamics of Traumatized Neuron/Glial Cell Subnetworks" [†] *Nonlinear Sciences Gordon Research Conference* (Jun 2005)

### WORKSHOPS & TUTORIALS

"Practical Data Visualization" [*‡] *Duke University Marine Lab* (Jun 2014 & 2015)

"Designing Academic Figures" [*‡] *Duke iiD, Data+* (Jun 2015)

"Visual Communication for Scientific Research" [*‡] *Medical Physics Graduate Student Symposium* (Mar 2014)

"Using Gephi for Network Analysis and Visualization" [‡] *Duke Data and Visualization Services* (Feb 2014, Sept 2014)

"Scientific Posters" [*‡] *Duke Chemistry* (Jun 2013)

"Figures, Tables and Captions for Print" [*] *Medical Physics Graduate Student Symposium* (Feb 2010 & 2012)

"Scientific Graphics with Adobe Illustrator: Don't be scared – really, it's fun!" [*] *Duke Social Sciences Research Institute* (Nov 2008)

[* invited, † peer reviewed, ‡ in collaboration with Duke Data and Visualization Services]

## Selected Publications

JOURNAL ARTICLES

R Etemadpour, E Monson, L Linsen (2013) "The Effect of Stereoscopic Immersive Environments on Projection-Based Multi-dimensional Data Visualization", *Information Visualization*, 389–397

C Chan, M Billard, SA Ramirez, H Schmidl, E Monson & TB Kepler (2013) "A Model for Migratory B Cell Oscillations from Receptor Down-Regulation Induced by External Chemokine Fields", *Bull Math Biol* 75, (1), 185–205.

AJ Catllá, DG Schaeffer, TP Witelski, EE Monson & AL Lin (2008) "On Spiking Models for Synaptic Activity and Impulsive Differential Equations", *SIAM Review* 50, 553–569.

AC Horton, B Rácz, EE Monson, AL Lin, RJ Weinberg & MD Ehlers (8 Dec 2005) "Polarized Secretory Trafficking Directs Cargo for Asymmetric Dendrite Growth and Morphogenesis", *Neuron* 48, 757–771.

E Monson & R Kopelman (2004) "Non-Classical Kinetics of an Elementary A+B→C Reaction-Diffusion System Showing Effects of a Speckled Initial Reactant Distribution and Self-Segregation: Experiments", *Phys. Rev. E* 69 (2), 021103.

H Xu, F Yan, EE Monson, & R Kopelman (2003) "Room-temperature preparation and characterization of poly (ethylene glycol)-coated silica nanoparticles for biomedical applications", *J. Biomed. Mater. Res. Part A* 66A, 870–879.

NK Kerner, B Black, E Monson & L Meeuwenberg (2002) "Training Instructors to Facilitate Collaborative Inquiry", *J. Student Centered Learning* 1, 29–36.

E Monson & R Kopelman (2000). "Observation of laser speckle effects and nonclassical kinetics in an elementary chemical reaction", *Phys. Rev. Lett.* 85, 666–669.

HA Clark, SLR Barker, M Brasuel, MT Miller, E Monson, S Parus, ZY Shi, A Song, B Thorsrud, A Kopelman, A Ade, W Meixner, B Athey, M Hoyer, D Hill, R Lightle & MA Philbert (1998) "Subcellular optochemical nanobiosensors: probes encapsulated by biologically localised embedding (PEBBLEs)", *Sens. Actuator B-Chem.* 51, 12–16.

AL Lin, E Monson & R Kopelman (1997) "Nonclassical dimension-dependent kinetics of a photobleaching reaction in a focused laser beam 'phototrap'", *Phys. Rev. E* 56, 1561–1566.

E Monson, G Merritt, S Smith, JP Langmore & R Kopelman (1995) "Implementation of an NSOM System for Fluorescence Microscopy", *Ultramicroscopy* 57, 257–262.

BOOK CHAPTERS

E Monson (2012) "Delving into Image Features", in FC Frankel & AH DePace, *Visual Strategies (ch. Interactive Graphics)*. Yale University Press, New Haven & London, 124–127.

E Monson, M Brasuel, MA Philbert & R Kopelman (2003) "PEBBLE Nanosensors for *in vitro* Bioanalysis", in T Vo-Dinh (ed.) *Biomedical Photonics Handbook*. CRC Press, Boca Raton, FL, 59.1–59.14.

CONFERENCE PROCEEDINGS

EE Monson, G Chen, R Brady & M Maggioni (2010) "Data representation and exploration with Geometric Wavelets", IEEE Symposium on Visual Analytics Science and Technology (VAST), 243–244.

E Monson & R Kopelman (2001) "Observation of Laser Speckle Effects in an Elementary Chemical Reaction", in JM Drake, J Klafter, PE Levitz, M Urback (eds.) *Materials Research Society Symposium*, Boston, MA, T7.27.21–T27.27.26.

E Monson, AL Lin & R Kopelman (1996) "The anomalous diffusion-limited reaction kinetics of a phototrapping reaction", *Materials Research Society Symposium*, Boston, MA.

PATENTS

J Anker, E Monson, R Kopelman & MA Philbert (2003) "Modulated Chemical Sensors", *Patent Application No. 10/419,033*, Regents of the University of Michigan, USA.

R Kopelman, HA Clark, E Monson, S Parus, MA Philbert & B Thorsrud (2002) "Optical fiberless sensors", *US Patent No. 6,379,955*, Regents of the University of Michigan, USA.

# Curriculum Vitae

| | |
|---|---|
| **PERSONAL INFORMATION** | **Name**: Thomas J. Nechyba |

**Address**:   Department of Economics     or     Social Science Research Institute
318 Social Science Building           230D Gross Hall
419 Chapel Drive, Box 90097          140 Science Drive
Duke University                         Duke University
Durham, NC 27708                  Durham, NC 27708

**Phone**:    (919) 660-1826                   (919) 681-6590
**E-mail**:   nechyba@duke.edu
**Date of Birth**: March 27, 1968
**Family**:   Married 6/20/1992 (Stacy L. Nechyba)
               Three children: Eleanor Li, Jennifer Li, Katherine Wu (Ages 15, 14, 12)
**Citizenship**: USA and Austria (dual citizenship)

**PRESENT EMPLOYMENT & AFFILIATIONS**

**Duke University**, Durham, NC
Professor of Economics: January 1, 2003 – Present
Professor of Public Policy Studies (secondary appointment): July 1, 2003 – Present
Associate Professor of Economics and Public Policy Studies (with tenure): July 1, 1999 – December 31, 2002

**National Bureau of Economic Research**, Cambridge, MA
Research Associate: Spring 2000 – Present
Faculty Research Fellow: Spring 1995 – Spring 2000

**Lincoln Institute of Land Policy**, Cambridge, MA
Board of Directors: July 2010 – Present

**Excelsior Charter School**, Durham, NC
Board of Directors, 2014 – Present

**Institute for Economic Research (CESifo)**, University of Munich, Germany
Research Fellow: 2001 – Present

**PAST EMPLOYMENT & AFFILIATIONS**

**Stanford University**, Stanford, CA
Assistant Professor of Economics: July 1, 1994 – April 30, 2000
(on leave September 1, 1998 – April 30, 2000)

**Hoover Institution on War, Revolution and Peace**, Stanford, CA
National Fellow: September 1998 – August 1999

**Fundacao Getulio Vargas**, Rio de Janeiro, Brazil
Visiting Professor: June 1998 – September 1998

**Center for Economic Studies (CES)**, University of Munich, Germany
Visiting Fellow: Fall 2000

**EDUCATION**

**University of Rochester**, Rochester, NY
Ph.D. in Economics: May 22, 1994
Dissertation: *Fiscal Federalism and Local Public Finance: A General Equilibrium Approach with Voting* (Marcus Berliant and Robert Strauss, Advisors)

**University of Florida**, Gainesville, FL
B.A. in Economics: July 1989 (with High Honors)

| UNIVERSITY LEADERSHIP & SERVICE | **Social Science Research Institute**, Duke University |
|---|---|

**Social Science Research Institute**, Duke University
Director: July 1, 2012 – Present
Advisory Board: Fall 2004 – 2008

**Bass Connections**, Duke University
Director, Education & Human Development Theme: January 2013 – Present
Founding Faculty Advisory Board: Fall 2011 – Present

**Economics Center of Teaching (EcoTeach)**, Duke University
Director: January 2011 – Present
Founding Director: July 1, 2000 – May 31, 2003

**Department of Economics**, Duke University
Duke Financial Economics Steering Committee, 2011 – Present
Department Chair: June 1, 2003 – May 31, 2009
Chair, Senior Recruiting Committee: Summer 2002 – Spring 2003
Chair, Departmental External Review Committee: Spring 2002 – Spring 2003
Chair, Micro-Theory Recruiting Committee: Fall 1999 – Spring 2000
Executive Committee: September 1999 – June 2009; January 2011 – Present
Director of Undergraduate Studies: September 2001 – June 2003
Chair, Committee to Reform Undergraduate Education: Fall 1999 – Spring 2000

**Academic Council**, Duke University
Elected Member: July 1, 2000 – June 30, 2004

**Arts & Sciences Council**, Duke University
Elected Member: September 1, 2003 – August 31, 2004
Task Force on Budget & Finance: Fall 2001 – Spring 2002
Social Sciences Planning Group: Fall 2001 – Spring 2002

**Department of Economics**, Stanford University
Associate Director of Undergraduate Studies: Fall 1996 – Spring 1997
Chair, Committee to Reform Major: Fall 1996 – Spring 1997

**EXTERNAL LEADERSHIP & SERVICE**

**Editorial Board Service**
*American Economic Review*: 2001 – 2007
*American Economic Journals: Policy*: 2007 – Present
*BE Journal of Economic Analysis and Policy*: 2006 – Present
*International Tax and Public Finance*: 2003 – Present
*Journal of Economic Literature*: 2013 – Present
*Journal of Public Economic Theory*: 1997 – 2002
*Regional Science and Urban Economics*: 2006 – Present

**Conference Organization**

- Program Committee: Econometric Society Summer Meetings, 2007

- Program Committee: "Understanding the 1990's: The Economy in Long-Run Perspective", Duke/UNC, March 26-27, 2004

- Chair, Program Committee and Host: 4th International Conference on Public Economic Theory, June 12-15, 2003 (Program published as "Program for the fourth APET International Conference on Public Economics," *Journal of Public Economic Theory* 6(1), 181-91, 2004)

- Scientific Organizing Committee: International Institute of Public Finance (IIPF) Congress, Helsinki, Summer 2002

- Chair, Program Committee: Interdisciplinary Conference on "Education and Incentives", Duke University, December 7-8, 2001

- Chair, Program Committee, Stanford Institute for Theoretical Economics (SITE): "Intergovernmental Competition in Public Economics", August 11-15, 1997 and July 6-10, 1998

- Chair, Program Committee, Stanford Institute for Theoretical Economics (SITE): "Competition in Urban Economics", July 13-17, 1998

**Referee Service**

*American Economic Review, Journal of Political Economy, Econometrica, Quarterly Journal of Economics, European Economic Review, International Economic Review, Review of Economics and Statistics, Review of Economic Studies, American Political Science Review, Journal of Monetary Economics, Journal of Public Economics, Journal of Urban Economics, National Tax Journal, Regional Science and Urban Economics, International Tax and Public Finance, Journal of Economic Theory, Games and Economic Behavior, Journal of Economic Literature, Journal of Business and Economic Statistics, Southern Economic Journal, Economic Design, Education Economics, Journal of Economic Education, Bulletin of Economic Research, Contemporary Economic Policy, Economics of Governance, Developmental Psychology, National Science Foundation Proposal Reviews, Social Sciences and Humanities Research Council of Canada*

**DISTINCTIONS**
.

**Awards**

- Duke University Scholar-Teacher of the Year, 2007

- Kenneth G. Elzinga Distinguished Teaching Award, Southern Economics Association, 2004

- Fuchsberg–Levine Family Bass Chair, Duke University, July 2002 – June 2007

- Dean's Award for Distinguished Teaching, Stanford University, 1996

- Voted among "Top 10 Professors at Stanford", *Stanford Review*, 1995

- First Place, Outstanding Dissertation Award, Association for Public Policy and Management, 1994

- Second Place, 24th Annual Competition for Outstanding Doctoral Dissertation in Government Finance and Taxation, National Tax Association, 1994

- President's (Edward Peck Curtis) Award for Excellence in Undergraduate Teaching, University of Rochester, 1994

- Conibear Prize for Best Third Year Paper, University of Rochester, 1992

**Honors**

- J. Fish Smith and Lillian F. Smith Lecture, Brigham Young University, 2006

- Congress for the New Urbanism Keynote Lecture, 2006
- Keynote Address, CESifo Public Sector Conference, University of Munich, 2004
- Bass Society of Fellows, Duke University, 2002 – Present
- National Fellowship, Hoover Institution, Stanford University, 1998-99
- Sigrid Banks Fellowship, Stanford University, 1998-99
- Keynote Speaker, New Zealand Association of Economists, 1998

**Grants**

- Hewlett-Packard, "Technology-Based Education Interventions," 2013-15. (PI)
- NSF (SBR-0339000), "An Empirical Investigation of Peer Effects and of Household Responses to School Policy Changes," (with J. Vigdor) 2004–2009 (PI)
- Lincoln Institute of Land Policy, "Urban Sprawl," (with Randy Walsh), 2002–2004 (PI)
- Spencer Foundation, "The Role of Peers, Parental Choices and Neighborhoods," (with J. Vigdor and N. Hill), 2000–2004 (Co-PI)
- NSF (SBR-9905706), "Efficient and Equitable Delivery of Education in a District-Based Public School System," 1999–2001 (PI)
- NSF (SBR-9809269), "Vouchers, Migration and School Quality," 1998–1999 (PI)
- New Zealand Ministry of Education, "The Impact of Family and Community Resources on Educational Outcomes," 1998–1999 (PI)
- Lincoln Institute of Land Policy, "Land Taxes as State Tax Bases," 1998–1999 (PI)
- Lincoln Institute of Land Policy, "The Revenue Potential for Land Value Taxation," 1997–1998 (PI)
- Center for Economic Policy Research (CEPR), "Education and Welfare Policy," 1994–1996

**PUBLICATIONS** **Books**

1) *Microeconomics: An Intuitive Approach with Calculus*, Cengage Learning (Second Edition), January 2016).
2) *Microeconomics: An Intuitive Approach*, Cengage Learning, (Second Edition), January 2016).
3) *The Effect of Family and Community Resources on Education Outcomes*, New Zealand Government, Ministry of Education (with D. Older-Aguilar and P. McEwan), 1999.

**Book Reviews and Comments**

1) "Nonfiscal Residential Zoning" (Calabrese, S, D. Epple and R. Romano), commented on in *The Tiebout Model at Fifty*, Lincoln Institute Press (2006).
2) *IQ and the Wealth of Nations* (Lynn, R. and T. Vanhanen) reviewed in *Journal of Economic Literature* 42 (2004), pp. 220–21.

3) "Land Taxation in New York City: A General Equilibrium Analysis" (A. Schwatz), commented on in *Urban Issues and Public Finance: Essays in Honor of Dick Netzer* (2004).

4) *Handbook of Public Economics: Vol. 3* (A. Auerbach and M. Feldstein, eds.) reviewed in *Journal of Economic Literature* 41 (2003), pp. 1299–1301.

5) "Immigrant Children and New York City Schools: Segregation and Its Consequences," commented on in *Brookings–Wharton Papers on Urban Affairs,* pp. 208–11 (2002).

6) *State Sales and Income Taxes: An Economic Analysis* (G. Zodrow) reviewed in *Journal of Regional Science* (2001).

7) *Social Welfare and Individual Responsibility* (Schmidtz, D. and R. Goodin) reviewed in *Economics and Philosophy* (2000).

8) *Local Government Tax and Land Use Policies in the United States: Understanding the Links* (H. Ladd) reviewed in *Regional Science and Urban Economics* 29 (1999), pp. 547–52.

9) *Competitive Governments: An Economic Theory of Politics and Public Finance* (A. Breton) reviewed in *Journal of Economic Literature* 35 (1997), pp. 2062–64.


**Published Papers**

1) "Mobilizing the Private Sector in the United States: A Theoretical Overview," in *School Choice International: Exploring Public-Private Partnerships* (Chakrabarti, R. and P. Peterson, eds.), MIT Press, pp. 47–70 (2009).

2) "The Social Context of Vouchers," in *Handbook of Research on School Choice* (Berends, M., Springer, M., Ballou, D. and H. Walberg, eds.), Routledge, pp. 289–308 (2009).

3) "Alternative School Finance Strategies," *St. Louis Federal Reserve*, 2008.

4) "School Finance, School Choice and Residential Segregation," *CESifo Economic Studies*, (2005).

5) "Peer Effects in North Carolina Public Schools," in *Schools and the Equal Opportunity Problem* (Woessmann, L. and P. Peterson, eds.), MIT Press, pp. 73–102 (2007) (with Jacob Vigdor).

6) "The Efficiency and Equity of Tiebout in the United States: Taxes, Services and Property Values," in *Land Policies and Their Outcomes* (Ingram, G. and Y. Hong, eds.), Lincoln Institute Press (2007).

7) "Income and Peer Quality Sorting in Public and Private Schools," in *Handbook of Economics of Education* (Hanushek, E. and F. Welch, eds.), North Holland, pp. 1327–68 (2006).

8) "Urban Sprawl," *Journal of Economic Perspectives* 18(4), pp. 177–200 (2004) (with Randy Walsh).

9) "School Competition and School Quality in the U.S.", *CESifo DICE Report – Journal of Institutional Comparison* 4, pp. 3–8 (2004).

10) "Fiscal Decentralization," in *Handbook of Regional and Urban Economics* (Henderson, V. and J. Thisse, eds.), North Holland, pp. 2426 – 80 (2004) (with Dennis Epple).

11) "Prospects for Achieving Equity or Adequacy in Education: The Limits of State Aid in General Equilibrium," in *Helping Children Left Behind: State Aid and the Pursuit of Educational Equity* (Yinger, J., ed.), MIT Press, pp. 111–45 (2004).

12) "Centralization, Fiscal Federalism and Private School Attendance," *International Economic Review* 44(1), pp. 179–204 (2003).

13) "School Finance, Spatial Income Segregation and the Nature of Communities," *Journal of Urban Economics* 54(1), pp. 61–88 (2003).

14) "What Can Be (and What Has Been) Learned from General Equilibrium Simulation Models of School Finance," *National Tax Journal* 56(2), pp. 387–414 (2003).

15) "Public School Finance and Urban School Policy: General Versus Partial Equilibrium Analysis," *Brookings-Wharton Papers on Urban Affairs*, pp. 139–70 (2003).

16) "Introducing School Choice into Multidistrict Public School Systems," in *The Economics of School Choice* (C. Hoxby, ed.), University of Chicago Press, pp. 145–94 (2003).

17) "The Benefit and the New View: Where Do We Stand 25 Years Into the Debate?", in *Property Taxation and Local Public Finance* (W. Oates, ed.), Lincoln Institute Press, pp. 113–21 (2002).

18) "Social Approval, Values and AFDC: A Re-Examination of the Illegitimacy Debate," *Journal of Political Economy*, pp. 638–72 (2001).

19) "How Does a Community's Demographic Composition Alter Its Fiscal Burdens," in *Demographic Change and Fiscal Policy* (Auerbach, A. and R. Lee, eds.), Cambridge University Press, pp. 101–48 (2001) (with Thomas MaCurdy).

20) "Mobility, Targeting and Private School Vouchers," *American Economic Review* 90(1), pp. 130–46 (2000). [Reprinted in *The International Library of Critical Writings in Economics: The Economics of Schooling and School Quality* (Mark Blaug, ed.), Edward Elgar Publishing, (2004).]

21) "School Finance Reform: Introducing the Choice Factor," in *City Schools: Lessons from New York* (Ravitch, D. and J. Viteritti, eds.), John Hopkins University Press, pp. 367–92 (2000) (with Michael Heise).

22) "School Finance Induced Migration Patterns: The Impact of Private School Vouchers," *Journal of Public Economic Theory* 1(1), pp. 5–50 (1999).

23) "Community Choice and Local Public Services: A Discrete Choice Approach," *Regional Science and Urban Economics* 28(1), pp. 51–74 (1998) (with Robert Strauss).

24) "An Economic Framework for Assessing the Fiscal Impact of Immigration," in *The Immigration Debate: Studies on the Economic, Demographic and Fiscal*

*Effects of Immigration* (Smith, J. and B. Edmonston, eds.), National Academy Press, (1998) (with Jay Battacharaya and Thomas MaCurdy).

25) "Public School Finance and Vouchers in a General Equilibrium Tiebout World," *Proceedings of the 90th Annual Conference of the National Tax Association*, pp. 119–25 (1998).

26) "Replacing Capital Taxes with Land Taxes: Efficiency and Distributional Implications with an Application to the US," in *Land Value Taxation: Can It and Will It Work Today?* (D. Netzer, ed.), Lincoln Institute Press, (1998).

27) "Existence of Equilibrium and Stratification in Local and Hierarchical Public Goods Economies with Property Taxes and Voting," *Economic Theory* 10, pp. 277–304 (1997).

28) "Local Property and State Income Taxes: The Role of Interjurisdictional Competition and Collusion," *Journal of Political Economy* 105(2), pp. 351–84 (1997).

29) "Computable General Equilibrium in Local Public Finance and Fiscal Federalism: Applications to Local Taxation, Intergovernmental Aid and Educational Vouchers," in *Fiscal Aspects of Evolving Federations* (D. Wildasin, ed.), Cambridge University Press, pp. 168–93 (1997).

30) "A Computable General Equilibrium Model of Intergovernmental Aid," *Journal of Public Economics* 62, pp. 363–97 (1997).

31) "Tax Competition in Federal Systems: Political Accountability and Financial Constraints," in *The New Federalism* (Ferejohn, J. and B. Weingast, eds.), Hoover Institution Press, pp. 3–61 (1997) (with Ron McKinnon).

32) "Fiscal Federalism and Local Public Finance: A Computable General Equilibrium (CGE) Approach," *International Tax and Public Finance* 3, pp. 215–31 (1996).

33) "Fiscal Federalism and Local Public Finance: A General Equilibrium Approach with Voting," *Proceedings of the 86th Annual Conference of the National Tax Association*, pp. 136–41 (1995).

34) "The Southern Wage Gap, Human Capital and the Quality of Education," *Southern Economic Journal* 57, pp. 308–22 (1990).

# JEROME P. REITER

Department of Statistical Science, Duke University
Box 90251, Durham, NC 27708
phone: 919 668 5227. email: jerry@stat.duke.edu.
January 19, 2016

## EDUCATION

Ph.D. in Statistics, Harvard University, 1999.

A.M. in Statistics, Harvard University, 1996.

B.S. in Mathematics, Duke University, 1992.

## DISSERTATION

"Estimation in the Presence of Constraints that Prohibit Explicit Data Pooling."
Advisor: Donald B. Rubin.

## POSITIONS

Academic Appointments

Professor of Statistical Science, Duke University, 2015 - present.

Mrs. Alexander Hehmeyer Professor of Statistical Science, Duke University, 2013 - 2015.

Mrs. Alexander Hehmeyer Associate Professor of Statistical Science, Duke University, 2010 - 2013.

Associate Professor of Statistical Science, Duke University, 2009 - 2010.

Assistant Professor of Statistical Science, Duke University, 2006 - 2008.

Assistant Professor of the Practice of Statistics and Decision Sciences, Duke University, 2002 - 2006.

Lecturer in Statistics, University of California at Santa Barbara, 2001 - 2002.

Assistant Professor of Statistics, Williams College, 1999 - 2001.

Other Appointments

Associate/Deputy Director of Information Initiative at Duke, Duke University, 2013 - present.

Social Sciences Research Institute Data Services Core Director, Duke University, 2010 - 2013.

Interim Director, Triangle Research Data Center, 2006.

Senior Fellow, National Institute of Statistical Sciences, 2002 - 2005.

## ACADEMIC HONORS

Gertrude M. Cox Award, 2014.

W. J. Youden Award in Interlaboratory Testing, 2013.

Bass Chair for Excellence in Undergraduate Teaching and Research, Duke University, 2010 - 2015.

Fellow of the American Statistical Association, 2011.

Cathie Marsh Fellowship, Centre for Census and Survey Research, University of Manchester, 2010.

Elected member of International Statistical Institute, 2007.

Recipient of Alumni Distinguished Undergraduate Teaching Award, Duke University, 2007.

## GRANTS AND FUNDED RESEARCH

Arthur P. Sloan Foundation, "Computational Methods and Tools for Creating High Quality Household Microdata", 1/2016 - 7/2017, $108,903. Role: Principal Investigator.

National Science Foundation (DIBBs), "An Integrated System for Public/Private Access to Large-Scale, Confidential Social Science Data," 1/2015 - 12/2017, $1,498,683. Role: Principal Investigator.

National Science Foundation (NCRN), "The Triangle Census Research Network: Supplemental Award 2," 10/2013 - 9/2016, $199,737. Role: Principal Investigator.

National Science Foundation (NCRN), "The Triangle Census Research Network: Supplemental Award 1," 10/2012 - 9/2016, $322,027. Role: Principal Investigator.

U.S. Bureau of the Census, "Creating a Synthetic Public Use File for the Longitudinal Business Database," 9/2012 - 9/2014, $117,919 plus $30,000 discount on Duke's costs for Triangle Census Research Data Center. Role: PI on Intergovernmental Personnel Agreement.

National Science Foundation (NCRN-RCN), "Coordination of the NSF-Census Research Network," 7/2012 - 6/2017, $748,577 total, $0 to Duke. Role: Co-Investigator (PI: Alan Karr).

National Science Foundation (NCRN), "The Triangle Census Research Network," 10/2011 - 9/2016, $2,997,591 total, $2,144,766 to Duke. Role: Principal Investigator.

National Science Foundation (MMS), "Multiple Imputation Methods for Handling Missing Data in Panel Studies with Refreshment Samples," 6/2011 - 5/2013, $160,000. Role: Principal Investigator.

U.S. Bureau of the Census, "Creating a Synthetic Public Use File for the Longitudinal Business Database," 9/2010 - 9/2012, $95,828 plus $30,000 discount on Duke's costs for Triangle Census Research Data Center. Role: PI on Intergovernmental Personnel Agreement.

National Science Foundation (TC): "Large: Collaborative Research: Practical Privacy: Metrics and Methods for Protecting Record-level and Relational Data," 7/2010 - 7/2015, $583,169. Role: Principal Investigator.

National Science Foundation (SES): "Joint NSF-Census-IRS Workshop on synthetic data and confidentiality protection," 7/2009 - 6/2010, $18,480. Role: Co-Investigator.

National Institutes of Health (NIA): "R21: Sharing Confidential Datasets With Geographic Identifiers Via Multiple Imputation," 3/2009 - 1/2011, $351,780. Role: Principal Investigator.

National Science Foundation (MMS), "Methodology for Improving Public Use Data Dissemination Via Multiply-Imputed, Partially Synthetic Data," 6/2008 - 5/2011, $180,000. Role: Principal Investigator.

Environmental Protection Agency, "Southern Center for Environmentally Driven Disparities in Birth Outcomes," 5/2007 - 5/2011, $7,735,620. Role: Co-Investigator (PI: Marie Lynn Miranda).

National Institutes of Health (NIDA), "The Alcohol Pharmacology Education Partnership," 8/2005 - 7/2010, $1,250,000. Role: Co-investigator (PI: Rochelle Schwartz-Bloom).

National Science Foundation (ITR), "Information Technology Challenges for Secure Access to Confidential Social Science Data," 10/2004 - 10/2008, $2,938,000 total, $247,903 to Duke. Role: Senior Scientist (PI: John Abowd).

U.S. Bureau of the Census, "Using Synthetic Data to Protect Confidentiality of People in Group Quarters in the American Community Survey," 9/2005 - 9/2007, $24,000. Role: Principal Investigator.

U.S. Bureau of the Census, "Synthetic Data Methods for the Survey of Income and Program Participation Public Use File," 6/2005 - 6/2006, $25,000. Role: Principal Investigator.

U.S. Bureau of the Census, "Assessing Disclosure Risks for Remote Access Model Servers," 6/2004 - 7/2005, $25,000. Role: Principal Investigator.

Transportation Research Board, 2005. Commissioned to write paper on threats to data confidentiality in public-use transportation data (with David Banks), $10,000.

National Institute of Statistical Sciences, "Data Confidentiality, Data Quality, and Data Integration for Federal Databases: Foundations to Software Prototypes," 8/2002 - 8/2004, $25,000. Role: Senior Fellow.

National Academy of Sciences Panel on Confidential Data Access for Research Purposes, 2003. Commissioned to write paper on measuring disclosure risks, $5,000.

U.S. Bureau of the Census, "Research on Synthetic Data," 8/2002 - 8/2003, $16,000. Role: Principal Investigator.

## PROFESSIONAL SERVICE

Editorial Appointments

Editorial Board, *Transactions on Data Privacy*, 2013 - present.
Associate Editor, *Journal of Privacy and Confidentiality*, 2006 - present.
Associate Editor, *Survey Methodology*, 2004 - 2013.
Associate Editor, *Journal of Statistical Theory and Practice*, 2009 - 2013.
Associate Editor, *Journal of the American Statistical Association*, 2006 - 2013.
Advisory Board, *The Encyclopedia of Measurement and Statistics*, 2005.

Advisory Boards and Expert Panels

University of Michigan Safe Designs Project Advisory Board, 2012 - present.
Wisconsin Longitudinal Study Genetic Advisory Board, 2010 - 2013.
Technical Expert Panel for CMS Comparative Effectiveness Research Public Use Data Pilot Project, 2010 - 2011.
Panel Member, Office of Civil Rights Workshop on Reconsidering the HIPAA De-identification Standard, 2010.

Technical Expert Panel for the Patient Safety Organization Project, Agency for Healthcare Research and Quality, 2008.

National Center for Education Statistics Configuration of Longitudinal Studies and Data Integration Task Force, 2007 - 2008.

National Center for Education Statistics Confidentiality Task Force, 2006 - 2007.

Membership on National Academy of Science Committees, Panels, and Expert Meetings

Committee: Standing Committee for the American Opportunity Study — Phase 1, 2015 - present.

Expert Meeting: BEA Expert Meeting on Exploiting Commercial Data for Official Economic Statistics, 2015.

Panel: Addressing Priority Technical Issues for the Next Decade of the American Community Survey, 2012 - 2015.

Committee: Long-Term Stewardship of Safety Data from the Second Strategic Highway Research Program, 2012 - 2013.

Expert Meeting: Redaction Strategy for the Airline Pilots Safety Data, 2008.

Panel: Collecting, Storing, Accessing, and Protecting Social Survey Data Containing Biological Measures, 2008 - 2010.

Panel: Re-engineering the Survey of Income and Program Participation, 2006 - 2009.

Panel: Confidentiality Issues Arising from the Integration of Remotely Sensed Data with Social Science Survey and Other Self-Identifying Data, 2005 - 2007.

Positions in American Statistical Association

Statistical Ambassadors Program Working Group, 2015 - present.

Youden Award Selection Committee, 2015 - present.

NC-ASA Chapter President, 2014, and Vice President, 2013.

COSGB Fiscal Oversight Committee, 2012 - present.

Secretary/Treasurer, Section on Bayesian Statistical Sciences, 2012 - 2015.

Scientific and Public Affairs Advisory Committee, 2012 - present. Chair 2015 - present.

Committee on Privacy and Confidentiality in Statistics, 2006 - 2012. Chair 2009 - 2012.

Chair, Section on Statistics in Sports, 2008.

Program Chair for Statistics in Sports, 2005 Joint Statistical Meetings.

Program Chair for General Methodology, 2001 Joint Statistical Meetings.

Conference and Workshop Program Committees

Local Scientific Coordinator, SAMSI Program on Computational Methods in Social Science, 2013 - 2014.

NSF-NBER Time Series Conference, 2010.

Total Survey Error Workshops, 2005 and 2008.

International Conference on Data Mining Workshop on Privacy and Security Aspects of Data Mining, 2005.

Honors Examiner in Statistics and Probability, Swarthmore College, 2008.

Student Paper Judge, International Biometric Society (WNAR) Annual Meeting, 2005.

BMRD Study Section Member (ad hoc), National Institutes of Health.

Proposal Review Panel Member, National Science Foundation (five panels).

Proposal Review Panel Member, Institute of Education Sciences.

Proposal and Project Reviewer for National Science Foundation, National Research Council, Institute of Medicine, Census Bureau Center for Economic Studies, Agency for Healthcare Research and Quality DEcIDE program, Statistics New Zealand, Mathematics of Information Technology and Complex Systems Network of Centres of Excellence (Canada).

Referee. Journals include *American Journal of Epidemiology*; *American Political Science Review*; *Annals of Applied Statistics*; *Biometrical Journal*; *Biometrics*; *Biometrika*; *Biostatistics*; *BMC Medical Research Methodology*; *Computational Statistics and Data Analysis*; *Data and Knowledge Engineering*; *Demography*; *Epidemiology*; *Health Services and Outcomes Research Methodology*; *IERI Monographs*; *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*; *Journal of the American Statistical Association*; *Journal of Biomedical Informatics*, *Journal of Business and Economic Statistics*; *Journal of Causal Inference*; *Journal of Computational and Graphical Statistics*; *Journal of Law, Medicine, and Ethics*; *Journal of Official Statistics*; *Journal of Privacy and Confidentiality*; *Journal of Probability and Statistics*; *Journal of Quantitative Analysis in Sports*; *Journal of the Royal Statistical Society Series A*; *Journal of Statistical Planning and Inference*; *Journal of Statistical Software*; *Journal of Statistical Theory and Practice*; *Journal of Statistics Education*; *Journal of Survey Statistics and Methodology*; *Management Science*; *Marketing Science*; *Metron*; *PLoS One*; *Psychological Methods*; *Statistical Methodology*; *Public Opinion Quarterly*; *Sociological Methods and Research*; *Stat*; *Statistical Science*; *Statistics and Computing*; *Statistics in Medicine*; *Statistics, Policy, and Politics*; *Survey Methodology*; *The American Statistician*; *The Stata Journal*; *Transactions on Data Privacy*; *Transactions on Knowledge Discovery from Data*.

## UNIVERSITY SERVICE

Duke Transportation Advisory Committee, 2014 - 2015.

Duke University Population Research Institute Executive Committee, 2012 - 2015.

Faculty Compensation Committee, 2012 - 2015. Committee chair, 2013 - 2015.

Academic Council Representative, 2012 - 2014 and 2007 - 2009.

Duke Initiative on Survey Methods Advisory Board, 2010 - present.

Triangle Census Research Data Center Advisory Board, 2010 - present.

Advisory Committee on Bass Professorships, 2011 - 2014. Committee chair, 2013 - 2014.

Program for Advanced Research in the Social Sciences Advisory Board, 2004 - 2014.

Arts & Sciences Faculty Assessment Committee, 2011 - 2013.

Masters Degree Coordinator, Department of Statistical Science, 2011 - 2013.

Social Science Research Institute Planning Committee, 2012.

Provost's Informational Futures Committee, 2012.

Committee on Revising the Quantitative Studies Requirement, 2010 - 2011.

Social Sciences Research Institute Steering Committee Member, 2009 - 2011.

Arts & Sciences Council Representative, 2004 - 2007.

Undergraduate Pre-Major Advisor, 2003 - 2007.

Director of Undergraduate Studies, ISDS, 2002 - 2006.

Task Force on Data Analysis in the Social Sciences, 2005.

Task Force on Program in Data Analysis and Computation, 2005.

Presenter, Duke Faculty Forum on Teaching and Technology, 2012; Duke Up Close, 2005, 2006, 2008; Duke School Days, 2004, 2005, 2008, 2010; Duke Family Days, 2007; Duke Homecoming Academic Program, 2007; Duke NC-SC Open House, 2007, 2008; Duke Founders' Day Recognition Celebration, 2007; Jacksonville Duke Alumni Association, 2008.

## STUDENTS AND OTHER ADVISEES

Habilitation Committee Member

> Joerg Drechsler, University of Munich, Germany, 2015.

Postdoctoral Associate Mentor

> Andres Barrientos, 2015 - present; Mauricio Sadinle, 2015 - present; Maria DeYoreo, 2014 - present; Hang Joon Kim, 2012 - 2015; Daniel Manrique-Vallier, 2010 - 2013; Lane Burgette, 2009 - 2011.

Ph.D. Dissertation Committee Advisor

> Tracy Schifeling, expected 2016.

> Lan Wei, expected 2016.

> Nicole Dalzell, expected 2017.

> Jingchen Hu, "Dirichlet Process Mixture Models for Nested Categorical Data," 2015.

> David McClure, "Relaxations of Differential Privacy and Risk/Utility Evaluations of Synthetic Data and Fidelity Measures," 2015.

> Thais Paiva, "Multiple Imputation Methods for Nonignorable Nonresponse, Adaptive Survey Design, and Dissemination of Synthetic Geographies," 2014.

> Jared Murray, "Some Recent Advances in Non- and Semiparametric Bayesian Modeling with Copulas, Mixtures, and Latent Variables," 2013.

> Yajuan Si, "Nonparametric Bayesian Methods for Multiple Imputation of Large Scale Incomplete Categorical Data in Panel Studies," 2012.

> Scott Schwartz, "Bayesian Mixture Modeling Approaches for Intermediate Variables and Causal Inference," 2010 (co-advisor).

> Robin Mitra, "Bayesian Methods to Impute Missing Covariates for Causal Inference and Model Selection," 2008.

> Satkartar Kinney, "Model Selection and Multivariate Inference Using Data Multiply Imputed for Disclosure Limitation and Nonresponse," 2007.

> Christine Kohnen, "Using Multiply Imputed, Synthetic Data to Facilitate Data Sharing," 2005.

Ph.D. Dissertation Committee Member

Tsuyoshi Kunihama, Wenzhao Lian (Electrical and Computer Engineering), 2015; Wen-jie Lu (Mechanical Engineering and Materials Science), 2014; Ann-Sophie Charest (Statistics, Carnegie Mellon University), Lenka Bustikova (Political Science), Lynn Lin, 2012; Simone Gray, Hao Wang, 2010; Chunlin Ji, 2009; Joyee Ghosh, Liang Zhang, 2008.

Masters Thesis Committee Member

Sohae Oh (Masters in Statistical and Economic Modeling), Olenrewaju Akande (Masters in Statistical and Economic Modeling, chair), Wenzhao Lian (Electrical and Computer Engineering), 2015; Yiting Deng (Fuqua School of Business, chair), Melissa Tosiano (Integrated Environmental Health and Toxicology Program), 2012; Arturas Rozenas (Political Science, chair), 2010; Jacob Montgomery (Political Science, chair), 2009.

Undergraduate Research Advisor

Honors thesis advisor: Christine Cheng, 2013; Jane Zhu, 2007 (Program II); Tyler Mc-Cormick, 2005 (Program II).

Statistical science majors (non-Honors): Emily Hadley, Lori Kim, 2015; Josh Ruffin, 2014; Drew Cannon, 2012; Siyu Zheng, 2011; Michael Lyngaas, 2010.

Independent studies: Haley Miller, 2015; Lori Kim, Courtney Jones, 2014; Christine Cheng, 2012; Gregory Caiola, 2009; Joshua Koster, 2008; Stephanie Guan, 2005; Orcun Unlu, 2005.

## TEACHING EXPERIENCE

"Bayesian Approaches to Survey Sampling," Duke, Spring 2010.

"Bayesian and Modern Data Analysis," Duke, Spring 2011, Fall 2011, Spring 2012.

"Complex Surveys and Missing Data Methods," Duke, Fall 2003.

"Data Analysis and Statistical Inference," Duke, Fall 2002, Spring 2003, Fall 2003, Fall 2008.

"Data Analysis for Undergraduate Research," Duke, Spring 2006.

"Design and Analysis of Surveys," Duke, Spring 2008, Spring 2010, Spring 2013, Spring 2015.

"Fundamentals of Modern Statistical Modeling," Duke, Spring 2009.

"Probability and Statistical Inference," Duke, Fall 2004, Spring 2004, Fall 2005, Spring 2007.

"Regression Analysis," Duke, Fall 2006, Spring 2013, Fall 2013, Spring 2014.

"Statistical Analysis with Missing Data," Duke, Fall 2008.

"Undergraduate Research Seminar in Statistical Science," Duke, Spring 2011, Fall 2014, Spring 2015.

"Actuarial Statistics I," UCSB, Winter 2001.

"Actuarial Statistics II," UCSB, Spring 2002.

"General Statistics," UCSB, Fall 2001, Winter 2001.

"Mathematics of Compound Interest," UCSB, Fall 2001.

"Seminar and Projects in Statistical Consulting," UCSB, Spring 2002.

"Elementary Statistics and Data Analysis," Williams, Fall 1999.

"Regression and Forecasting," Williams, Fall 2000.

"Statistics and Data Analysis," Williams, Spring 2000, Spring 2001.

"The Mathematics and Statistics of Surveys," Williams, Fall 1999.

## PEER-REVIEWED PUBLICATIONS

1. McClure, D. and Reiter, J. P. (forthcoming), "Assessing disclosure risks for synthetic data with arbitrary intruder knowledge," *Statistical Journal of the International Association of Official Statistics*.

2. Mitra, R. and Reiter, J. P. (forthcoming), "A comparison of two methods of estimating propensity scores after multiple imputation," *Statistical Methods in Medical Research*.

3. Schifeling, T. and Reiter, J. P. (forthcoming), "Incorporating marginal prior information in latent class models," *Bayesian Analysis*.

4. Si, Y., Reiter, J. P., and Hillygus, D. S. (forthcoming), "Bayesian latent pattern mixture models for handling attrition in panel studies with refreshment samples," *Annals of Applied Statistics*.

5. Wei, L. and Reiter, J. P. (forthcoming), "Releasing synthetic magnitude microdata constrained to fixed marginal totals," *Statistical Journal of the International Association of Official Statistics*.

6. Carrig, M. M., Manrique-Vallier, D., Ranby, K., Reiter, J. P., and Hoyle, R. (2015), "A multiple imputation-based method for the retrospective harmonization of data sets," *Multivariate Behavioral Research*, 50, 383 - 397.

7. Kim, H. J., Cox, L., Karr, A. F., Reiter, J. P. and Wang, Q. (2015), "Simultaneous edit-imputation for continuous microdata," *Journal of the American Statistical Association*, 110, 987 - 999.

8. Kim, H. J., Karr, A. F., and Reiter, J. P. (2015), "Statistical disclosure limitation in the presence of edit rules," *Journal of Official Statistics*, 31, 121 - 138.

9. Quick, H., Holan, S. H., Wikle, C. K., Reiter, J. P. (2015),"Bayesian marked point process modeling for generating fully synthetic public use data with point-referenced geography,"*Spatial Statistics*, 14, 439 - 451.

10. Schifeling, T., Cheng, C. Hillygus, S., and Reiter, J. P., (2015), "Accounting for nonignorable unit nonresponse and attrition in panel studies with refreshment samples," *Journal of Survey Statistics and Methodology*, 3, 265 - 295.

11. Si, Y., Reiter, J. P., and Hillygus, D. S. (2015), "Semi-parametric selection models for potentially non-ignorable attrition in panel studies with refreshment samples," *Political Analysis*, 23, 92 - 112.

12. Siddique, J., Reiter, J. P., Brincks, A., Gibbons, R., Crespi, C., and Brown, C. H. (2015), "Multiple imputation for harmonizing non-commensurate measures in individual participant data meta-analysis," *Statistics in Medicine*, 34, 3399 - 3414.

13. Godin, E. A., Kwiek, N., Sikes, S. S., Halpin, M. J., Weinbaum, C. A., Burgette, L. F., Reiter, J. P., and Schwartz-Bloom, R. D. (2014), "Alcohol pharmacology education partnership: Using chemistry and biology concepts to educate high school students about alcohol," *Journal of Chemical Education*, 91, 165 - 172.

14. Grochowski, C. O., Cartmill, M., Reiter, J., Spaulding, J., Haviland, J., Valea, F., Thibodeau, P. L., McCorison, S., and Halperin, E. C. (2014), "Anxiety in first year medical students taking gross anatomy," *Clinical Anatomy*, 27, 835 - 838.

15. Kim, H. J., Reiter, J. P., Wang, Q., Cox L. H., and Karr, A. F. (2014), "Multiple imputation of missing or faulty values under linear constraints," *Journal of Business and Economic Statistics*, 32, 375 - 386.

16. Kinney, S. K., Reiter, J. P., and Miranda, J. (2014), "SynLBD 2.0: Improving the synthetic Longitudinal Business Database," *Statistical Journal of the International Association for Official Statistics*, 30, 129 - 135.

17. Manrique-Vallier, D. and Reiter, J. P. (2014), "Bayesian estimation of discrete multivariate latent structure models with structural zeros," *Journal of Computational and Graphical Statistics*, 23, 1061 - 1079.

18. Manrique-Vallier, D. and Reiter, J. P. (2014), "Bayesian multiple imputation for large-scale categorical data with structural zeros," *Survey Methodology*, 40, 125 - 134.

19. Paiva, T., Chakraborty, A., Reiter, J. P., Gelfand, A. E. (2014), "Imputation of confidential data sets with spatial locations using disease mapping models," *Statistics in Medicine*, 33, 1928 - 1945.

20. Reiter, J. P., Wang, Q., and Zhang, B. (2014), "Bayesian estimation of disclosure risks in multiply imputed, synthetic data," *Journal of Privacy and Confidentiality*, 6:1, Article 2.

21. Burgette, L. and Reiter, J. P. (2013), "Multiple-shrinkage multinomial probit models with applications to simulating geographies in public use data," *Bayesian Analysis*, 8, 453 - 478.

22. Deng, Y., Hillygus, D. S., Reiter, J. P., Si, Y., and Zheng, S. (2013), "Handling attrition in longitudinal studies: The case for refreshment samples," *Statistical Science*, 28, 238 - 256.

23. Ghosh, J. and Reiter, J. P. (2013), "Secure Bayesian model averaging for horizontally partitioned data," *Statistics and Computing*, 23, 311 - 322.

24. Hu, J., Mitra, R., and Reiter, J. P. (2013), "Are independent parameter draws necessary for multiple imputation?" *The American Statistician*, 67, 143 - 149.

25. Si, Y. and Reiter, J. P. (2013), "Nonparametric Bayesian multiple imputation for incomplete categorical variables in large-scale assessment surveys," *Journal of Educational and Behavioral Statistics*, 38, 499 - 521.

26. Burgette, L. F. and Reiter, J. P. (2012), "Modeling adverse birth outcomes via confirmatory factor quantile regressions," *Biometrics*, 68, 92 - 100.

27. Burgette L. F. and Reiter, J. P. (2012), "Nonparametric Bayesian multiple imputation for missing data due to mid-study switching of measurement methods," *Journal of the American Statistical Association*, 107, 439 - 449.

28. Drechsler, J. and Reiter, J. P. (2012), "Combining synthetic data with subsampling to create public use microdata files for large scale surveys," *Survey Methodology*, 38, 73 - 79.

29. Manrique-Vallier, D. and Reiter, J. P. (2012), "Estimating identification disclosure risk using mixed membership models," *Journal of the American Statistical Association*, 107, 1385 - 1394.

30. Reiter, J. P. and Kinney, S. K. (2012), "Inferentially valid, partially synthetic data: Generating from posterior predictive distributions not necessary," *Journal of Official Statistics*, 28, 583 - 590.

31. McClure, D. and Reiter, J. P. (2012), "Towards providing automated feedback on the quality of inferences from synthetic datasets," *Journal of Privacy and Confidentiality*, 4:1, Article 8.

32. McClure, D. and Reiter, J. P. (2012), "Differential privacy and statistical disclosure risk measures: An investigation with binary synthetic data," *Transactions on Data Privacy*, 5:3, 535 - 552.

33. Orr, S. T., Reiter, J. P., James, S. A., and Orr, C. A. (2012), "Maternal health prior to pregnancy and preterm birth among urban, low income black women in Baltimore: The Baltimore Preterm Birth Study." *Ethnicity and Disease*, 22, 85 - 89.

34. Reiter, J. P. (2012), "Bayesian finite population imputation for data fusion," *Statistica Sinica*, 22, 795 - 811.

35. Reiter, J. P. (2012), "Statistical approaches to protecting confidentiality for microdata and their effects on the quality of statistical inferences," *Public Opinion Quarterly*, 76, 163 - 181.

36. Schwartz, S., Li, F., and Reiter, J. P. (2012), "Sensitivity analysis for unmeasured confounding in principal stratification," *Statistics in Medicine*, 31, 949 - 962.

37. Wang, H. and Reiter, J. P. (2012), "Multiple imputation for sharing precise geographies in public use data," *Annals of Applied Statistics*, 6, 229 - 252.

38. Burgette, L., Reiter, J. P., and Miranda, M. L. (2011), "Exploratory quantile regression with many covariates: An application to adverse birth outcomes," *Epidemiology*, 22, 859 - 866.

39. Drechsler, J. and Reiter, J. P. (2011), "An empirical evaluation of easily implemented, non-parametric methods for generating synthetic datasets," *Computational Statistics and Data Analysis*, 55, 3232 - 3243.

40. Kinney, S. K., Reiter, J. P., Reznek, A. P., Miranda, J., Jarmin, R. S., and Abowd, J. M. (2011), "Towards unrestricted public use business microdata: The synthetic Longitudinal Business Database," *International Statistical Review*, 79, 363 - 384.

41. Mitra, R. and Reiter, J. P. (2011), "Estimating propensity scores with missing covariate data using general location mixture models," *Statistics in Medicine*, 30, 627 - 641.

42. Petrin, A., Reiter, J. P., and White, T. K. (2011), "The impact of plant-level resource reallocations and technical progress on U.S. macroeconomic growth," *Review of Economic Dynamics*, 14, 3 - 26. Also published as NBER Working Paper No. 16700.

43. Reiter, J. P. (2011), "Data confidentiality," *Wiley Interdisciplinary Reviews: Computational Statistics*, 3, 450 - 456.

44. Reiter, J. P. and Kinney, S. K. (2011), "Sharing confidential data for research purposes: A primer [invited commentary]," *Epidemiology*, 22, 632 - 635.

45. Schwartz-Bloom, R. D., Halpin, M. J., and Reiter, J. P. (2011), "Teaching high school chemistry in the context of pharmacology helps both teachers and students learn," *Journal of Chemical Education*, 88, 744 - 750.

46. Si, Y. and Reiter, J. P. (2011), "A comparison of posterior simulation and inference by combining rules for multiple imputation," *Journal of Statistical Theory and Practice*, 5, 335 - 347.

47. Burgette, L. and Reiter, J. P. (2010), "Multiple imputation via sequential regression trees," *American Journal of Epidemiology*, 172, 1070 - 1076.

48. Cailoa, G. and Reiter, J. P. (2010), "Random forests for generating partially synthetic, categorical data," *Transactions on Data Privacy*, 3:1, 27 - 42.

49. Drechsler, J. and Reiter, J. P. (2010), "Sampling with synthesis: A new approach to releasing public use microdata samples of census data," *Journal of the American Statistical Association*, 105, 1347 - 1357.

50. Kinney, S. K. and Reiter, J. P. (2010), "Tests of multivariate hypotheses when using multiple imputation for missing data and disclosure limitation," *Journal of Official Statistics*, 26, 301 - 315.

51. Kinney, S., Reiter, J. P., and Berger, J. O. (2010), "Model selection when multiple imputation is used to protect confidentiality in public use data," *Journal of Privacy and Confidentiality*, 2, Article 2.

52. Reiter, J. P. (2010), "Multiple imputation for disclosure limitation: Future research challenges," *Journal of Privacy and Confidentiality*, 1:2, Article 7.

53. Reiter, J. P. and Drechsler, J. (2010), "Releasing multiply-imputed, synthetic data generated in two stages to protect confidentiality," *Statistica Sinica*, 20, 405 - 422.

54. Zhou, X. and Reiter, J. P. (2010), "A note on Bayesian inference after multiple imputation," *The American Statistician*, 64, 159 - 163.

55. Drechsler, J. and Reiter, J. P. (2009), "Disclosure risk and data utility for partially synthetic data: An empirical study using the German IAB Establishment Survey," *Journal of Official Statistics*, 25, 589 - 603.

56. Karr, A. F., Lin, X., Reiter, J. P. and Sanil, A. P. (2009), "Privacy preserving analysis of vertically partitioned data using secure matrix protocols," *Journal of Official Statistics*, 25, 125 - 138.

57. Kinney, S. K. and Reiter, J. P. (2009), "Inferences for two stage multiple imputation for nonresponse," *Journal of Statistical Theory and Practice*, 3, 307 - 318.

58. Kohnen, C. N. and Reiter, J. P. (2009), "Multiple imputation for combining confidential data owned by two agencies," *Journal of the Royal Statistical Society, Series A*, 172, 511 - 528.

59. Marchenko, Y. V. and Reiter, J. P. (2009), "Improved degrees of freedom for multivariate significance tests obtained from multiply-imputed, small sample data," *The Stata Journal*, 9, 388 - 397.

60. Miranda, M. L., Kim, D., Reiter, J. P., Overstreet, M. A., Maxson, P. (2009), "Environmental contributors to the achievement gap," *NeuroToxicology*, 30, 1019 - 1024.

61. Reiter, J. P. (2009), "Using multiple imputation to integrate and disseminate confidential microdata," *International Statistical Review*, 77, 179 - 195.

62. Reiter J. P., Oganian A., and Karr A. F. (2009), "Verification servers: enabling analysts to assess the quality of inferences from public use data," *Computational Statistics and Data Analysis*, 53, 1475 - 1482.

63. Reiter, J. P. and Mitra, R. (2009), "Estimating risks of identification disclosure in partially synthetic data," *Journal of Privacy and Confidentiality*, 1.1, 99 - 110.

64. Woo, M., Reiter, J. P., Oganian, A., and Karr, A. F. (2009), "Global measures of data utility for microdata masked for disclosure limitation," *Journal of Privacy and Confidentiality*, 1.1, 111 - 124.

65. Montgomery, J. M., Cooper, A., Reiter, J. P., and Guan, S. (2008), "A comparison of respondents and non-respondents on dimensions of political activity," *International Journal of Public Opinion Research*, 20, 494 - 506.

66. Orr, S. T., James, S. A., and Reiter, J. P. (2008), "Unintended pregnancy and prenatal behaviors among urban, black women in Baltimore, Maryland: The Baltimore preterm birth study," *Annals of Epidemiology*, 18, 545 - 551.

67. Reiter, J. P. (2008), "Multiple imputation when records used for imputation are not used or disseminated for analysis," *Biometrika*, 95, 933 - 946.

68. Reiter, J. P. (2008), "Selecting the number of imputed datasets when using multiple imputation for missing data and disclosure limitation," *Statistics and Probability Letters*, 78, 15 - 20.

69. Reiter, J. P. (2008), "Statistics in sports: Current and future research trends," *STAtOR*, 9:2, 4 - 7.

70. Woo, M., Reiter, J. P., and Karr A. F. (2008), "Estimation of propensity scores using generalized additive models," *Statistics in Medicine*, 27, 3805 - 3816.

71. Dolber, P. C., Gu, B., Zhang, X., Fraser, M. O., Thor, K. B., and Reiter, J. P. (2007), "Activation of the external urethral sphincter central pattern generator by a 5-HT1A serotonin receptor agonist in rats with chronic spinal cord injury," *American Journal of Physiology: Regulatory, Integrative, and Comparative Physiology*, 292, R1699 - R1706.

72. Ghosh, J., Reiter, J. P., and Karr, A. F. (2007), "Secure computation with horizontally partitioned data using adaptive regression splines," *Computational Statistics and Data Analysis*, 51, 5813 - 5820.

73. Gu, B., Thor, K. B., Reiter, J. P., and Dolber, P. C. (2007), "Effect of 5HT-1 serotonin receptor agonists on noxiously stimulated micturition in cats with chronic spinal cord injury," *Journal of Urology*, 177, 2381 - 2385.

74. Karr, A. F., Fulp, W. J., Vera, F., Young, S. S., Lin, X., and Reiter, J. P. (2007), "Secure privacy-preserving analysis of distributed databases," *Technometrics*, 49, 335 - 345.

75. Kwiek, N. C., Halpin, M. J., Reiter, J. P., Hoeffler, L. A., and Schwartz-Bloom, R. D. (2007), "Pharmacology in the high school classroom," *Science*, 317 (September 2007), 1871 - 1872 (with supplemental material).

76. Orr, S. T., Blazer, D. G., James, S. A., and Reiter, J. P. (2007), "Depressive symptoms and indicators of maternal health status during pregnancy," *Journal of Women's Health*, 16, 535 - 542.

77. Orr, S. T., Reiter, J. P., Blazer, D. G., and James, S. A. (2007), "Maternal prenatal pregnancy-related anxiety and spontaneous preterm birth in Baltimore, Maryland," *Psychosomatic Medicine*, 69, 566 - 570.

78. Reiter, J. P. (2007), "Small-sample degrees of freedom for multi-component significance tests with multiple imputation for missing data," *Biometrika*, 94, 502 - 508.

79. Reiter, J. P. and Raghunathan, T. E. (2007), "The multiple adaptations of multiple imputation," *Journal of the American Statistical Association*, 102, 1462 - 1471.

80. Hill, J. L. and Reiter, J. P. (2006), "Interval estimation of treatment effects when using propensity score matching," *Statistics in Medicine*, 25:13, 2230 - 2256.

81. Karr, A. F., Kohnen, C. N., Oganian, A., Reiter, J. P. and Sanil, A. P. (2006), "A framework for evaluating the utility of data altered to protect confidentiality," *The American Statistician*, 60, 224 - 232.

82. Reiter, J. P., Raghunathan, T. E., and Kinney, S. (2006), "The importance of the sampling design in multiple imputation for missing data," *Survey Methodology*, 32.2, 143 - 150.

83. Gomatam, S., Karr, A. F., Reiter, J. P., and Sanil, A. P. (2005), "Data dissemination and disclosure limitation in a world without microdata: A risk-utility framework for remote access servers," *Statistical Science*, 20, 163 - 177.

84. Karr, A. F., Feng, J., Lin, X., Sanil, A. P., Young, S. S., and Reiter, J. P. (2005), "Secure analysis of distributed chemical databases without data integration," *Journal of Computer-Aided Molecular Design*, 19, 739 - 747.

85. Karr, A. F., Lin, X., Sanil, A. P., and Reiter, J. P. (2005), "Secure regressions on distributed databases," *Journal of Computational and Graphical Statistics*, 14, 263 - 279.

86. Reiter, J. P. (2005), "Estimating risks of identification disclosure for microdata," *Journal of the American Statistical Association*, 100, 1103 - 1113.

87. Reiter, J. P. (2005), "Releasing multiply-imputed, synthetic public use microdata: An illustration and empirical study," *Journal of the Royal Statistical Society, Series A*, 168, 185 - 205.

88. Reiter, J. P. (2005), "Significance tests for multi-component estimands from multiply-imputed, synthetic microdata," *Journal of Statistical Planning and Inference*, 131 (2), 365 - 377.

89. Reiter, J. P. (2005), "Using CART to generate partially synthetic public use microdata," *Journal of Official Statistics*, 21, 441 - 462.

90. Reiter, J. P. and Kohnen, C. N. (2005), "Categorical data regression diagnostics for remote servers," *Journal of Statistical Computation and Simulation*, 75, 889 - 903.

91. Reiter, J. P., Zanutto, E. L., and Hunter, L. W. (2005), "Analytical modeling in complex surveys of work practices," *Industrial Labor Relations Review*, 89, 82 - 100.

92. Gu, B., Olejar, K., Reiter, J. P., Thor, K. B., and Dolber, P. C. (2004), "Inhibition of bladder activity by 5HT1 serotonin receptor antagonists in cats with chronic spinal cord injury," *Journal of Pharmacology and Experimental Therapeutics*, 310:3, 1266 - 1272.

93. Gu, B., Reiter, J. P., Schwinn, D. A., Smith, M. P., Korstanje, C., Thor, K. B., and Dolber, P. C. (2004), "Effects of alpha1-adrenergic receptor subtype selective antagonists on lower urinary tract function in rats with bladder outlet obstruction," *Journal of Urology*, 172, 758 - 762.

94. Karr, A. F., Lin, X., Sanil, A. P., and Reiter, J. P. (2004), "Analysis of integrated data without data integration," *Chance*, 17:3 (Summer 2004), 27 - 30.

95. Reiter, J. P. (2004), "New approaches to data dissemination: A glimpse into the future," *Chance*, 17:3 (Summer 2004), 12 - 16.

96. Reiter, J. P. (2004), "Should teams walk or pitch to Barry Bonds?" *Baseball Research Journal*, 32, 63 - 69.

97. Reiter, J. P. (2004), "Simultaneous use of multiple imputation for missing data and disclosure limitation," *Survey Methodology*, 30, 235 - 242.

98. Sanil, A. P., Karr, A. F., Lin, X., and Reiter, J. P. (2004), "Privacy preserving regression modeling via distributed computation," *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 677 - 682.

99. Raghunathan, T. E., Reiter, J. P., and Rubin, D. B. (2003), "Multiple imputation for statistical disclosure limitation," *Journal of Official Statistics*, 19, 1 - 16.

100. Reiter, J. P. (2003), "Inference for partially synthetic, public-use microdata sets," *Survey Methodology*, 29, 181 - 188.

101. Reiter, J. P. (2003), "Model diagnostics for remote access regression servers," *Statistics and Computing*, 13, 371 - 380.

102. Reiter, J. P. (2002), "Satisfying disclosure restrictions with synthetic data sets," *Journal of Official Statistics*, 18, 531 - 544.

103. Reiter, J. P. (2000), "Borrowing strength when explicit data pooling is prohibited," *Journal of Official Statistics*, 16, 295 - 319.

104. Reiter, J. P. (2000), "Using statistics to determine causal relationships," *The American Mathematical Monthly*, 107, 24 - 32.

## BOOK CHAPTERS AND INVITED PUBLICATIONS

105. Hu, J., Reiter, J. P., and Wang, Q. (2014), "Disclosure risk evaluation for fully synthetic data," *Privacy in Statistical Databases*, edited by J. Domingo-Ferrer, Lecture Notes in Computer Science 8744, Heidelberg: Springer, 185 - 199.

106. Karr, A. F. and Reiter, J. P. (2014), "Using statistics to protect privacy," in *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, edited by J. Lane, V. Stodden, S. Bender, and H. Nissenbaum, Cambridge University Press, 276 - 295.

107. Reiter, J. P. (2012), "Discussion: Bayesian perspectives and disclosure risk assessment," *International Statistical Review*, 80, 373 - 375.

108. Machanavajjhala, A. and Reiter, J. P. (2012), "Big privacy: Protecting confidentiality in big data," *XRDS Crossroads*, 19:1, 20 - 23.

109. Reiter, J. P. (2012), "Protecting data confidentiality in public release datasets: Approaches based on multiple imputation," *Handbook of Statistics: Bioinformatics*, edited by C. R. Rao, R. Chakraborty, and P. K. Sen, Elsevier, 533 - 545.

110. Reiter, J. P. (2012), "Statistical disclosure limitation research and practice: Fascinating and growing areas of importance," *CHANCE*, vol. 25:1.

111. Reiter, J. P. (2011), "Commentary on article by Gates," *Journal of Privacy and Confidentiality*, vol 3:2.

112. Reiter, J. P. (2011), "Statistical approaches to protecting confidentiality in public use data," *International Encyclopedia of Statistical Science*, edited by M. Lovric, Springer, 1386 - 1388. Reprinted online in *StatProb: The Encyclopedia Sponsored by Statistics and Probability Societies*.

113. Reiter, J. P. (2008), "Protecting data confidentiality in public release datasets: Approaches based on multiple imputation," *The Imputation Bulletin*, 8:2, 1 - 6.

114. Drechsler, J. and Reiter, J. P. (2008), "Accounting for intruder uncertainty due to sampling when estimating identification disclosure risks in partially synthetic data," *Privacy in Statistical Databases (Lecture Notes in Computer Science 5262)*, edited by J. Domingo-Ferrer and Y. Saygin, 227 - 238.

115. Reiter, J. P. (2007), "Survey error," *Encyclopedia of Statistics in Quality and Reliability*, edited by F. Ruggeri, R. Kennett, and F. W. Faltin, Chichester, UK: John Wiley & Sons, 1952 - 1954.

116. Karr, A. F., Lin, X., Reiter, J. P., and Sanil, A. P. (2006), "Methods of secure computation and data integration," *Monographs of Official Statistics: Worksession on Statistical Data Confidentiality*, edited by P. Diaz Munoz and H. Brungger, Eurostat, 217 - 226.

117. Karr, A. F., Lin, X., Sanil, A. P., and Reiter, J. P. (2006), "Secure statistical analysis of distributed databases," *Statistical Methods in Counterterrorism: Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication*, edited by A. Wilson, G. Wilson, and D. H. Olwell, New York: Springer, 237 - 262.

118. Mitra, R. and Reiter, J. P. (2006), "Adjusting survey weights when altering identifying design variables via synthetic data," *Privacy in Statistical Databases (Lecture Notes in Computer Science)*, edited by J. Domingo-Ferrer and L. Franconi, Berlin: Springer, 177 - 188.

119. Reiter, J. P. (2006), "Multiple imputation for missing data," *Encyclopedia of Measurement and Statistics*, edited by N. Salkind, Sage Publications..

120. Banks, D. and Reiter, J. P. (2005), "Confidentiality issues related to transportation use of census data for transportation planning: Preparing for the future," commissioned paper for the Transportation Research Board.

121. Hill, J. L., Reiter, J. P., Zanutto, E. (2004), "A comparison of experimental and observational data analyses," *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, edited by A. Gelman and X. Meng, New York: Wiley, 49 - 60.

122. Reiter, J. P. (2003), "Estimating probabilities of identification for microdata," commissioned paper for the National Academy of Sciences Panel on Confidential Data Access for Research Purposes.

123. Reiter, J. P. (2001), "Borrowing strength without explicit data pooling," in *Monographs of Official Statistics: Bayesian Methods With Applications to Science, Policy, and Official Statistics*, edited by E. George and N. Photis, Eurostat, 439 - 448.

## OTHER PUBLICATIONS

124. Reiter, J. P. (2011), "Introduction to Chapter 5," in *Advances in Quantitative Methods*, edited by C. R. Rao and S. Gupta, Grace Scientific Publishing, 297 - 298.

125. Reiter, J. P. (2011), Book review of "Bayesian Ideas and Data Analysis," *The American Statistician*, 65, 284.

126. National Research Council (2010), *Conducting Biosocial Surveys: Collecting, Storing, Accessing, and Protecting Biospecimens and Biodata*, edited by R. Hauser, M. Weinstein, R. Pool, and B. Cohen, Panel on Collecting, Storing, Accessing, and Protecting Biological Specimens and Biodata in Social Surveys, National Academies Press. (Contributing author.)

127. National Research Council (2009), *Reengineering the Survey of Income and Program Participation*, edited by C. Citro and J. Scholz, Panel on the Census Bureau's Reengineered Survey of Income and Program Participation, National Academies Press. (Contributing author.)

128. Reiter, J. P. (2008), "Letter to the editor," *Journal of Official Statistics*, 24, 319 - 320.

129. Reiter, J. P. (2008), Book review of "Data Quality and Record Linkage Techniques," *Journal of the American Statistical Association*, 103, 881.

130. Kohnen, C. N. and Reiter, J. P. (2007), "Manual for JMP-IN." Accompanies *Mind on Statistics: Third Edition* by J. Utts and R. Heckard.

131. National Research Council (2007), *Putting People on the Map,* edited by M. Gutman and P. Stern, Panel on Confidentiality Issues Arising From the Integration of Remotely-Sensed and Self-Identifying Data, National Academies Press. (Contributing author.)

132. Reiter, J. P. (2007), Book review of "Statistical Matching: Theory and Practice," *Journal of Official Statistics*, 23, 130 - 132.

133. Reiter, J. P. (2006), "Teaching the fundamentals of statistics with sports data," *Proceedings of the International Conference on Teaching Statistics*.

134. Stangl, D., Banks, D., House, L., and Reiter, J. P. (2006), "Progressive mastery testing: Does it increase learning and retentions? Yes and no," *Proceedings of the International Conference on Teaching Statistics*.

135. Reiter, J. P. (2005), "Bayesian methods in statistical disclosure limitation," *ISBA Bulletin*, 12(4), 6.

136. Biemer, P. and Reiter, J. P. (2005), "Total survey error workshop," *Newsletter of the ASA Survey Research Methods Section*, 21 (July), 2 - 3.

137. Kohnen, C. N. and Reiter, J. P. (2004), "Sharing confidential data among multiple agencies using multiply imputed, synthetic data," *ASA Proceedings of the Joint Statistical Meetings.*

138. Reiter, J. P. (2004), "One question interview: Walking Barry Bonds," *Duke Dialogue*, 19:5 (May 28, 2004), 7.

139. Reiter, J. P. (2004), Book review of "Teaching Statistics Using Baseball," *INFORMS Transactions on Education*, 5 (Sept. 2004).

140. Reiter, J. P. (2004), "Introductory statistics animations," Wiley. Flash animations of statistical topics to accompany introductory statistics text books.

141. Reiter, J. P., Kohnen, C. N., Karr, A. F., Lin, X., and Sanil, A. P. (2004), "Secure regression for vertically partitioned, partially overlapping data," *ASA Proceedings of the Joint Statistical Meetings.*

142. Reiter, J. P. (2003), "Using CART to generate partially synthetic, public use microdata," *Proceedings of the Federal Committee on Statistical Methodology Research Conference.*

143. Reiter, J. P. and Kohnen, C. N. (2003), "Generalized linear model diagnostics for remote servers," *Proceedings of the Association for Survey Computing Conference*, 2003.

144. Reiter, J. P. (2002), Book reviews of "Finite Population Sampling and Inference: A Prediction Approach" and of "Topics in Survey Sampling," *Journal of the American Statistical Association*, 97, 357 - 358.

145. Reiter, J. P. (2002), Book review of "Fundamentals of Modern Statistical Methods," *Journal of Official Statistics*, 18, 313 - 315.

146. Reiter, J. P. (2002), "Should teams walk or pitch to Barry Bonds?" *By the Numbers*, 12 (November 2002), pp. 7-11. (Newsletter of the Statistical Analysis Committee of the Society for American Baseball Research).

147. Reiter, J. P. (2001), "Using synthetic data sets to satisfy disclosure restrictions," *Proceedings of the Federal Committee on Statistical Methodology Research Conference.*

148. Reiter, J. P. (2001), "Motivating students' interest in statistics through sports," *ASA Proceedings of the Joint Statistical Meetings.*

149. Reiter, J. P. (2000), "Satisfying disclosure restrictions with synthetic data sets," *ASA Proceedings of the Section on Government Statistics and Section on Social Statistics*, 276 – 281.

150. Reiter, J. P. (1999), "Stimulating interest in quantitative courses with activity-based lessons," in *Voices of Experience*, edited by M. Winkelmes and J. Wilkinson, Derek Bok Center for Teaching and Learning, Harvard University, 47 - 54.

151. Reiter, J. P. (1998), "Estimation in the presence of external constraints that prohibit explicit data pooling," *ASA Proceedings of the Section on Survey Research Methods*, 599 – 604.

## MANUSCRIPTS UNDER REVIEW

De Yoreo, M., Reiter, J. P., and Hillygus, D. S., "Nonparametric Bayesian models with focused clustering for mixed ordinal and nominal data," submitted to *Bayesian Analysis.*

Fosdick, B. K., De Yoreo, M. and Reiter, J. P. "Categorical data fusion using auxiliary data," submitted to *Annals of Applied Statistics.*

Hu, J., Reiter, J. P., and Wang, Q. "Dirichlet process mixture models for nested categorical data," submitted to *Bayesian Analysis.*

Kim, H. J., Reiter, J. P., and Karr, A. F. "Simultaneous edit-imputation and disclosure limitation for business establishment data," submitted to *Journal of the Royal Statistical Society, Series A.*

Murray, J. S. and Reiter, J. P. "Multiple imputation of missing categorical and continuous values via Bayesian mixture models with local dependence," submitted to *Journal of the American Statistical Association.*

Manrique-Vallier, D. and Reiter, J. P. "Bayesian simultaneous edit and imputation for multivariate categorical data," submitted to *Journal of the American Statistical Association.*

Paiva, T. and Reiter, J. P. "Stop or continue data collection: A nonignorable missing data approach for continuous variables," submitted to *Journal of Official Statistics.*

White, T. K., Reiter, J. P. and Petrin, A. "Plant-level productivity and imputation of missing data in U.S. Census manufacturing data," submitted to *Review of Economics and Statistics.*

## INVITED PRESENTATIONS AT PROFESSIONAL CONFERENCES

"Bayesian modeling and multiple imputation to simultaneously handle missing and erroneous values and protect confidentiality in Census Bureau data," 65th Birthday Conference for Rod Little, University of Michigan, October 2015.

"Relationships between data quality and confidentiality," Total Survey Error Workshop, Baltimore, MD September 2015.

"Statistical methods for protecting confidentiality of data with personally identifiable information," Conference on Big Data and HIV, NIAID, Rockville, MD, July 2015.

"Making large-scale, confidential data available for secondary analysis," Fields Institute, Toronto, Canada, April 2015.

"Protecting confidentiality in an era with no privacy," AISC 2014, Greensboro, NC, October 2014.

"Generating and releasing synthetic data: Lessons learned and future directions," MITRE workshop, Mclean, VA, July 2014.

"Providing public access to confidential, big social science data," ISNIE conference, Duke University, NC, June 2014.

"Multiple imputation via flexible, joint models," ENAR conference, Baltimore, MD, March 2014.

"Statistical disclosure limitation and edit imputation," Federal Committee on Statistical Methodology research conference, Washington DC, November 2013.

"Protecting data confidentiality in an era without privacy," Plenary talk, UNC-G Regional Mathematics and Statistics Conference, November 2013.

"Multiple imputation: Theory and practice," Workshop, University of Maryland Baltimore County, April 2013.

"Data fusion via multiple imputation," Joint Statistical Meetings, San Diego, CA, August 2012.

"The multiple adaptations of multiple imputation," Plenary talk at Political Methodology XII Conference, Chapel Hill, NC, July 2012.

"Protecting confidentiality of public use data: Common approaches and their impacts on statistical analysis," Academy Health Annual Research Meeting, Orlando, FL, June 2012.

"Public use business establishment data: Protecting confidentiality and providing utility with synthetic data," International Conference on Establishment Surveys IV, Montreal, Canada, June 2012.

"Adapting multiple imputation to protect confidentiality," International Chinese Statistical Association Applied Statistics Symposium, Indianapolis, June 2010.

"Statistical methods for disclosure risk assessment," Office of Civil Rights Workshop on Reconsidering the HIPAA De-identification Standard, Washington DC, March 2010.

"Reconciling measures of confidentiality risk from statistics and computer science," IPAM workshop on Privacy and Statistical Learning, UCLA, February 2010.

"Using multiple imputation to protect participants' confidentiality when sharing data," International Conference on Health Policy Statistics, Washington DC, January 2010.

"Easily implemented, nonparametric synthesizers based on algorithmic methods from computer science," NSF-Census-IRS workshop on synthetic data and confidentiality protection, Washington, DC, July 2009.

"Multiple imputation when some records used for imputation are not used or disseminated for analysis," Conference on Sample Surveys and Bayesian Statistics, Southampton, England, August 2008.

"Synthetic data methods: A discussion of the most pressing challenges," Joint Meeting of the Statistical Society of Canada, May 2008.

"The future of synthetic data," NISS workshop on Data Confidentiality: The Next Five Years, Washington DC, May 2008.

"Evaluating the disclosure risks of reporting quality measures to the public," United Nations Economic Commission for Europe workshop on data confidentiality, Manchester, England, December 2007.

"Using two stage multiple imputation to correct measurement error with external validation data," International Statistical Institute meetings, Lisbon, Portugal, August 2007.

"The risks and benefits of releasing measures of data quality," Workshop on Data Access, Nuremberg Germany, August 2007.

"Protecting confidentiality in public use data by releasing synthetic datasets: Experiences with genuine applications," invited panel, Joint Statistical Meetings, Salt Lake City, August 2007.

"Alternative approaches to data dissemination," Population Association of America conference, New York, March 2007.

"Using multiple imputation for data integration and dissemination," Workshop on Combining and Enhancing Data, Manchester, UK, January 2007.

"Adjusting survey weights when using partially synthetic data," Privacy in Statistical Databases conference, Rome, Italy, December 2006.

"Making public use, synthetic files of longitudinal establishment data," Comparative Analysis of Enterprise Data conference, Washington, DC, September 2006.

"Using multiple imputation to handle missing data and disclosure limitation," invited talk, Workshop on data confidentiality, Nurnberg, Germany, August 2006.

"Can and should we teach Bayesian statistics in Stat 101?" roundtable coffee, Joint Statistical Meetings, August 2006.

"Discussion of session on statistical disclosure limitation," Joint Statistical Meetings, August 2006.

"Identification disclosure in social science research," Data Sharing Workshop for Behavioral and Social Studies that Collect Genetic Data, National Institute on Aging, August 2006.

"Teaching the fundamentals of statistics with sports data," International Conference on Teaching Statistics, Brazil, July 2006.

"Alternative approaches to data dissemination and data sharing," Conference on Quantitative Methods and Statistical Applications in Defense and National Security, RAND, February 2006.

"Methods for secure computation and data integration," United Nations Economic Commission for Europe Workshop on Data Confidentiality, Geneva, Switzerland, November 2005.

"Some approaches and challenges in secure statistical analysis," Statistical and Applied Mathematical Sciences Institute Workshop on National Defense and Homeland Security, September 2005.

"Methods for secure computation and data integration," Joint Statistical Meetings, Minneapolis, August 2005.

"Generating partially synthetic data when there are missing data," Computer Science-Statistics Workshop on Privacy and Confidentiality, Italy, July 2005.

"Discussion of the National Election Surveys: The potential of multiple imputation," panel, Social Science Research Institute, Duke University, January 2005.

"Disclosure limitation via synthetic, multiply-imputed data sets," International Conference on the Future of Statistical Theory, Practice and Education, Hyderabad, India, December 2004.

"Disclosure limitation via synthetic, multiply-imputed data sets," Research Triangle Institute Fellows Symposium, Durham, NC, December 2004.

"Multiple imputation for missing data in surveys with complex designs, " Statistical and Applied Mathematical Sciences Institute Workshop on Latent Variables in the Social Sciences, September 2004.

"Disclosure risk and data utility for remote access servers," Joint Statistical Meetings, Toronto, August 2004.

"Disclosure limitation via synthetic, multiply-imputed data sets." International Indian Statistical Association conference, University of Georgia, May 2004.

"Data confidentiality, data integration, and secure computation," Electronic Frontier Foundation panel, National Institute of Statistical Sciences, April 2004

"Disclosure limitation via partially synthetic, multiply-imputed data sets," International Biometric Society (ENAR) Annual Meeting, Pittsburgh, March 2004.

"Is it better to walk or pitch to Barry Bonds?" Institute of Mathematical Statistics Meeting on Statistics and Sports, Worcester Polytechnic Institute, November 2003.

"Disclosure limitation by releasing multiply imputed, public-use microdata," New Researchers Conference, Davis, CA., July 2003.

"Estimating probabilities of identification for microdata," National Academy of Sciences Panel on Data Access workshop, Washington DC, October 2003.

"Borrowing strength without explicit data pooling," International Society for Bayesian Analysis Conference, Greece, June 2000.

## OTHER PRESENTATIONS

"Data access and data confidentiality: Where are we and where are we going?" Commissioner's Invited Lecture Series, Bureau of Labor Statistics, December 2015.

"A simple way to incorporate prior information on margins in Bayesian latent class models," Dept. of Biostatistics, UNC Chapel Hill, October 2015.

"Protecting confidentiality in an era with no privacy," BSSR Lecture Series, National Institutes of Health, April 2015.

"Multiple imputation via Bayesian mixture models," NISS workshop on missing data, Bureau of Labor Statistics, Washington, DC, October 2014.

"Imputation of missing data via flexible mixture modeling," AISC 2014, Greensboro, NC, October 2014.

"An integrated approach to providing access to confidential social science data," Center for Improving Methods for Quantitative Policy Research, Northwestern University, October 2014.

"Sharing confidential data in era with no privacy," Gertrude Cox award lecture, Washington Statistical Society, Washington DC, June 2014.

"Missing data in longitudinal studies: The case for refreshment samples," Helen Barton Lecture Series in Computational Mathematics, University of Greensboro, Dept. of Mathematics and Statistics, April 2014.

"Bayesian methods for complex surveys," Research Triangle Institute International, January 2014.

"Multiple adaptations of multiple imputation," Dept. of Biostatistics, Brown University, April 2013.

"Protecting confidentiality by releasing simulated public use data sets," Institute for Quantitative Social Science, Harvard University, April 2013.

"Protecting confidentiality in public use data with multiple imputation," Dept. of Political Science, Princeton University, March 2013.

"Multiple imputation: Some novel applications," Prevention Science and Methodology Group, Durham, NC (remote talk), December 2012.

"Protecting confidential data with geographic identifiers: Approaches based on multiple imputation," Dept. of Statistics, University of Missouri Columbia, November 2012.

"Missing data in longitudinal studies: The case for refreshment samples," Center for Developmental Science, UNC-Chapel Hill, October 2012.

"An overview of the challenges of statistical disclosure limitation," AISC conference, UNC-Greensboro, October 2012.

"An overview of multiple imputation: Theory and practice," Research Triangle Institute International, March 2012.

"Protecting confidentiality in public use data with geographic identifiers," National Cancer Institute, August 2011.

"Some novel adaptations of multiple imputations," Department of Mathematics and Statistics, University of Maryland–Baltimore County, November 2010.

"Understanding synthetic data," short course, Centre for Census and Survey Research, University of Manchester, UK, October 2010.

"Some novel adaptations of multiple imputation," Centre for Census and Survey Research, University of Manchester, UK, October 2010.

"Sampling with synthesis: A new approach to releasing public use microdata samples of census data," Summer at Census talk, U. S. Bureau of the Census, Washington, DC., July 2010.

"Novel adaptations of multiple imputations," Department of Statistics, The Ohio State University, May 2010.

"Measuring disclosure risks in statistical databases," Westat, March 2010.

"Handling missing data with multiple imputation: A primer for quantitative social scientists," Quantitative Psychology Program, UNC Chapel Hill, September 2009.

"Issues in disclosure risk assessment for tabular data," CNSTAT workshop on confidentiality criteria for statistics from the NSF Survey of Earned Doctorates, National Academy of Sciences, May 2009.

"Multiple imputation when some records used for imputation are not used or disseminated for analysis," Department of Biostatistics, Duke University, December 2008.

"Multiple imputation when some records used for imputation are not used or disseminated for analysis," Social Science Research Institute, Duke University, September 2008.

"Preserving confidentiality in shared data: A statistician's perspective," Department of Computer Science, UNC Chapel Hill, August 2008.

"The multiple adaptations of multiple imputation," Department of Statistics, University of Wollongong, Australia, August 2008.

"Using multiple imputation to protect confidentiality in public use data," Australian Bureau of Statistics, July 2008.

"Multiple imputation when some records used for imputation are not used or disseminated for analysis," Australian Bureau of Statistics, July 2008.

"The multiple adaptations of multiple imputation," Heinz School, Carnegie Mellon University, March 2008.

"Protecting confidentiality of public use genetic data: What can we learn from the experiences of statistical agencies," invited talk, Department of Biostatistics, University of Alabama at Birmingham, January 2008.

"The multiple adaptations of multiple imputation," Center for Statistics in the Social Sciences, University of Washington, November 2007.

"Analysis of missing data in social science research," short course, Social Sciences Research Institute, Duke University, September 2007, February 2007, April 2006.

"Estimating risks of identification disclosure with partially synthetic data," Bureau of the Census, Washington DC, June 2007.

"Analysis with missing data," Center for Health Promotion and Disease Prevention, UNC Chapel Hill, April 2007.

"Using multiple imputation to protect confidentiality in public use data," Center for Demography and Health of Aging, University of Wisconsin, February 2007.

"Multiple imputation for disclosure limitation: Why don't Rubin's 1987 rules apply?" IAB, Nurnberg, Germany, January 2007.

"Discussion of papers on synthetic data," Washington Statistical Society seminar series, Washington, D.C., November 2006.

"The multiple applications of multiple imputation," University of Pennsylvania, October 2006.

"Disclosure limitation in microdata using multiple imputation," National Center for Education Statistics, Washington, D.C., July 2006.

"Disclosure limitation for ACS microdata using multiple imputation," Bureau of the Census, Washington, D.C., June 2006.

"The multiple applications of multiple imputation," Department of Statistics, University of South Carolina, April 2006.

"Handling confidentiality concerns with multiple imputation," University of Southampton, March 2006.

"Handling confidentiality concerns and missing data simultaneously with multiple imputation," Department of Statistics, Cornell University, February 2006.

"Multiple imputation for disclosure limitation: Why don't Rubin's 1987 rules apply," Center for Economic Studies, U.S. Bureau of the Census, October 2005.

"Disclosure limitation via synthetic, multiply-imputed data sets," Department of Statistics, Pennsylvania State University, September 2005.

"Are we teaching the right things in introductory statistics," poster session, U.S. Conference on Teaching Statistics, May 2005.

"Discussion of the National Election Surveys: The potential of multiple imputation," panel discussion, Social Science Research Institute, Duke University, January 2005.

"What do students retain from introductory statistics: Lessons from a study in statistical education," Teaching and Learning Club, Duke University, October 2004.

"Disclosure limitation via synthetic, multiply-imputed data sets," Los Alamos National Labs, August 2004.

"Protecting confidentiality by releasing simulated microdata: An informal discussion of benefits, limitations, and practical experiences," roundtable luncheon, Joint Statistical Meetings, Toronto, August 2004.

"Disclosure limitation via synthetic, multiply-imputed data sets," Department of Statistics, Iowa State University, April 2004.

"Protecting confidentiality by releasing synthetic, public-use datasets," Social Science Research Institute, Duke University, March 2004.

"A primer on designing studies to determine causal effects," Teaching and Learning Club, Duke University, February 2004.

"Teaching quantitative material," Center for Teaching, Learning, and Writing Teaching Breakfast, Duke University, February 2004.

"Releasing partially synthetic, public-use microdata to protect confidentiality: An investigation of data generation by CART," Federal Committee on Statistical Methodology Research Conference, Washington DC, November 2003.

"Protecting confidentiality by releasing synthetic microdata," INFORMS Annual Meeting, Atlanta, October 2003.

"Model diagnostics for remote access servers," Association for Survey Computing Conference, Warwick, England, September 2003.

"Inference for multiply-imputed, partially synthetic microdata," Joint Statistical Meetings, San Francisco, August 2003.

"The feasibility of using synthetic microdata for public release data," Joint Statistical Meetings, August 2002.

"The feasibility of using synthetic microdata for public release data, " Federal Committee on Statistical Methodology Conference, November 2001.

"Motivating students' interest in statistics through sports," Joint Statistical Meetings, August 2001.

"Satisfying disclosure restrictions with synthetic data sets," U.S. Bureau of the Census, May 2001.

"Satisfying disclosure restrictions with synthetic data sets," Joint Program in Survey Methodology, April 2001.

"Satisfying disclosure restrictions with synthetic data sets," Sigma Xi Honor Society Research Lecture, Williams College, April 2001.

"What is Bayesian statistics?" Sigma Xi Honor Society Research Lecture, Williams College, April 2001.

"Borrowing strength without explicit data pooling," contributed talk, Interface 2000 Conference, New Orleans, April 2000.

# Jana Schaich Borg, PhD

Duke University, Department of Psychiatry and Behavioral Sciences
361 Bryan Research Building, Durham, NC 27710
janaschaichborg@gmail.com

---

## RESEARCH SUMMARY

I use high-dimensional neural recordings in rodents, functional magnetic resonance imaging in human participants, and computational models of neural data to understand how and why we make decisions in social settings. I use the insights gleaned from this research to develop strategies to improve social interactions and optimize the outcome of decisions that affect other people.

---

## EDUCATION

**Duke University**                                                                                                     **Durham, NC**
    Postdoctoral Associate                                                                          **(present)**
    Postdoctoral advisor: Kafui Dzirasa

**Stanford University**                                                                                                 **Palo Alto, CA**
    PhD in Neuroscience                                                                             **degree awarded January, 2013**
    Dissertation: **"**Uncovering the Neural Correlates of Intersubjective Avoidance in a Novel Rat Model"
    Advisors: Luis de Lecea (primary) and William Newsome (secondary)

**Dartmouth College**                                                                                                   **Hanover, NH**
    Bachelor of Arts in Philosophy of Neuroscience (custom designed major), with Honors   **degree awarded October 2002**

**Université Lumière II**                                                                                               **Lyon, France**
    Education exchange in French literature                                                          **Spring 2000**

**University of Edinburgh**                                                                                             **Edinburgh, Scotland**
    Education exchange in philosophy                                                                **Fall 2000**

---

## RESEARCH EXPERIENCE

**Department of Psychiatry and Behavioral Sciences, and Kenan Institute for Ethics at Duke University**       **Durham, NC**
*Laboratory of Kafui Dzirasa, MD, PhD*                                                                          **January 2013-present**

- Designed wireless headstage that permits up to 31 channels of simultaneous electrical recordings from the brains of pair-housed rats
- Conceptualized and implemented strategy to use the elastic net penalized regression framework and sparse principal components analysis to infer spatially-distributed neural networks involved in rat empathic avoidance
- Devised and performed a new surgical strategy that permits electrical recordings from 22 brain areas of the mouse brain simultaneously (previously published studies have recorded from a maximum of nine areas simultaneously)
- Created *in vivo* neuroimaging method for localizing metal electrodes implanted into the rat brain through registering CT images to a high-resolution MR histology atlas
- Co-authored Duke Medicine "Autism and Beyond" app hosted by the Apple App Store that uses automated machine learning of videos collected through mobile devices to improve early diagnosis of human neuropsychiatric disease
- Developed novel computational methods for inferring dynamic brain networks in collaboration with Drs. David Dunson (Statistical Science) and Lawrence Carin (Electrical and Computer Engineering)
- Co-designing a "Moral Artificial Intelligence" using machine learning and game theoretic models in collaboration with Drs. Vincent Conitzer (Computer Science) and Walter Sinnott-Armstrong (Philosophy)

**Departments of Psychiatry and Neurobiology at Stanford University**                                      **Stanford, CA**
*Laboratory of Luis de Lecea, PhD and William Newsome, PhD*                                   **October 2006-December 2012**
  - Created five new tests of mice and rodent social decision-making, including tests of empathy and altruism
  - Designed and programmed new method for automated histological cell counting across large areas of tissue
  - Discovered social decision-making circuit in rats that overlaps with the social decision-making circuit in humans using self-designed behavioral tests and histological cell counting methods
  - Developed and implemented *in vivo* optogenetic and DREADDs strategies to manipulate selective neural circuits in awake behaving rats and mice
  - Collected electrophysiological recording in awake behaving monkeys performing value-based decisions
  - Designed and implemented neuroimaging and behavioral paradigms to test neural dysfunction during social decision-making in incarcerated human psychopaths

**Olin Neuropsychiatry Research Center, Institute of Living at Hartford Hospital/Yale University**       **Hartford, CT**
*Research Assistant and Lab Manager, laboratory of Dr. Kent Kiehl*                                   **Summer 2004 – Summer 2006**
  - Designed, implemented, and analyzed behavioral, event-related potential (ERP), electroencephalography (EEG), and functional magnetic resonance imaging (fMRI) experiments to identify emotional and decision-making deficits in psychopathic, schizophrenic, bipolar, and cocaine-abusing clinical populations
  - Managed recruitment of, and multi-modal data collection from, over 300 human participants
  - Managed and implemented collaborative project to use EEG and ERP data to evaluate linguistic abilities of Beluga whales at Mystic Aquarium

**Kewalo Basin Marine Mammal Laboratory**                                                            **Honolulu, HI**
*Intern, laboratory of Louis Herman, PhD*                                                        **Summer 2003 – Winter 2004**
  - Designed methods for testing working memory, language, numerosity, inhibition, imitation, and theory of mind in dolphins

**Department of Philosophy and Department of Psychological and Brain Sciences**                       **Hanover, NH**
**at Dartmouth College**                                                                        **Fall 2002 –Summer 2003**
*Research Assistant under Walter Sinnot-Armstrong, PhD, Scott Grafton, MD, Michael Gazzaniga, PhD*     **Winter 2004**
*Todd Handy, PhD, and David Turk, PhD*
  - Designed, implemented, and analyzed behavioral, event-related potential (ERP), electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and split-brain experiments to investigate the role of emotion and cognition in moral reasoning, the neural substrates of "the self", and the function of attention in motor activation elicited by graspable objects
  - Junior Fellow, "The Psychology and Biology of Morality" Dartmouth Humanities Center Institute, April-May 2004. A 10-week institute that brought together 10 interdisciplinary scholars to discuss biological and philosophical insights into the human capacity for moral reasoning

**Senior Honor's Thesis at Dartmouth College, Department of Philosophy and Department of**             **Hanover, NH**
**Psychological and Brain Sciences**                                                            **Fall 2001-Summer 2002**
*Advisors: Walter Sinnot-Armstrong, PhD, and Scott Grafton, MD*
  - Designed, implemented, and analyzed behavioral and functional magnetic resonance imaging (fMRI) experiment to assess the role of consequences, action, and intention on explicit moral judgments

**Brookhaven National Laboratory, Chemistry Department**                                              **Upton, NY**
*Research Assistant under Stephen Dewey, PhD, Madina Gerasimov, MD, Joanna Fowler, PhD,*               **Summer 2001**
*and Nora Volkow, MD*
  - Targeted brain microdialysis in awake behaving rats to assess the effects of drugs of abuse on neural transmitter signaling

**American Cyanamid**                                                                               **Princeton, NJ**
*Research Assistant under Dr. Jonathan Hyashi*                                                       **Summer 1998**
  - Implemented patch- and current-clamp recordings of *manduca sexta* neuronal cultures to study the mechanism of action of insecticides

## FUNDING AND AWARDS

- Duke Blue Ribbon Teamwork Award for "Autism and Beyond" app to screen for childhood autism (2015)
- Competitive $100,000 advance from Coursera to co-teach Big Data specialization (2015). Projections expect the courses to be taken by hundreds of thousands of students across the world (http://memp.pratt.duke.edu/news/721)
- "Measuring Implicit Moral Attitudes", Co-Investigator, Duke Institute for Brain Sciences Incubator Award, Duke University (2014-2016). $60,000, renewed for $75,000
- Duke Emerging Leaders Institute Fellow (2015). Selected through a competitive process to undergo six weeks of leadership and management training
- Co-wrote "Bayesian Learning of the Neural Circuit of Empathy." Big Data Information Initiative at Duke (iiD) Research Incubator Award, Duke University (2013-2015). Competitive award for $50,000/yr awarded to support postdoctoral salaries
- Fundamental and Translational Postdoc Fellowship Award, Duke Institute for Brain Sciences and the National Institutes of Health (2013-2014). Competitive award for 1 year of postdoctoral financial support
- Fellow, Stanford Interdisciplinary Group on Neuroscience and Society (2010-2012)
- Bio-X Travel Subsidy Award (2011). Competitive $1,000 award to support invitation to speak at an international conference
- $70,000 dissertation research award from Imitatio /The Thiel Foundation (2010)
- National Science Foundation Graduate Fellowship (2006)

## PUBLICATIONS

**Note: publications written in blue are included in the dossier. The number in parentheses indicates their order of importance.**

**(7) Schaich Borg, J.** "Of Mice and Men: The Influence of Animal Models of Empathy and Social Decision-Making on Human Models of Morality", invited chapter for "Moral Brains: The Neuroscience of Morality", ed. Matthew Liao, expected 2016.

Hashemi, J., Campbell, K., Carpenter, K., Qiu, Q., Tepper, M., Espinosa, S., **Schaich Borg, J.**, Marsan, S., Calderbank, R., Baker, J.P., Egger, H., Dawson, G., Sapiro, G., (2015) A scalable app for measuring autism risk behaviors in young children: A technical validity and feasibility study. In proceedings of *International Conference on Wireless Mobile Communication and Healthcare (Mobihealth)*.

**(8) Schaich Borg, J.**\*, Vu, M.\*, Badea, C., Badea, A., Johnson, G. A., Dzirasa, K. (2015) "Localization of metal electrodes in the intact rat brain using registration of 3-D micro-computed tomography images to a magnetic resonance histology atlas." *eNeuro* 2(4), ENEURO-0017.

Strohminger, N., Caldwell, B., Cameron, D., **Schaich Borg, J.**, & Sinnott-Armstrong, W. (2014). Implicit Morality: A Methodological Survey. In *Experimental Ethics: Toward an Empirical Moral Philosophy*, 133.

**(9)** Carlson, D. E., **Schaich Borg, J.**, Dzirasa, K., & Carin, L. (2014). On the relations of LFPs & Neural Spike Trains. In *Advances in Neural Information Processing Systems*: 2060-2068.

**(10)** Ulrich, K. R., Carlson, D. E., Lian, W., **Schaich Borg, J.**, Dzirasa, K., & Carin, L. (2014). Analysis of Brain States from Multi-Region LFP Time-Series. In *Advances in Neural Information Processing Systems*: 2483-2491.

**(6) Schaich Borg, J.**, Kahn, R., Sinnott-Armstrong, W., Kurzban, R., Robinson, P.H., Kiehl, K. A. (2013) "Subcomponents of Psychopathy have Opposing Contributions to Punishment Judgments." *Journal of Personality and Social Psychology* 105(4):667-687.

**(5) Schaich Borg, J.** and Sinnott-Armstrong, W. (2013) "Moral Reasoning in Psychopaths." In *Handbook of Psychopathy and Law* (Oxford University Press).

**(2) Schaich Borg, J.**, Sinnott-Armstrong, W., Calhoun, V. D., Kiehl, K. A. (2011). Neural basis of moral verdict and moral deliberation. *Social Neuroscience* 6(4): 398-413.

Rolls, A., **Schaich Borg, J.**, de Lecea, L. (2010). Sleep and metabolism: Role of hypothalamic neuronal circuitry. *Clinical Endocrinology and Metabolism* 24(5):817-828.

Cope, L.M., **Schaich Borg, J.**, Harenski C.L., Sinnott-Armstrong, W., Lieberman, D., Nyalakanti, P., Calhoun, V., & Kiehl, K.A. (2010). Hemispheric asymmetries during processing of immoral stimuli. *Frontiers in Evolutionary Neuroscience* 2:1-14.

Carter, M. E., **Schaich Borg, J.**, de Lecea, L. (2009). The brain hypocretins and their receptors: mediators of allostatic arousal. *Current Opinion in Pharmacology* 9(1):39-45.

**Schaich Borg, J.** (2008). "Impaired moral reasoning in psychopaths? Response to Kent Kiehl*".* In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development* (MIT Press).

**(3) Schaich Borg, J.**, Lieberman, D., and Kiehl, K. A. (2008). Infection, Incest, and Iniquity: Investigating the Neural Correlates of Disgust and Morality. *Journal of Cognitive Neuroscience* 20(9):1529-1546.

**(4) Schaich Borg, J.**, Hynes, C., Sinnott-Armstrong, W., Van Horn, J. D.; Grafton, S. T. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience* 18(5):803-17.

Handy, T. C., Tipper, C. M., **Schaich Borg. J.**, Grafton, S. T., Gazzaniga, M. S. (2006). Motor experience with graspable objects reduces their implicit analysis in visual- and motor-related cortex. *Brain Research* 1097(1):156-66.

Handy, T. C., **Schaich Borg, J.**, Turk, D. J., Tipper, C., Grafton, S. T.; Gazzaniga, M. S. (2005). Placing a tool in the spotlight: Spatial attention modulates visuomotor responses in cortex. *NeuroImage* 26(1):266-76.

Marsteller, D. A., Gerasimov, M. R., Schiffer, W. K., Geiger, J. M., Barnett, C. R., **Schaich Borg, J.**, Scott, S., Ceccarelli, J., Volkow, N. D., Molina, P. E., Alexoff, D. L., Dewey, S. L. (2002). Acute Handling Stress Modulates Methylphenidate-induced Catecholamine Overflow in the Medial Prefrontal Cortex. *Neuropsychopharmacology* 27(2):163-70.


**Publications in preparation or in submission*:***

**(1) Schaich Borg, J.**, Lin, L., Srivastava, S., Dunson, D., Dzirasa, K., de Lecea, L. "Distinct frequencies of neural synchrony encode rat empathic avoidance of other rats' pain." (submitted)

Campbell, K., Carpenter, K., Hashemi, J., Espinosa, S., Harris, A., **Schaich Borg, J.**, Chang, Z., Qiu, Q., Tepper, M., Calderbank, R., Baker, J.P., Sapiro, G., Egger, H.L., Dawson, G. "Computer Vision Detects Delayed Social Orienting in Toddlers with Autism." (under review)

Fede, S. J., Harenski, C. L., **Schaich Borg, J.**, Sinnott-Armstrong, W., Rao, V., Caldwell, B., Nyalakanti, P., Koenigs, M., Decety, J., Kosson, D., Calhoun, V. C., Kiehl, Kent A. "Abnormal fronto-limbic engagement in incarcerated stimulant users during moral processing." (under review)

Kumar, S., Hughes, D., Katz, B. M., Johnson, K. E., **Schaich Borg, J.**, Lin, L., Deisseroth, K., Jaaro-Peled, H., Wetsel, W. C., Dunson, D., Sawa, A., Dzirasa, K. "Cortico-striatal gating dysfunction in mice with impairment of the DISC1 gene locus cortico-striatal circuit function modulates PPI." (submitted)

Fede, S. J., **Schaich Borg, J.**, Nyalakanti, P. K., Cope, L. M., Harenski, C. L., Sinnott-Armstrong, W., Koenigs, M., Calhoun, V. C., Kiehl, K. A. "Distinct patterns of positive and negative moral processing in psychopathy." (under review)

**Schaich Borg, J.***, Rangarajan, V.*, Parvizi, J. "Differential electrophysiological response during moral deliberation and moral judgment in the human posterior cingulate, parietal cortex, and medial prefrontal cortex."

**Schaich Borg, J.**, Hynes, C., Sinnott-Armstrong, W., Van Horn, J.D.; Grafton, S.T. "Moral decision-making…the second time around.*"*

---

# ACADEMIC PRESENTATIONS

**Schaich Borg, J.** "The Amygdala as a Network 'Coordinator' in Rat Empathic Avoidance." Association for Psychological Science Annual Convention, New York City, NY. May 24, 2015.

**Schaich Borg, J.** "Moral Judgment in Psychopaths." Invited speaker at the Moral Psychology Research Group (an invitation-only interdisciplinary group of world scholars in moral judgment), New Orleans, LA. November 8, 2014.

**Schaich Borg, J.** and Dzirasa, K. "Innovation through Integration: Bridging the Translational Divide." Duke Psychiatry and Behavioral Sciences Grand Rounds, Duke University School of Medicine, Durham, NC. August 21, 2014.

**Schaich Borg, J.** "The Brain's Disgust with Morality." Invited speaker at the Moral Psychology Research Group (an invitation-only interdisciplinary group of world scholars in moral judgment), Durham, NC. April 26, 2014.

**Schaich Borg, J.** "The Classification and Anatomy of Moral Judgments." Invited speaker at the "Morality and the Cognitive Sciences" conference, Riga, Latvia. May 7, 2011.

**Schaich Borg, J.** "Recent Developments in the Neuroscience of Moral Decision-making: Implications for Ethics and Law?" Invited speaker at the Fourth International Legal Ethics Conference, Stanford, CA. July 17, 2010.

**Schaich Borg, J.** "A Primer on the Neural Basis of Moral Decision Making." Invited speaker for "Neuroimaging and the Law" workshop, Halifax, Canada. May 20, 2010.

**Schaich Borg, J.** and Sinnott-Armstrong, W. "Moral Judgments by Psychopaths?" Invited speaker at the MacArthur Foundation Psychopathy and the Law Symposium (a pre-conference to the Society for the Scientific Study of Psychopathy annual meeting), New Orleans, LA. April 16, 2009.

**Schaich Borg, J.** Invited Panelist for "Neuroscience and the Courts: The Implications of Advances in Neurotechnology", held by the Stanford Technology Law Review, Stanford Law School, Stanford, CA. February 27, 2009.

**Schaich Borg, J.** "Questioning the role of the Ventromedial Prefrontal Cortex in the Diminished Moral Capacity of Psychopaths." Invited speaker at the MacArthur Foundation Law and Neuroscience Project meeting, Stanford, CA. January 24, 2008.

**Schaich Borg, J.,** Lieberman, D., and Kiehl, K. "The Common and Distinct Neural Correlates of Pathogen and Socio-moral Disgust." Speaker at Neural Systems of Social Behavior Conference, Austin, TX. May 12, 2007.

**Schaich Borg, J.** "Morality and Ethics in the Brain: Applications for Clinical Cognitive Neuroscience." Psychopharmacology/Neuroscience Grand Rounds at Hartford Hospital, CT. February 23, 2006

**Schaich Borg, J.** and Sinnott-Armstrong, W. "Consequences, Action, and Intention as Factors in Moral Judgments." Invited speaker for the Program of Neuroscience Seminar series, Princeton University. September 25, 2005

**Schaich Borg, J.** and Sinnott-Armstrong, W. "Morality in the Brain." Invited speaker at the Center for Bioethics at Columbia University. February 24, 2005

_____
# TEACHING EXPERIENCE

**"Excel to MySQL: Analytic Techniques for Business"**                    **July 2015 to present**
*A "Specialization" of Five related Massive Open Online Courses (MOOCs) hosted by Coursera*
  - Funded by a competitive $100,000 advance from Coursera
  - Projections expect the courses to be taken by hundreds of thousands of students across the world (current enrollment in January, 2015 is over 100,000 students)
  - *https://www.coursera.org/specializations/excel-mysql*
  - Details of individual courses provided below

**Business Metrics for Data-Driven Companies (secondary instructor)**          **September, 2015 - September, 2018**
*An introduction to Business Intelligence and financial analytics metrics*
  - Co-developed learning objectives, course syllabus, and assessment tools

**Mastering Data Analysis in Excel (secondary instructor)**          **October, 2015 - October, 2018**
*An introduction to predictive models in Excel*
  - Co-developed learning objectives and course syllabus

**Data Visualization and Communication with Tableau (primary instructor)**          **November, 2015 - November, 2018**
*An introduction to analysis strategies and data communication techniques using Tableau visualization software*
  - Negotiated a partnership with Tableau (www.tableau.com) to provide free Tableau Professional accounts to all students for one year (normal cost is $1999 per person)
  - Arranged a collaboration with local start-up Dognition (www.dognition.com) to provide students with access to a real-world data set of over 1 million rows in size
  - Created over 40 separate videos of course content describing how to analyze large data sets, how to use the visualization software Tableau, and how to communicate analysis findings to non-technical audiences, with featured interviews from senior analysts employed by Gartner, Airbnb, and Square
  - Authored online assessments based on instructed analysis of Dognition data set
  - Designed a final project that required students to present a 5-minute pitch to Dognition executives about a business recommendation derived from their data analysis

**Managing Big Data with MySQL (primary instructor)**          **January, 2016 - January, 2019**
*An introduction to writing SQL queries with business "Big Data" databases*
  - Negotiated a partnership with Teradata and Walton School of Business, University of Arkansas to provide students with free access to a Dillard's department store database containing over 100 million rows of data, and free access to Teradata Viewpoint's SQL scratchpad interface
  - Led team at Duke's Office of Information Technology to create a web-based, instructional interface (www.jupyter.org) that would allow tens of thousands of students to simultaneously access read-only copies of a MySQL database housing donated data from local start-up Dognition
  - Recorded over 20 videos introducing relational database concepts, and featuring interviews from senior software engineers and analysts employed by Box, Airbnb, and Dognition
  - Authored over 100 pages of online interactive exercises that walk students through how to implement SQL queries
  - Created online assessments based on SQL queries of the Dillard's and Dognition databases

**Excel to MySQL: Analytic Techniques for Business Capstone Project**          **February, 2016 - February, 2019**
**(co-primary instructor)**
*Final project of the specialization where students have to pitch a business recommendation to the executive suite of a fictional real estate company based on their analysis of a real estate database. Top projects will receive feedback from Data Science team at Airbnb.*
  - Co-designed course vision and learning objectives
  - Co-created mock real-estate data set for project
  - Co-authored extensive rubric to evaluate the quality of the skills, analysis, and presentation exhibited in final project

# Mariano Tepper `<mariano.tepper@duke.edu>`

Work address:
Duke University, Department of Electrical and Computer Engineering
Campus box 90291
Durham, North Carolina 27708

Home address:
3611 University Dr Apt 17K
Durham, North Carolina 27707

Tel: +1 919 641 4558 (Mobile)
http://www.marianotepper.com.ar

## EDUCATION

**PhD in Computer Science**                                                                 2011
Department of Computer Science, Universidad de Buenos Aires, Argentina
- Title: Detecting clusters and boundaries: a twofold study on shape representation.
  Advisors: A. Almansa, and M. Mejail.
- With honors.
- CONICET (National Scientific and Technical Research Council, Argentina) scholarship

**Master in Mathematics, Vision and Learning**                                              2007
École Normale Supérieure de Cachan, France
- "Bien" mention.
- ALFA Program "Computer Vision Foundations and Applications" scholarship.

**Computer Science Licenciate degree**                                                      2006
Universidad de Buenos Aires, Argentina
- With honors.

## WORK EXPERIENCE

**Duke University**                                                                  2014 - present
Project Manager
- "Lives of Things" Bass Connections Project "The Museum Inside Out: Interactive 3D Exhibition Design and Programming"

**Information Initiative at Duke & Department of Electrical and Computer Engineering,**  2012 - present
**Duke University**
Postdoctoral Associate

**Department of Electrical and Computer Engineering, University of Minnesota**        2011 - 2012
Postdoctoral Research Associate

## PUBLICATIONS

**Journal papers:**

M. Tepper, G. Sapiro, "Compressed nonnegative matrix factorization is fast and accurate." IEEE Transactions on Signal Processing, 2016 (accepted)

M. Tepper, G. Sapiro, "A bi-clustering framework for consensus problems." SIAM Journal on Imaging Sciences, 2014

J. Hashemi, M. Tepper, T. Vallin Spina, A. Esler, V. Morellas, N. Papanikolopoulos, H. Egger, G. Dawson, G. Sapiro, "Computer vision tools for low-cost and non-invasive measurement of autism-related behaviors in infants." Autism Research and Treatment, 2014

A. Morelli Andres, S. Padovani, M. Tepper, J. Jacobo, "Face recognition on partially occluded images using compressed sensing." Pattern Recognition Letters, 2014

M. Tepper, P. Musé, A. Almansa, M. Mejail, "Finding contrasted and regular edges by a contrario detection of periodic subsequences." Pattern Recognition, 2013

M. Tepper, P. Musé, A. Almansa, "On the role of contrast and regularity in perceptual boundary saliency." Journal of Mathematical Imaging and Vision, 2013

M. Tepper, P. Musé, A. Almansa, M. Mejail, "Automatically finding clusters in normalized cuts." Pattern Recognition, 2011

## Technical reports:

M. Tepper, G. Sapiro, "A short-graph Fourier transform via personalized PageRank vectors." arXiv:1510.02700, 2015

M. Tepper, G. Sapiro, "Compressed nonnegative matrix factorization is fast and accurate." arXiv:1505.04650, 2015

T. Vallin Spina, M. Tepper, A. Esler, V. Morellas, N. Papanikolopoulos, A. Falcão, G. Sapiro, "Video human segmentation using fuzzy object models and its application to body pose estimation of toddlers for behavior studies." arXiv:1305.6918, 2013

J. Hashemi, T. Vallin Spina, M. Tepper, A. Esler, V. Morellas, N. Papanikolopoulos, G. Sapiro, "Computer vision tools for the non-invasive assessment of autism-related behavioral markers." arXiv:1210.7014, 2012

M. Tepper, G. Sapiro, "L1 splines for robust, simple, and fast smoothing of grid data." arXiv:1208.2292, 2012

Z. Tang, A. Castrodad, M. Tepper, G. Sapiro, "Are you imitating me? Unsupervised sparse modeling for group activity analysis from a single video." arXiv:1208.5451, 2012

M. Tepper, M. Mejail, P. Musé, A. Almansa, "Boruvka meets nearest neighbors." hal-00583120, 2011

M. Tepper, P. Musé, A. Almansa, "Meaningful clustered forest: an automatic and robust clustering algorithm." arXiv:1104.0651, 2011

## Peer-Reviewed Conferences:

M. Tepper, G. Sapiro, "A short-graph Fourier transform via personalized PageRank vectors." ICASSP, 2016 (accepted)

M. Tepper, G. Sapiro, "From local to global communities in large networks through consensus." CIARP, 2015

J. Hashemi, K. Campbell, K. Carpenter, A. Harris, Q. Qiu, M. Tepper, S. Espinosa, J. Schaich Borg, S. Marsan, R. Calderbank, J. Baker, H. Egger, G. Dawson, G. Sapiro, "A scalable app for measuring autism risk behaviors in young children: A technical validity and feasibility study." MOBIHEALTH, 2015

A. Newson, M. Tepper, G. Sapiro, "Low-rank spatio-temporal video segmentation." BMVC, 2015

M. Tepper, A. Newson, P. Sprechmann, G. Sapiro, "Multi-temporal foreground detection in videos." ICIP, 2015

M. Tepper, G. Sapiro, "Intersecting 2D lines: A simple method for detecting vanishing points." ICIP, 2014

M. Tepper, G. Sapiro, "All for one, one for all: Consensus community detection in networks." ICASSP, 2014

M. Tepper, G. Sapiro, "Ants crawling to discover the community structure in networks." CIARP, 2013

M. Tepper, P. Musé, A. Almansa, M. Mejail, "Boruvka meets nearest neighbors." CIARP, 2013

I. Ramírez, M. Tepper, "Bi-clustering via MDL-based matrix factorization." CIARP, 2013

Z. Tang, M. Tepper, G. Sapiro, "Reflective symmetry detection by rectifying randomized correspondences." BMVC, 2013

M. Tepper, G. Sapiro, "Fast L1 smoothing splines with an application to Kinect depth data." ICIP, 2013

J. Hashemi, T.V. Spina, M. Tepper, A. Esler, V. Morellas, N. Papanikolopoulos, G. Sapiro, "A computer vision approach for the assessment of autism-related behavioral markers." ICDL-EpiRob, 2012

A. Asiaee Taheri, M. Tepper, A. Banerjee, G. Sapiro, "If you are happy and you know it... tweet." CIKM, 2012

M. Tepper, G. Sapiro, "Decoupled coarse-to-fine matching and nonlinear regularization for efficient motion estimation." ICIP, 2012

A. Morelli Andres, S. Padovani, M. Tepper, M. Mejail, J. Jacobo, "Randomized face recognition on partially occluded images." CIARP, 2012

M. Tepper, P. Musé, A. Almansa, M. Mejail, "Finding edges by a contrario detection of periodic subsequences." CIARP, 2012

P. Negri, M. Tepper, D. Acevedo, J. Jacobo, M. Mejail, "Multiple clues for license plate detection and recognition." CIARP, 2010

M. Mottalli, M. Tepper, M. Mejail, "A contrario detection of false matches in iris recognition." CIARP, 2010          .

E. Cura, M. Tepper, M. Mejail, "Content-based emblem retrieval using zernike moments." CIARP, 2010

M. Tepper, D. Acevedo, N. Goussies, J. Jacobo, M. Mejail, "A decision step for shape context matching." ICIP, 2009

M. Tepper, F. Gómez, P. Musé, A. Almansa, M. Mejail, "Morphological shape context: semi-locality and robust matching in shape recognition." CIARP, 2009


**Conference Presentations:**

A. Dodson, M. Tepper, J. Hashemi, "Projecting polychromy: The art and science of displaying medieval sculpture." Southeastern College Art Conference, 2015

W. F. Broderick, R. M. Carter, M. Tepper, J. F. Gariepy, M. L. Platt, G. Sapiro, and S. A. Huettel, "A multi-variate pattern analysis investigation of strategic thinking and deception in a dynamic, competitive game." Neuroscience, 2015

M. Tepper, G. Sapiro, "Compressed nonnegative matrix factorization is fast and accurate." Duke Workshop on Sensing and Analysis of High-Dimensional Data, USA, 2015

M. Tepper, G. Sapiro, "A bi-clustering formulation of multiple model estimation." Duke Workshop on Sensing and Analysis of High-Dimensional Data, USA, 2013

J. Hashemi, T.V. Spina, M. Tepper, A. Esler, V. Morellas, N. Papanikolopoulos, G. Sapiro, "A computer vision approach for aiding in the autism observation scale for infants test: Automatic monitoring of visual attention." American Academy of Child and Adolescent Psychiatry Annual Meeting, USA, 2012

T.V. Spina, J. Hashemi, M. Tepper, A. Esler, V. Morellas, N. Papanikolopoulos, G. Sapiro, "Automatically aiding the assessment of atypical motor patterns for ASD diagnosis." American Academy of Child and Adolescent Psychiatry Annual Meeting, USA, 2012

M. Tepper, G. Sapiro, "Ants performing patchwork for detecting community structure in networks." SIAM Annual Meeting, USA, 2012

F. Gómez, M. Tepper, "Shape context and semi-locality." Escuela de Ciencias de las Imágenes (ECImag), Argentina, 2009

E. Cura, M. Tepper, "Trademark retrieval using Zernike moments." Escuela de Ciencias de las Imágenes (ECImag), Argentina, 2009

M. Tepper, "Visual quantification of the Copan hieroglyphic stairway deterioration." ALFA workshop in Computer Vision Foundations and Applications, France, 2008

## INVITED TALKS

"All for one and one for all: A bi-clustering model for reaching consensus in grouping problems" at Universidad de Buenos Aires, Argentina, 2015

"Matrix factorization for big data: From video analysis to consensus problems" at Comcast Labs, USA, 2015

"A bi-clustering framework for consensus problems" at North Carolina State University, USA, 2015

"A bi-clustering framework for consensus problems" at Telecom ParisTech, France, 2014

"A bi-clustering framework for consensus problems" at Université Paris Descartes, France, 2014

"A bi-clustering framework for consensus problems" at École Normale Supérieure de Cachan, France, 2014

"The proximity gestalt: A computational quest" at Universidad de la Rep'ública, Uruguay, 2011

## TEACHING EXPERIENCE

**Department of Electrical and Computer Engineering, Duke University, USA**     2013 - 2016
Lead Teaching and Preproduction Assistant
- MOOC "Image and video processing: From Mars to Hollywood with a stop at the hospital." `https://www.coursera.org/course/images` (91.000 students in total)

**Department of Electrical and Computer Engineering, Duke University, USA**     2014
Instructor
- "3d Design & Programming for Arts & Medicine"

**Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina**     2009
Teaching Assistant
- "Algorithms and Data Structure I"
- "Robot Vision."

**Imaging Sciences School (ECImag), Facultad de Ciencias Exactas. Universidad Nacional del Centro, Argentina**     2009
Teaching Assistant
- "Geometrical Structure Detection based on Computational Gestalt. Applications to shape recognition."

**Imaging Sciences School (ECImag) 2008. Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina**     2008
Teaching Assistant
- "Geometrical Structure Detection based on Computational Gestalt. Applications to shape recognition"

**Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina**     2005
Teaching Assistant
- "Introduction to Digital Image Processing"

## STUDENT SUPERVISION

Graduate students
- W. Broderick, Research Assistant in Computational Neuroscience, Duke University, 2014-2015.
- J. Hashemi, Research Assistant in Electrical and Computer Engineering, University of Minnesota, 2011-2012.

Undergraduate students
- A. Rajkovic. Summer intern in Electrical and Computer Engineering, Duke University, 2015.

Licenciate thesis advisor
- "Face Recognition with automatic occlusion handling". A. Andrés y S. Padovani. Co-advisor M. Mejail. Universidad de Buenos Aires, Argentina, 2011.
- "Maximally Stable Extremal Regions in the Tree of Shapes". F. Gómez. Co-advisor: M. Mejail. Universidad de Buenos Aires, Argentina, 2010.
- "Analysis of Zernike moments as shape descriptors". E. Cura. Co-advisor: M. Mejail. Universidad de Buenos Aires, Argentina, 2010.

## SERVICE TO PROFESSION

Developer of the linear algebra module in Dask (`http://dask.readthedocs.org`), which enables parallel computing through task scheduling and blocked algorithms. This allows developers to write complex parallel algorithms and execute them in parallel either on a modern multi-core machine or on a distributed cluster.

Reviewer for:
- IEEE Transactions on Image Processing
- SIAM Journal on Imaging Sciences
- Pattern Recognition
- Neurocomputing
- International Journal of Pattern Recognition and Artificial
- Image Processing On Line
- Latin American Conference of Informatics (CLEI)
- Ibero-American Conference on Pattern Recognition (CIARP)
- National Agency of Research and Innovation (ANII, Uruguay)
- Sectorial Committee of Scientific Research (CSIC, Uruguay)
- Army Research Office (ARO)

## FUNDING

Contributed in writing the proposal "Network Inference and Analysis of Big Dynamic Data," National Geospatial-Intelligence Agency (NGA) Academic Research Program, PI Guillermo Sapiro, 2016. Funded.

Contributed parts of the proposal "Advanced Computational Video for Leave Behind Detection," U.S. Department of Homeland Security and the Johns Hopkins University Applied Physics Laboratory, PIs Guillermo Sapiro and Lawrence Carin, 2012-2014. Funded.

Contributed in writing the proposal "Mathematical Models for Visual Perception and Subpixel Computer Vision," STIC-AmSud scientific-technological cooperation programme between ENS Cachan (France), Universidad de Buenos Aires (Argentina), Télécom ParisTech (France), Universidad de la República (Uruguay), 2011-2013. for Argentina. Funded.

## LANGUAGES

- Spanish, native speaker.
- French, bilingual (Baccalauréat général, série Scientifique, option physique-chimie).
- English, fluent.

## SKILLS

- Programming: comfortable with Python, Matlab, C, C++, Java, HTML. In the past, I have programmed in Smalltalk, Prolog, Haskell, Assembler, SQL, Javascript, and other languages.
- Software architecture. Database design. Design patterns.
- Operating systems: Linux, Windows, Mac OS X.

# Thomas Peter Witelski

Department of Mathematics  
Duke University  
Box 90320  
Durham, North Carolina 27708-0320

(919) 660-2841 (office)  
(919) 660-2821 (fax)  
witelski@math.duke.edu  
http://www.math.duke.edu/faculty/witelski

## EDUCATION

**Cooper Union School of Engineering** 1987 – 1991  
*B.S. in Engineering*, May 1991, Summa Cum Laude  
Concentration: Interdisciplinary Engineering and Mathematical Methods

**California Institute of Technology** 1991 – 1995  
*Ph.D. in Applied Mathematics*, June 1995  
Thesis area: Applied mathematics for nonlinear diffusion models  
Thesis advisor: Donald S. Cohen

**Massachusetts Institute of Technology, Mathematics Dept.** 1995 – 1998  
Postdoctoral Advisor: D. J. Benney

## ACADEMIC EXPERIENCE

**Duke University** 1998 – Present  
*Department of Mathematics, Full Professor (2011–Present)*  
*Department of Mechanical Engineering and Materials Science, secondary appointment (2012–Present)*

**Massachusetts Institute of Technology, Mathematics Dept.** 1995 – 1998  
*NSF Postdoctoral Fellow/Applied Mathematics Instructor*, Advisor: D. J. Benney

**Caltech, Applied Mathematics Dept.** 1991 – 1995  
*NSF Graduate Fellow/Graduate Research Assistant*, Advisor: D. S. Cohen

**Los Alamos National Laboratory** Summer 1992  
*Graduate Research Assistant, Computer division*, Advisor: P. S. Hagan

**Woods Hole Oceanographic Institution** Summer 1991  
*Geophysical Fluid Dynamics Summer Fellow*, Advisor: E. A. Spiegel

**Argonne National Laboratory** Summer 1990  
*Undergraduate Research Assistant, Engineering Physics*, Advisor: C. B. Reed

## VISITING POSITIONS

**University of Oxford, Mathematical Institute** Spring 2013  
*OCCAM Visiting Fellow*, Oxford Centre for Collaborative Applied Mathematics

**University of Oxford, Mathematical Institute** 2007–2008  
*Reader in Applied Mathematics*, Oxford Centre for Industrial and Applied Mathematics

## PROFESSIONAL ORGANIZATIONS
- SIAM: Society of Industrial and Applied Mathematics
- AMS: American Mathematical Society

## AWARDS
- 2003–2008 NSF DMS CAREER: Math. Modeling of Microfluidic Dynamics and Transport
- 2000–2003 Alfred P. Sloan Foundation Fellowship
- 1995–1998 NSF Postdoctoral Fellowship (MIT, Math)
- 1992–1995 NSF Graduate Fellowship (Caltech, Applied Math)

## Synergistic activities

- Associate Director, Statistical and Applied Mathematical Sciences Institute (SAMSI), 2014–present
- **Journal of Engineering Mathematics**, *co-Editor-in-Chief*, 2011–2015
- **Journal of Mathematical Analysis and Applications**, *Division Editor*, 2009–2014
- Editorial boards: **European Journal of Applied Mathematics**, **Discrete and Continuous Dynamical Systems Series B**
- Conferences organized:
  - Mathematical Problems in Industry (MPI) 2016 workshop, Duke University, June 13-17, 2016
  - Dynamics Days 2016, organizing committee, Jan 7–10, 2016, Durham, North Carolina
  - Flows and Patterns: The Physics of Fluids, Granular Materials, and Soft Matter, organizing committee, Oct 12-13, 2013, Duke University
  - BIRS workshop 12w5035: Thin liquid films and fluid interfaces: models, experiments and applications, organizing committee, Dec 9–14, 2012, Banff, Canada
  - Dynamics Days 2011, organizing committee, Jan 5–8, 2011, Chapel Hill, North Carolina
- Director of Graduate Studies, Dept of Mathematics, Duke University, 2009–2011
- Executive Committee of the Graduate Faculty, Duke University, 2010–2012

## Selected publications

- Hangjie Ji and T.P. Witelski, Finite-time Thin Film Rupture Driven by Generalized Evaporative Loss, submitted to Physica D (December 2015, in review) [arXiv:1601.03625]
- P. Sanaei, G. W. Richardson, T. Witelski, L.J. Cummings, Flow and fouling in a pleated membrane filter, **Journal of Fluid Mechanics**, vol. 795 (2016), pp. 36-59
- J. A. Dijksman, S. Mukhopadhyay, C. Gaebler, T. P. Witelski and R. P. Behringer, Obtaining self-similar scalings in focusing flows, **Physical review E**, vol. 92 no. 4 (2015), pp. 043016
- T.P. Witelski and M. Bowen, Methods of Mathematical Modelling, Springer Undergraduate Mathematics Series (SUMS) (2015), 305 pp
- N.S. Hall Taylor, I.J. Hewitt, J.R. Ockendon, T.P. Witelski, A new model for disturbance waves, **International Journal of Multiphase Flow**, vol. 66 (2014), pp. 38-45
- S.J.Chapman, P.H.Trinh and T.P. Witelski, Exponential asymptotics for thin film rupture, **SIAM Journal on Applied Mathematics**, vol. 73 no. 1 (2013), pp. 232-253.
- C. Breward, E. Aydemir, T.P. Witelski, The effect of polar lipids on tear film dynamics, **Bulletin of Mathematical Biology**, vol. 73 (2011), pp. 1171-1201
- M. Aguareles, S. J. Chapman, T. Witelski, Motion of spiral waves in the complex Ginzburg-Landau equation, **Physica D**, vol. 239 no. 7 (2010), pp. 348-365

## Graduate students

(Total 7) Andrew Feist (2003–2005), Ryan Haskett (2002–2006), Michael Gratton, Northwestern U. (2004–2008), Mihaela Froehlich (2005–2009), Harrison Potter (2010–2015), Hangjie Ji (2012–Present), Weifan Liu (2015–Present)

## Postdocs

(Total 7) Mark Bowen, Waseda (2000–2001), Karl Glasner, U-Arizona (2001–2002), Linda Smolka, Bucknell University (2002–2004), Sandra Wieland (2004), Anand Jayaraman (2004–2005), Rachel Levy, Harvey Mudd College (2005–2007), Anne Catlla, Wofford College (2005–2008)

## Recent collaborators

| | | |
|---|---|---|
| R. P. Behringer, Duke | R. Levy, Harvey Mudd | L. B. Smolka, Bucknell |
| M. Bowen, Waseda (Japan) | A. Münch, Humboldt U. (Berlin) | L. N. Virgin, Duke |
| C. J. Breward, Oxford | D. G. Schaeffer, Duke | B. Wagner, WIAS (Berlin) |
| S. J. Chapman, Oxford | D. S. Schwendeman, RPI | |

## Appendix C: Statements of Support

Following are letters of support from various units around the university. Since near final drafts of the proposal have circulated for the past 9 months for feedback and comment, some of the attached letters were written in Spring of 2015, and some of the letter-writers no longer hold the administrative positions they held then. However, in each such case (indicated by *), the successor similarly supports the proposal.

Jonathan Mattingly (on behalf of the Mathematics Department)

Ron Parr (on behalf of the Department of Computer Science)

Michael Platt (on behalf of the Duke Institute for Brain Sciences)*

Sunshine Hillygus (on behalf of the Duke Initiative on Survey Methods)

Richard Newell (on behalf of the Energy Initiative)

Pat Bayer (on behalf of the Department of Economics)*

William Boulding (on behalf of the Fuqua School of Business)

Robert Kramer (on behalf of the Duke Global Health Institute)

Victoria Szabo (on behalf of Information Science + Studies Program)

Eric Toone (on behalf of the Initiative in Innovation and Entrepreneurship)

Scott Huettel (on behalf of the Department of Psychology and Neuroscience)

Elizabeth Merwin  
Janice Humphreys  
Rachel Richesson      (on behalf of the School of Nursing)  
Constance Johnson  
Ryan Shaw

Eduardo Bonilla-Silva (on behalf of the Department of Sociology)

Merlise Clyde (on behalf of the Department of Statistics)

Susan Roth (on behalf of the Office of the Vice-Provost of Interdisciplinary Studies)*

Larry Carin (on behalf of the Office of the Vice-Provost for Research)

Ed Balleisen (on behalf of the Office of the Vice-Provost for Interdisciplinary Studies)

Sandy Connolly, Vice Dean for Finance and Administration, A&S

Linda Burton and Dan Kiehart, Deans for Social Science and Natural Science in A&S

August 21, 2016

From:  Jonathan Mattingly
       Chair of the Mathematics Department
       Duke University

Re: Proposed Masters Program in Data Science

As the Chair of the Mathematics Department, I write in enthusiastic support of this proposal for the creation of a Masters Program in Data Science. It will provide an important opportunity for the MS students, for the mathematics department, for its undergraduate and PhD students, and for Duke.

The Duke Mathematics department is unusual among mathematics department. It represents a seamless blending of core and applied mathematics, with world class experts in probability, image/signal processing. multi-scale data representations/analysis, and mathematical modeling in domains ranging from biomedical to cellular biology to molecular dynamics to fluid mechanics.

Additionally, the Mathematics Department has a long proven track record of vertical integration in both research and education. The department is nationally recognized for its training of graduate students in instruction and communication.

These unique strengths of the Mathematics Department have yet to be leveraged in any of the Masters Programs at Duke. Furthermore the addition of this MS Program will improve our other mathematics programs and Duke in general.

The following summarize the synergies between the proposed MS Program and the interests of the Mathematics Department:

- developing a program which can deliver a modern degree in the mathematical sciences which connects disciplinary knowledge to modern mathematical modeling through the integrated use and understanding of data

- improving the mathematical literacy and sophistication of established and budding data scientists and of society in general.

- providing a track for an MS degree in the mathematical sciences that mixes a substantive modeling component with a deep understanding of the analysis and use of data.

- creating a broader and larger community which will improve both our PhD and undergraduate curriculum

- creating a vibrant interdisciplinary graduate student research community which mixes both professional and more traditional PhD students

- increasing the attractiveness of the mathematics doctoral program to potential students, especially those from non-traditional backgrounds extending our reputation for excellence in teacher-training to other areas of professional development.

I will now explain the connection between the MDS proposal and each of these ideas in turn.

The MDS will create an MS program that will fit nicely into the educational and research program of the Mathematics Department. It will allow the exceptionally interdisciplinary Mathematics Department to train a new class of mathematical and data experts who will populate a wide and diverse range of positions in enterprises and research groups.

Involvement of the Mathematics Department will help improve the mathematical and technical literacy of even the most applied and problem or policy oriented MS students. It will also provide a mathematical modeling track, which will make it unique among the MS programs at Duke, and which will make it stand out when compared to many other data science programs nationally. None of the current Duke MS programs provide a track with general training in mathematical modeling like what is envisaged by the Mathematics Department.

Partnership with MDS will lead to new mathematics classes and diversify student participation in existing courses. The Mathematics Department will take responsibility for two of the core courses in the proposed MS program, and this course development is synergistic with the transformation of the Mathematics curriculum in Linear Algebra that we are initiating in Fall 2016. Ezra Miller, Tom Witelski, and Paul Bendich are directing this transformation, and Paul will use the Data Expedition program at iiD to infuse new content into the curriculum. The courses under development will primarily be at the mathematics undergraduate and non-mathematics MS level. This will further enrich the undergraduate curriculum both because of the additional courses possible but also because it will bring a more diverse group of students into the class room. Students with professional experience and orientation will add to the diversity and hence the quality of the classes.

The Mathematics Department has a proven track record of offering Graduate Mini-Courses (Math 790-92) to graduate students in social sciences, biology, and others who need their mathematical background refreshing. Mini-Courses on the books include differential calculus, discrete probability, and integral calculus. Modules to be offered in Spring 2017 include linear algebra, ordinary differential equations and modeling case studies. These resources will be available to MS students in need of a mathematics refresher. The graduate program in Mathematics at Duke already does an impressive job preparing its students for academic research. Interaction with the MDS students will help broaden this preparation. We envision interactions of doctoral students with MDS students, and possibly doctoral students from other quantitative STEM fields, to explore a data-driven research project. This will enrich all involved. It will provide even the most theoretical mathematics PhD student both a vision of a different possible career path as well as exposure to new classes of possible research problems.

The MDS program will open up more connections between Duke and the entities outside the university that use data analytics. We hope to use these connections to open up lines

of dialogue between our doctoral students and local companies, schools, and governmental agencies. This will certainly make our graduate program very attractive to prospective students, especially those that want to see a clear path between their research and benefits to their community.

The Mathematics Department has excellent teacher training program which is promoted as model for other universities. We hope that the MDS program will help our teaching faculty interact with their counterparts in other quantitative STEM fields. In particular, it will help them enrich this training with the many other skills demanded by the jobs that our students may take after graduation. By adding these new industrial and professional training options we plan to build model program in professional training beyond traditional academic option for our PhD student. We expect that this will make the PhD program more attractive to students from non-traditional backgrounds. I would like to close by stating that the Mathematics Department is fully supportive of the financial model outlined in the proposed MS Program. I can also share that I have been working with Sandy Connolly and the Chairs of Computer Science and Statistics to develop a framework for MS funding that would hopefully apply a finicial model similar to that porposed more broadly.

Sincerely,

Jonathan C. Mattingly (`jonm@math.duke.edu`)
Chair of Mathematics Department
Prof. of Mathematics & of Statistical Science

OFFICE OF THE CHAIR
DEPARTMENT OF COMPUTER SCIENCE
D315 LEVINE SCIENCE RESEARCH CENTER
BOX 90129

TELEPHONE (919) 660-6548
FAX (919) 660-6502
EMAIL CHAIR@CS.DUKE.EDU
WWW.CS.DUKE.EDU

November 2, 2015

Professor Robert Calderbank
317 Gross Hall
Duke University
Durham, NC 27708

Dear Robert:

I am pleased to support the proposed MS in Data Science (MSDS) submitted by the Social
Science Research Institute (SSRI) and the Information Initiative at Duke (iiD).

Several of our faculty teach courses that may be of interest to MSDS students and we will
welcome these students, though we will give priority to computer science students if space is
limited.

This academic year we are expecting Cynthia Rudin to join the Computer Science Department
(joint with ECE) as an Associate Professor with tenure. Cynthia works in Machine Learning,
starting from applications, and using them to illuminate the research frontier in her discipline.
One example is her development of algorithms to learn predictive classifiers that are also
interpretable (*Falling Rule Lists)*. We expect her research direction to be of significant interest to
industry, as well as to clinical departments in the School of Medicine. We also expect that her
graduate courses will be very attractive to students in the MSDS program as well as to students
in our MS programs, both in Computer Science (MSC) and Economics and Computation
(MSEC).

A second example is Sudeepa Roy, an expert in databases who joined the Department in
September. She is interested in data mining and other topics related to exploratory data analysis.
The Information Initiative is working with her to develop project proposals for the Data+
program at iiD. In this program, MS students and Duke undergraduates from different disciplines
form teams to address data-driven approaches to interdisciplinary challenges.

The proposed MSDS program aims for interdisciplinary breadth rather than technical depth in
computer science and is complementary to our MSC and MSEC programs, which to offer a more
traditional academic curriculum. The proposed courses on Data Marshaling and Data
Visualization emphasize development of highly practical skillsets that could be useful to
computer science students working on projects. Department policy gives our DGS and DUS a
good deal of flexibility in allowing students count classes taken in other departments towards
their requirements. I cannot usurp their authority in this letter, but I expect that MSDS courses

will be viewed favorably. Of course, we also understand that priority can and should be given to MSDS students if space is limited.

I believe the MSDS program will provide a valuable educational opportunity to students, and I look forward to cooperating with you as the program develops.


Sincerely,

Ronald Parr
Chair and Professor
Department of Computer Science
Duke University

# DUKE
## INSTITUTE *for*
## BRAIN
## SCIENCES

**Michael L. Platt, PhD, Director**
**Professor, Departments of Neurobiology,**
Evolutionary Anthropology,
and Psychology & Neuroscience
Rm. B243F LSRC Building
Duke University
Durham, NC 27710

March 30, 2015

Robert Calderbank
Tom Nechyba
Duke University

Dear Tom and Robert,

I am enthusiastic about supporting your proposal for an MS degree in Data Science, particularly given the way you have set it up to be inclusive of all substantive areas. You correctly, in my view, identify data science as intellectually cross-cutting and the graduate program you have conceived clearly expresses this vision. With SSRI's infrastructure and iiD's depth in big data methods, you have a powerful base upon which to build the collaborations you envision.

The program becomes even more compelling for us in the Duke Institute for Brain Sciences (DIBS) in light of the connections with our faculty you have already nurtured. Helen Egger's exciting collaborations with iiD come to mind, as does SSRI's ongoing support of projects connecting to Geri Dawson's new Center for Autism and Brain Development. Both SSRI and iiD have collaborated with the Bass theme on Brain and Society that we house in DIBS (including my own project teams), another indication of how you are reaching well beyond your traditional domains. This creates the perfect foundation for a collaborative MS degree of the kind you propose. If you succeed, as I am sure you will, this will furthermore bring powerful tools to many disciplinary and interdisciplinary projects, including some critical projects important to the faculty community in DIBS.

I also believe that you have developed a sound business plan, the market for students is robust, and the returns on investment will be strong. I therefore wish you the best as you develop this program and offer my full support. If I can provide any additional information or can be of any further assistance, please do not hesitate to contact me.

All the best,

Michael Platt, PhD
Director, Duke Institute for Brain Sciences

# DUKE
# INSTITUTE *for*
# BRAIN
# SCIENCES

**Michael L. Platt, PhD, Director**
**Professor, Departments of Neurobiology,**
Evolutionary Anthropology,
and Psychology & Neuroscience
Rm. B243F LSRC Building
Duke University
Durham, NC 27710

**Duke** | INITIATIVE on SURVEY METHODOLOGY
*at the* Social Science Research Institute

**140 Science Drive
Campus Box 90989
Durham, NC 27708
*dism.ssri.duke.edu***

April 2, 2015

Dear Tom and Robert,

I am writing in enthusiastic support of the proposed MS in Data Science.  I have evaluated the proposed MS program with consideration to three different hats that I wear at Duke:  as Director of the Duke Initiative on Survey Methodology (DISM), as Associate Professor of Political Science, and as an organizer of an interdisciplinary working group with the Initiative in Information at Duke (iiD).  With each perspective in mind, it is clear that the MS in Data Science promises to enhance the research and learning experience of current and future students and faculty at Duke University.

The increasingly data-driven nature of government, industry, and academic work is well-recognized.  I have seen this first-hand – in federal agency decision-making, as a member of the Scientific Advisory Committee to the U.S. Census Bureau; in the successes (and failures) of students in the academic job market; and in my research collaborations with those outside academia (e.g., Facebook, Durham County School System). While many of the social and behavioral sciences routinely incorporate statistical methods as a standard part of their graduate training, departments often do not have the resources or expertise to provide advanced training in the methods and skills that are required for "big data" analysis.  More fundamentally, the type of training required for such research requires expertise from many different fields: computer science, statistics, and the social sciences.  The proposed MS in Data Sciences is built on this interdisciplinary model.

As Associate Professor of Political Science, I recognize the value that this MS program will offer to undergraduates and graduate students in the social sciences, whether their career interest is to become a top candidate for a PhD program or a professional working in government or private industry.   It is telling that the Duke undergraduate—a statistics and mathematics major— who worked as an RA on my iiD working group project  ("Using Big Data to Understand the American Electorate") was just admitted to the top PhD program *in political science*.   As director of the Duke Initiative on Survey Methodology, I understand the intellectual need for "big-data" researchers to receive broad training, not simply in machine-learning methods, but also in the social science methods that allow data analysts to make

inferences and solve substantive research problems.  As an iiD working group organizer, I understand the value of interdisciplinary collaboration—our group includes undergraduates, graduate students, post-doctoral fellows, and faculty from political science, public policy, statistics, and engineering and computer science.  Moreover, our large scale, interdisciplinary and data-intensive project serves to benefit from the infusion of students with these skills and interests.

In sum, this is a compelling proposal – I hope it will be approved.   If I can provide any further information, please don't hesitate to contact me at (919) 660-4341.

Sincerely,

D. Sunshine Hillygus
Director, Duke Initiative on Survey Methodology
Associate Professor, Political Science

March 30, 2015

Tom Nechyba, Director, Social Science Research Institute
Duke University, Durham, NC 27708

Dear Tom and Robert,

Thank you for sharing your proposal for an MS degree in Data Science with me. The more I think about the possibilities this would open for Energy education at Duke, the more enthusiastic I find myself becoming about the program. It truly is an innovative model that allows for many paths of engagement, with the unifying theme of the increased importance of information and data in the decisions we have to make.

First, I can see how this proposal would enhance our already existing collaboration on the Energy Data Analytics Lab. We are quickly amassing significant amounts of data from various private and public sector entities, data that can allow us to break new ground on questions around energy use, energy efficiency and alternative energy sources. These data are ideal for the kinds of capstone projects you discuss in your proposal.

Second, your observation about how data science MS students could become engaged in data-intensive Bass Connections projects and other research initiatives resonated greatly with me, particularly as you have made project management part of your proposed graduate program. The sustainability of programs like Bass Connections relies on the deeper integration of graduate programs such as this with undergraduate training, enhancing what you refer to as an apprenticeship model involving vertically integrated teams in which everyone learns.

Finally, I very much like the financial model underlying your proposal. It explicitly seeks to create new partnerships in ways that are very innovative and inviting to many different programs at Duke (including the Energy Initiative). I can see how this will lead to additional course offerings that will benefit not only MS students but also advanced undergraduates and PhD students in various programs. And it will do so without imposing unsupported teaching burdens on units; rather it invites them to share in the tuition revenue when they offer relevant classes.

I therefore congratulate you on a well-crafted proposal and wish you all the best in implementing it. Please know that you can count on my support along the way.

Sincerely,

Richard G. Newell
Director, Duke University Energy Initiative
Gendell Professor of Energy and Environmental Economics, Nicholas School of the Environment

Duke University
DURHAM
NORTH CAROLINA
27708-0097

Patrick Bayer                                                                          Chair and Professor
Department of Economics                                                      Telephone (919) 660-1815
Box 90097                                                                                Fax (919) 681-7984

Thomas Nechyba, Director
Social Science Research Institute
Duke University
Durham, NC 27708

April 29, 2015

Dear Tom,

I am writing this letter in support of the proposal for the MS in Data Science. This is a very well designed program that will serve as a tremendous model for interdisciplinary programs at Duke.

This program would provide a natural complement for the suite of analytical Master's programs that the Department of Economics has developed over the past decade including our core MA program, our MS in Quantitative Financial Economics, and our joint programs with the Departments of Statistical Science (MSEM), Computer Science (MSEC), and Political Science (MAPE).

I can see lots of obvious benefits to the department and its programs from collaboration with this new program - through course offerings, co-curricular activities, a larger data science culture and community at Duke, and coordination in advertising, teaching, recruiting, and placement.

In order to coordinate our activities and ensure that these complementary programs support rather than cannibalize one another, I would propose that we form a voluntary consortium of analytical/quantitative masters programs in the social and data sciences at Duke. Representatives of consortium programs could meet a couple of times a year to talk about priorities and to coordinate activities. I envision that the consortium would include all of the programs associated with Economics, Statistical Science, Computer Science, Political Science, Data Science and potentially others.

Joint activities could be funded by contributions from consortium members roughly in proportion to program size. I can see lots of areas for potential cooperation:

1. Shared advertising and outreach to expand the applicant pool
2. Targeted student recruiting aimed at increasing diversity of applicants
3. Developing a joint application - applicants could apply to multiple programs simultaneously with a ranking of preferences (so that we could coordinate admissions across programs) and decrease the cost to potential applicants of applying to multiple programs.

4. Career activities - training, internships, lecture series, placement, etc.

Regular meetings among the chairs and/or DGSs for programs would also provide an opportunity to discuss course offerings, teaching needs, joint hiring needs/opportunities, and other decisions at the level of department/institute.

Overall, I think the MS in Data Science offers a qualitatively different degree from the ones already on offer at Duke and through the kind of cooperation with existing programs described here would help to strengthen Duke's reputation in related fields and improve the quality of related programs.

Best Wishes,

**DUKE**
THE FUQUA
SCHOOL
OF BUSINESS

**William Boulding**
Dean
J.B. Fuqua Professor
of Business Administration

Duke University
The Fuqua School of Business
100 Fuqua Drive
Box 90120
Durham, NC 27708-0120 USA

Tel +1.919.660.7822
Fax +1.919.684.8742
www.fuqua.duke.edu

January 19, 2016

Dear Review Committee:

I am writing in support of the Master of Science in Data Science (MDS) program being proposed by the Social Science Research Institute (SSRI) and the Initiative for Information at Duke (iiD). As you may be aware, the Fuqua School of Business is proposing a Master of Quantitative Management (MQM) which we believe is highly complementary to the MDS program. We have discussed potential synergies between MDS and MQM with Professor Tom Nechyba (one of the proposers of the MDS program).

We believe the programs complement each other in several ways. First, the MQM program is more applied than the MDS program, and has a narrower focus on business-related data analytic career paths. Second, the MDS has a much more interdisciplinary focus than MQM. Third, and related to the second point, we believe the courses being developed for the MQM program will offer greater breadth of elective choice for students in the MDS program. While the MQM program does not currently anticipate MQM students taking courses outside of Fuqua, this is a feature that we might re-consider in the future, particularly if there was value elsewhere in the University for such a change. Fourth, and most importantly, having two programs with focus on data science and data analytics will draw together and strengthen the scholarship across Duke in this area.

In summary, I support the MDS proposal and think it would be a valuable addition to Duke's graduate offerings.
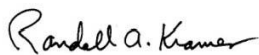
Sincerely,

William Boulding

Dear Tom and Robert,

I have followed the development of the proposed MS degree in data science with some interest and write to express my strong support for this program moving forward. Given the way it is structured, it promises to increase the offerings in applied data science across different areas, and will likely benefit both graduate and undergraduate students.

As you know, the Duke Global Health Institute (DGHI) has a number of education programs at both the graduate and undergraduate levels. Many of our undergraduates become involved in research, whether through Bass teams in Global Health or in other capstone experiences, and some of those experiences would be substantially enriched were they to include greater team expertise in data science. Already I am impressed with the Data+ and Data Expeditions programs you have started, as I think these will get more students interested in the possibilities of data science to connect to real world solutions. Having graduate students of the kind you describe participate in vertically integrated teams is very promising, as is the additional data science training that will likely be available to our best undergraduates.

We would also be interested in exploring potential course offerings in the electives that you plan for the MS degree. As I understand it, you have in mind a set of tools-based courses that will be offered in collaboration with units like Pratt, math, statistics and computer science, as well as more domain-specific electives that integrate the application of data science into topically focused areas. Connecting traditional global health students with data science students whose interests attract them to global health could be extremely powerful.

I am strongly supportive of your proposed degree, both because of its implications for DGHI and its likely impact on scaling Bass Connections in Global Health. I look forward to collaborating with you as you implement the program.

Sincerely,

Randall A. Kramer, Ph.D.
Deputy Director, Duke Global Health Institute
Professor of Environmental Economics and Global Health
Nicholas School of the Environment

Thomas Nechyba
Profesor of Economics and Director of SSRI
nechyba@duke.edu

22 January 2016

Dear Tom –

I am writing this letter in strong support of the proposed MS in Data Science. As the Director of the Information Science + Studies Program, as well as Co-Director of the Bass Connections Information, Society, and Culture theme, I can see many ways in which this program would work well with other programs and initiatives at Duke. Faculty and students involved in the Information Science + Studies Certificate, the MA track in Computational Media in Art, Art History & Visual Studies, and in the proposed Computational Media, Arts & Culture PhD could benefit from shared courses, workshops, and research projects. In addition, the Digital Humanities Initiative in the Franklin Humanities Institute would welcome the opportunity to connect these students to those interested in humanities applications of data science in PhD lab in Digital Knowledge.

Humanists are beginning to explore the possibilities for big data analysis of texts, images, archives, historic networks, and contemporary digital media resources, as well as in spatial media forms. Yet they often don't have the experience to do this work on their own. It is in everyone's interest that they do so with attention and connection to experienced practitioners in the field, and that we develop more sustained opportunities for mutual engagement, such as this MS. We can imagine, for example, electives in which critical thinking and communications around developing and analyzing domain-specific resources (image collections, literary texts, social media, historic archives) could take place, with students working on interdisciplinary project teams on ongoing projects. We can also envision co-developing Summer 0 opportunities focused around core skills development that would be of benefit to students interested in pursuing Digital Humanities as part of their graduate programs, or in alternate-academic careers post-graduation.

Sincerely,

Victoria Szabo

Victoria Szabo, PhD
Associate Research Professor in the Department of Art, Art History, and Visual Studies
Director of the Information Science + Studies Certificate Program
Director, Digital Humanities Initiative, Franklin Humanities Institute
Co-Director, Information, Society & Culture Theme, Bass Connections

1

**Eric J. Toone, Ph.D.**
Vice Provost and Director
Professor of Chemistry
140 Science Drive, Gross Hall Suite 102D
Campus Box 90467
Durham, NC 27708
phone: 919-681-3484
e-mail: eric.toone@duke.edu

March 24, 2015

Professor Robert Calderbank
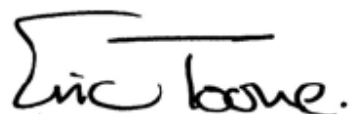317 Gross Hall
Duke University
Durham, NC  27708

Dear Robert:

I am pleased to provide my enthusiastic support for the proposed master's (MS) in Data Science submitted by the Social Science Research Institute (SSRI) and the Information Initiative at Duke (iiD). This proposal hits squarely on a number of important themes central to Duke's current strategic plan for curriculum development and our desire to foster a scholarly community engaged in interdisciplinary education. The interdisciplinary focus of this degree further supports our mission to graduate students prepared to address the most complex problems facing the world today.

I am particularly excited to see the emphasis on innovation in the application of quantitative methods and data analytics across a broad range of disciplines. Duke is uniquely positioned to lead in this arena and the emphasis will ensure that our strengths as a leading research institution will be central to the degree.  As noted in the proposal, the degree will not only develop technical competencies and a strong knowledge of data analysis and management, but also cultivate an understanding of the power of data-driven methodologies across the spectrum of disciplines, from the humanities and arts to the natural sciences, engineering, and medicine. I am confident it is in this regard that the degree will have the greatest impact on the students and, in turn, in the impact they then will have on the world.

The faculty members who developed this proposal have also created an novel and interesting funding model to encourage participation from across the university, allowing for the inclusion of a range of existing courses in the curriculum as well as for the development of new courses. This creative approach to funding will further promote the interdisciplinary aims of the degree and will result in a rich set of participants, both student and faculty.  I am excited by the possibilities of this funding model not only for this degree, but also for application in other parts of the university.

I am confident the MS in Data Science will be a meaningful and important addition to the university's academic offerings, and I warmly endorse your proposal.

With best personal regards,

*Eric Toone.*

**Scott A. Huettel, PhD**
**Jerry G. and Patricia Crawford Hubbard Professor**
**Chair, Department of Psychology and Neuroscience**

April 17, 2015

Tom Nechyba
Director, Social Science Research Institute
Duke University

Dear Tom,

I write to indicate my strong support for your proposal for a new MS in Data Science. I agree with the rationale for the program: Understanding how to manage, simplify, and draw inferences from data will underlie efforts to solve the major challenges of the coming decades. Students in this program will gain skills that will place them in high demand for a variety of positions, from research to policy.

I particularly appreciate two features of this program. First, it is expressly interdisciplinary, recognizing that data problems are not solved by rote application of technology; instead, a deep understanding of the structures of the underlying problem will be necessary and often gained through disciplinary analysis. Second, I am glad to see its focus on human decision makers (e.g., in the first course taken by students). The very label "big data" misses the mark, in my opinion, in that the data themselves are not an end goal; the critical advance comes when data drive meaningful inferences.

As the program matures, it could form natural connections both to the Psychology major and to our new certificate program in Decision Science. Both could provide pools of Duke undergraduates who see this master's program as a valuable extension to their studies.

I wish you (and Duke) great success with this program, and please do not hesitate to reach out if I can be of assistance in any way.

Best wishes,

Scott Huettel, Ph.D.

Jerry G. and Patricia Crawford Hubbard Professor
Chair, Department of Psychology and Neuroscience
Duke University

April 3, 2015

Social Science Research Institute and
The Initiative in Information at Duke"

To Whom It May Concern,

We are pleased to express our enthusiasm for the proposed Masters of Data Science degree program. With the increasing availability of both big data and advanced computing technologies, the culture of big data is ascendant in nearly every area of enquiry. This is especially true in healthcare. With the transition of medical records from paper to electronic, and the resulting outburst in captured data, more data scientists are urgently needed in many areas of healthcare, especially in nursing.

The increasing reliance of health care systems on information technology (IT) is opening up new opportunities for nursing and health informatics specialists who have expertise in both nursing practice and health care IT. Nurse informaticists play key roles in the development and implementation of health care IT systems in areas such as clinical documentation, computerized practitioner order entry, and electronic health records. Students in the nursing informatics major develop knowledge and skills in clinical information systems, strategic planning, project management, and a variety of technologies.

The promise of big data is that it enables us to raise new questions, create new knowledge about old problems, and create implementable solutions more rapidly. Our students now require techniques such as data and text mining, and predictive analytics, to make sense of the volume, variety, and velocity of healthcare data in the 21$^{st}$ century.

The proposed Masters of Data Science degree program is an opportunity for our programs and our School to collaborate. The Data Science program will offer courses that will likely be of much interest to our masters and doctoral students. Equally, our classes in health informatics will likely be of interest to Data Science students who are interested in the healthcare field and electronic health records.

There is a surging worldwide demand for data scientists and we believe that the Masters of Data Science is a needed and welcome addition to the Duke community. We are excited by this opportunity and are eager to collaborate with you and your team.
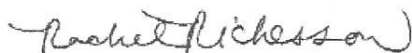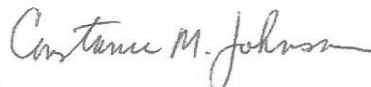
Sincerely,

*Elizabeth I Merwin*

Elizabeth Merwin, PhD, RN, FAAN
Ann Henshaw Gardiner Professor of Nursing
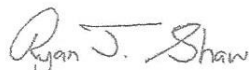Executive Vice Dean

*Janice Humphreys*

Janice Humphreys, PhD, RN, FAAN
Professor & Associate Dean for Academic Affairs

*Rachel Richesson*

Rachel Richesson, PhD
Associate Professor
Lead Faculty, Informatics Program

*Constance M. Johnson*

Constance Johnson, PhD, RN, FAAN
Associate Professor
Faculty Coordinator Systems Program

*Ryan J. Shaw*

Ryan J. Shaw, PhD, RN
Assistant Professor

# Duke University
DEPARTMENT OF SOCIOLOGY
BOX 90088
DURHAM, NC 27708-0088

EDUARDO BONILLA-SILVA
PROFESSOR AND CHAIR

TELEPHONE: (919) 660-5607
FAX: (919) 660-5623
E-MAIL: EBS@SOC.DUKE.EDU

1/11/16

Dear Dean Robert Calderbank:

I am writing to express my support for the MS in Data Science (MSDS) proposed by the Social Sciences Research Institute (SSRI) and the Information Initiative at Duke (iiD).

This is the era of unprecedented access to datasets from census, survey, geographic, satellite, administrative, communication, medical, and an ever-growing number of other sources. The Sociology Department has made information at scale a strategic priority and faculty led initiatives such as the Duke Network Analysis Center (DNAC) are flourishing, thanks in part to investments by SSRI and iiD. In fact resources contributed by iiD in 2014 made it possible for our Department to attract Chris Bail, a distinguished young scholar of social networks.

When Chris was hired, our Department expressed interest in development of a course in computational social science as part of his regular teaching assignment. We intended that this course would provide students with skills that would enable them to automate collection, analysis and visualization of social science data. This course is one of five courses in the core curriculum of the proposed MSDS program. The funding model that supports the MSDS program will make it possible for the Department to provide TA support for hands on data exploration that will benefit all our students. I anticipate that graduate students in Sociology, and some undergraduates too, will benefit from other skills based courses in the core curriculum.

I expect that the Data Dialog Seminar in the core curriculum will help connect social scientists and computational scientists. Datasets provide a common language, making it possible for computational scientists to learn what questions are important, and for social scientists to learn what questions can be solved. I am optimistic that this Seminar will bring social scientists and computational scientists together, and that it will be successful in catalyzing new collaborations.

I believe that the core curriculum will add value to the graduate program in Sociology, and I strongly endorse the proposed MSDS program

Eduardo Bonilla-Silva

Sincerely,

Eduardo Bonilla-Silva
Professor and Chair of Sociology
Duke University

1

**Professor Robert Calderbank**
**317 Gross Hall**
**Duke University**
**Durham  NC 27708**

**Dear Robert,**

I am pleased to support the proposed MS in Data Science (MSDS) submitted by the Social Science Research Institute (SSRI) and the Information Initiative at Duke (iiD).

The Department of Statistical Science recently introduced a MS in Statistical Science (MSS), a novel two year graduate degree program that provides a comprehensive education in statistical methods, computation, applications and relevant theory targeting students interested in careers in industry as well as preparing students for PhD programs.  The program is now in its second year and has attracted excellent students from over 500 applicants a year.  The proposed MSDS program aims for interdisciplinary breadth rather than depth in statistical science, and requirements of the program are complementary. The educational programs in Statistical Science will benefit from the proposed MSDS program, since some students may take MSDS courses as electives. The proposed MSDS courses on Data Marshaling and Data Visualization emphasize development of highly practical skillsets that are important for Data Scientists and will be useful to MS and PhD students in Statistical Science, as well as undergraduate majors pursuing projects leading to graduation with distinction. The Data Dialog Seminar which is already up and running and organized by Paul Bendich is designed to expose students to prospective clients, that is people with data who are looking for insight into their data. Students in our MSS program need to develop communication and project management skills as well as technical skills. The seminar provides a forum where they can start to understand and articulate client objectives, and then work with clients to translate these objectives into a delivered product or solution which could become part of their MSS portfolio, and provides additional opportunities beyond our Statistical Consulting Workshop  course for students to be involved in data driven modelling.

We expect that some students in the MSDS program will be interested in taking Statistical Science courses as part of their electives;  our courses on Modern Regression & Predictive Modelling, Programming for Statistical Science, Statistical Computing and Computation, DataFest, Probabilistic Machine Learning, Bayes and Modern Statistical Data Analysis, Statistical Networks, Statistical Learning and Bayesian Nonparametrics, among others will be of interest for students interested in the statistical side of Data Science.  The demand from PhD and MS students across the university for several of these courses exceeds course capacities, however, the proposed funding model  would

provide resources so that the Department would be able to open more sections and accommodate the demand.

I believe the MS in Data Science will provide valuable educational opportunities for students in a wide range of programs at Duke, including Statistical Science and look forward to working with you as the program evolves.

Sincerely

Merlise A Clyde

Department Chair
Professor of Statistical Science

March 30, 2015

Dear Robert and Tom,

I am delighted to add my enthusiastic support to the innovative MS in Data Science proposal you have shared with me. I know this is something you have been working on for the past two years, and I am thrilled that it has reached this stage. At a time when Duke is looking seriously into the wisdom of launching so many siloed Masters programs, this really stands out as a new approach that connects many units across schools in a shared enterprise and leverages these connections in ways that keep your own program from becoming siloed. I am very impressed.

Having read the proposal closely, I see the financial model as ideally suited to achieve the program you are trying to build. Rather than hiring your own faculty to establish the program, you are building in mechanisms to invite many different programs, with many complementary strengths, to help you create the connection from data science to data practice across a variety of domains. I can see how iiD and SSRI can play critical coordinating roles and jointly deliver the pathways along which many units will join the enterprise.

Indeed, your partnership in this effort makes a lot of sense. SSRI has much infrastructure as well as social science domains to bring to the program, and iiD has the nucleus of data science expertise and an initial set of connections that complement those of SSRI and extend to the humanities. Either of you could in principle establish a data science degree program, but it is the combination of the strengths you bring to this that makes it so distinctive, inviting and broad. Your already successful collaborations on a number of other fronts bode well for the success of this new collaboration.

You emphasize in your proposal the fact that the program is designed so as to invite departments and schools to become genuine partners – and to alleviate some of the pressure that smaller units are facing to establish their own Masters programs. Speaking for the interdisciplinary institutes and initiatives, I would furthermore emphasize how much these can also become real partners, particularly those that oversee Bass Connections.

As you know, we continue to search for new and innovative ways of making Bass Connections truly sustainable, and the expertise that your MS students will bring to so many Bass teams, combined with your focus on training them not only on technical fronts but also on practical matters like project management, could become a powerful new source of energy for Bass. In addition, while you are currently leading (or co-leading) two of the Bass themes, the importance of your MS program for Bass Connections extends across all five current themes and almost certainly would play a critical role in any new theme that might arise in the future. It is a prime example of how two programs – Bass and the MS in Data Science – will both be enriched by the existence of the other.

As you can tell, I truly am very enthusiastic about what you are building and wish you all the best as you bring this to fruition. If my office can be of any assistance along the way, please do not hesitate to call on us.

Best,

Susan Roth
Professor of Psychology
Vice Provost for Interdisciplinary Studies

# Duke
### UNIVERSITY

119 Allen Building
PO Box 90037
Durham, North Carolina 27708

**Lawrence Carin, Ph.D.**
Vice Provost for Research
Office of the Provost

Telephone: (919) 681-6438
Fax: (919) 668-5569

November 20, 2015

Robert Calderbank, Ph.D.
Director of the Information Initiative at Duke
CAMPUS Box 90984

Dear Robert:

I am pleased to report my enthusiastic support for the MS in Data Science proposed by the Social Sciences Research Institute (SSRI) and the Information Initiative at Duke (iiD).
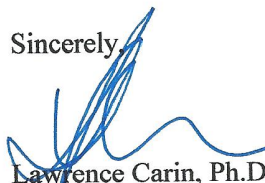
Research and education are in the early stages of major transformations, resulting largely from (i) increased availability of massive and rich datasets, (ii) advances in computational capacities and *big data* analytical techniques, and (iii) a recognition that complex questions are better answered by multidisciplinary teams. This is evident, for example, in major funding initiatives at NSF, DoE and DoD surrounding big data infrastructure and analytics. Universities at the leading edge of this transformation have the opportunity to make a substantial impact on the most pressing and complex challenges facing society. For example, I think that Duke Initiatives in Energy and the Environment will be more successful in impacting policy if stronger connections are developed with Information and Computation.

One of my priorities as Vice Provost for Research is to increase collaboration between the Duke Health System, the Pratt School of Engineering and the Trinity College of Arts and Sciences. Computation within the Life Sciences and Medicine is already important and will become absolutely essential in the very near future. This is not yet reflected in the training of graduate and undergraduate students. I see the core curriculum in the proposed MSDS program as starting to address the need for more skills based courses at the graduate level.

There is also a need for more robust and systematic connection between computational scientists and clinical practitioners. Datasets provide opportunity for ideas to flow back and forth between domain experts and computational experts. At the outset the computational experts often do not have a clear idea about what questions are interesting, and the domain experts often do not have a clear idea about what questions can be answered. In the proposed core curriculum I am particularly enthusiastic about the way the Data Dialog Seminar brings domain experts and data scientists together, and I believe that it will be successful in breaking down barriers to collaboration.

I anticipate that the Seminar will add value to many MS and PhD programs across the university, and strongly endorse the proposed MSDS program

Sincerely,

Lawrence Carin, Ph.D.
Vice Provost for Research
Professor of Computer Engineering

**Edward J. Balleisen**
Vice Provost for Interdisciplinary Studies

216 Allen Building
Box 90003
Durham, NC 27708

Telephone: (919) 684-1964
Fax: (919) 684-4421
eballeis@duke.edu

Feb. 9, 2016

Dear Robert and Tom,

Like my predecessor Susan Roth, I see the proposal for a new MS in Data Science as an exciting addition for Duke.  You have engaged in a remarkably thorough process of consultation, which has helped to shape an exemplary interdisciplinary endeavor that connects rather duplicates.  Roth's letter notes the strength of your innovative financial model, which engages departments and schools as full partners in a joint effort, and fairly distributes slices of tuition income from the MA back to those partners.  As Roth again stresses, this program has every prospect of creating new linkages across campus, and giving Bass Connections an additional jolt of student creativity and capacity.

Along so many others on campus, I have carefully read and digested last year's report on Master's Programs at Duke.  That report makes clear that we need to ensure that new Master's proposals will add something distinctive to our overall programmatic offerings, meet a clear student demand, and improve degree recipients' career opportunities.  The MS in Data Science meets these important criteria.  I am also impressed by your recognition of the opportunity to bring humanistic perspective to bear on the social construction of data, as well as training students in sophisticated statistical methods and the application of those methods to a wide array of real world problems.

I know that you are both committed to ongoing assessment and refinement of this innovative program, and that you look forward to deepening partnerships with A&S Departments and other Duke Schools. Facility with data analytics is an essential skill in the new world of big data.  I am convinced that this program will benefit the students who enroll in the MA, students in other degree programs who take one or more courses from the new offerings, the departments and schools whose faculty will offer core courses and electives, SSRI and iiD, Bass Connections, and Duke more generally.

As the new program launches, my office will be at your disposal to assist with any logistical issues that may crop up; to talk through early indications of the program's impact; and to think through ways to deepen cooperation with the key departments and schools.

All best, Ed

*Edward J. Balleisen*

Edward J. Balleisen
Vice Provost for Interdisciplinary Studies
Associate Professor of History and Public Policy

# Duke University

## Trinity College of Arts & Sciences

**DURHAM, NORTH CAROLINA 27708-0044**

Sandy P. Connolly
Senior Associate Dean Finance & Administration
Box 90044

Telephone: (919) 660-5264
Fax: (919) 684-8697
E-mail: sandy.connolly@duke.edu

## MEMORANDUM

**TO:**     Tom Nechyba
            Robert Calderbank

**FROM:**   Sandy P. Connolly

**SUBJECT:**   MS in Interdisciplinary Data Science

**DATE**:    September 13, 2016


Dear Robert and Tom,

Thank you for meeting over the summer about your proposed MS in Interdisciplinary Data Science and for sending me the revised proposal that you submitted to the ECGF.  This is to confirm what you already state in your response document to the ECGF regarding our discussions.

Several departments in A&S have raised the growing issue of students from one department's Masters program choosing courses in another department's program and thus imposing significant costs outside their home department. We are actively discussing ways of addressing this, and your financial model is directionally in line with our own discussions. Specifically, you have devised an approach to recognize these costs in your financial model, providing explicit course funding (for faculty participating in core courses) and setting aside $4,000 for each class a student takes in another department.

While our discussions are still ongoing, proposals that will emerge will likely contain similar elements to those you have included in your proposal. We hope that this will create an atmosphere of greater collaboration between departments, schools and institutes, as I know is your intention with your approach. We also appreciate the willingness you have expressed to modify your approach to complement whatever emerges in our discussions, as that seems appropriate.

Cc: Valeria Ashby
Linda Burton
Dan Kiehart

# Duke UNIVERSITY

Linda M. Burton, PhD
Dean of the Social Sciences
James B. Duke Professor of Sociology
Professor of African and African Studies

102 Allen Building
Box 90029
Durham, NC 27708-0029
Phone: 919-668-2746
Facsimile: 919-684-8503
socscidean@duke.edu

October 4, 2016

Dr. Tom Nechyba
Dr. Robert Calderbank
Social Science Research Institute
Duke University
Durham, NC 27708

Dear Tom and Robert,

We are excited to express our support for the innovative Masters in Interdisciplinary Data Science that you have proposed. As we learned at the lunch we had last spring (together with Richard Powell who was then Dean of Humanities), the degree paves the way for fostering innovative new partnerships across departments and schools while enriching both the undergraduate and the graduate experience. Your successful Data+ program has already demonstrated the potential for deep interdisciplinary data science to bring together faculty, students and outside partners in new ways, and we think this Masters degree will build on that success. It has similarly demonstrated the powerful way in which team-based and interdisciplinary data science can provide much needed professional development opportunities for graduate students.

As you describe in your revised proposal, the interdisciplinary focus of your proposal distinguishes it from most data science programs at peer institutions where such programs are typically housed in particular schools or disciplines. By branding it as a degree in *interdisciplinary* data science, you have left open the door for other, more disciplinarily focused data science programs to emerge at Duke and to connect to the larger data science landscape that SSRI and iiD are envisioning, and you are providing important pieces to help make such complementary efforts successful.

You have also thoughtfully tied the financial model to the underlying mission you have set for the degree, not only inviting partnerships across A&S and the wider campus but also explicitly budgeting to make such partnerships a win-win for all who engage. The Data+ experiment has already demonstrated the wide applicability of data science to virtually all the disciplines across the humanities, social sciences and natural sciences. It has furthermore demonstrated that data science is becoming an important tool for generating bridges across diverse "domain" disciplines (as you

call them). This is truly exciting and fully aligns with Duke's continuing effort to lead at the frontier of interdisciplinary innovation.
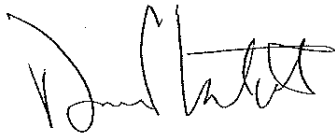
Finally, we are excited to see that some of the ideas that emerged during our lunch have made it into the revision of your proposal. For instance, we very much like the idea we jointly developed during our lunch to have open calls (in the form of RFPs) for collaborations both on enriching the curriculum with new electives and on developing research projects that can benefit from active engagement with your students as project leaders. This will no doubt link in new and innovative ways to all sorts of ongoing initiatives across A&S and the university.

We wish you the best of luck with this new program and look forward to working closely on intensifying our ongoing collaborations.

Sincerely,

*Linda M. Burton*

Linda M. Burton, Ph.D.
Dean of Social Sciences


Dan Kiehart, Ph.D
Dean of Natural Sciences

# Appendix D: Data Science Student Learning Outcomes Statement & Assessment Plan

**Student Learning Outcomes Statement**

Objectives with Outcomes:

Objective #1: To enable students to develop as **successful professionals** for highly competitive positions in industry, government, non-profits and academic departments – but with a nuanced perspective of both the power and the limits of data science.

Outcomes: By the time they graduate, students are expected to:
   a. Demonstrate proficiency in data science methods, tools, and applications, in research ethics, and in technologies used to explore novel and emerging problems;
   b. Contribute meaningfully to policy analysis, to basic or applied research drawing on the tools of data science, or to basic research advancing the field of data science.
   c. Secure employment that is challenging and respected within the profession.

Objective #2: To prepare students to be **effective researchers** in the field and to secure admissions to high-quality PhD programs.

Outcomes: By the time they graduate and in the following years, students are expected to:
   a. Secure admission to top PhD programs;
   b. Successfully advance through and graduate from their doctoral program;
   c. Develop effective research skills, including a nuanced sense of how data science fits into the broader picture of academic research; and,
   d. Develop presentation and teaching skills.

**Outcomes Assessment Plan**

Through the program evaluation team in SSRI, we will evaluate the MS program as it evolves. The general approach will be that of *developmental evaluation* – an approach to program evaluation appropriate to dynamic and changing programs where learning has to occur in real time and adjustments must be made as learning takes place. SSRI first adopted this approach in ongoing developmental evaluations of the Bass Connections Program, and we will apply simply methods here. As the program matures, the nature of each year's developmental evaluation will change – as it has in our evaluation of Bass Connections. But the purpose of the evaluations – to continually learn and improve – remains constant.

This developmental evaluation will furthermore go hand-in-hand with ongoing assessment, with assessment data feeding into the evaluation and the evaluation allowing us to fine-tune our assessment instruments. The following is therefore only the beginning of a plan for assessment, one that will develop over time as SSRI's evaluation team develops the ongoing developmental evaluation of the program.

OBJECTIVE 1: To enable students to develop as **successful professionals** for highly competitive positions in industry, government, and academic departments.

| Outcome | Evidence of Outcome | Frequency of Collection | Years Report Due |
|---|---|---|---|
| **Demonstrate proficiency in data science methods, tools, and applications, in research ethics, and in technologies used to explore novel and emerging problems** | End-of-program assessment | Annual | Annual |
| | Performance in Data Science courses | Semester | Annual |
| | Performance in advanced mathematics, computer science, and statistics courses | Semester | Annual |
| **Contribute meaningfully to policy analysis, to basic or applied research drawing on the tools of data science, or to basic research advancing the field of data science** | Research papers | Sporadic | Periodically |
| | Alumni reports | Biannual | Periodically |
| | Departmental payroll list of RAs and TAs | Annual | Annual |
| **Secure employment that is challenging and respected within the profession** | Survey of job placements of M.A. alumni | Annual | Annual |

OBJECTIVE 2: To prepare students to be **effective researchers** in the field and to secure admissions to high quality Ph.D. programs.

| Outcome | Evidence of Outcome | Frequency of Collection | Years Report Due |
|---|---|---|---|
| **Secure admission to top Ph.D. programs** | Student reporting in response to surveys | Annual | Annual |
| **Successfully advance through and graduate from their Ph.D. program** | Surveys of alumni who entered doctoral programs | Annual | Annual |

| Develop effective research skills | Student portfolios | Semester | Annual |
|---|---|---|---|
| **Develop presentation and teaching skills** | Student portfolios (which also will ask about RA and TA experiences) | | Annual |
| | TA evaluations | Annual | Periodically |
| | | Semester | |

**Appendix E**: Course Titles

ARTHIST     231.01 – History of Art and Markets (cross-listed with VMS 242.01)
                508S.01 – Art and Markets (cross-listed with VMS 567S.01)
                (541S.01 – Advanced Digital Art History)

CLST         544L – Digital Archeology

COMPSCI    316 - Introduction to Database System
                516 - Data-Intensive Computing Systems
                570 - Artificial Intelligence
                571D - Machine Learning
                579 – Statistical Data Mining

ECON        432S – Environmental Justice: The Economics of Race, Space and Pollution
                461 – How Markets Work: Theory, Evidence and Empirical Tools
                471 – Financial Markets and Investment
                514 – Fixed Income Markets and Quantitative Methods
                553 – Labor Economics
                554 – Urban Economics
                608 – Introduction to Econometrics
                612 – Time Series Econometrics
                613 – Applied Econometrics in Microeconomics
                673 – Mathematical Finance
                674 – Financial Derivatives
                703 – Econometrics I
                707 – Econometrics II
                881 – Various Special Topics Modules in Microeconomics
                883 – Various Special Topics Modules in Econometrics
                887 – Various Special Topics Modules in Financial Econometrics

MATH        361 – Mathematical Numerical Analysis
                404 – Mathematical Cryptography
                476S – Seminar in Mathematical Modeling
                541 – Applied Stochastic Processes
                545 – Introduction to Stochastic Processes
                561 – Numerical Linear Algebra, Optimization and Monte Carlo Simulation
                563 – Applied Computational Analysis
                573S – Modeling Biological Systems
                581 – Mathematical Finance

|        |                                                                      |
|--------|----------------------------------------------------------------------|
|        | 582 – Financial Derivatives                                          |
|        | 640 – Probability                                                    |

PSY
- 353S – Medical Decision Making
- 673S – Computer Models and Treatment of Psychiatric Disorders
- 681S – Genetics and Environment in Abnormal Behavior
- 685S – Biological Pathways to Psychopathology
- 768 – Applied Analysis of Variance
- 767 – Applied Correlation and Regression Analysis
- 768 – Applied Structural Equation Modeling
- 770 – Applied Multilevel Modeling
- 781S – Data Methods in Cognitive Psychology

POLSCI
- 230 – Networks in a Globalizing World
- 330 – Quantitative Political Inquiry and Evaluation
- 508S – Public Opinion and Behavior
- 550S – Predicting Politics: Counter Insurgency, Elections and Stability
- 551S – Voting Behavior
- 632 – Computational Political Economy
- 634 – Social Networks and Political Interdependence in Political Science
- 733 – Maximum Likelihood Methods
- 748 – Advanced Quantitative Research Methods in Political Science

PUBPOL
- 574 – Economic Evaluation of Sustainable Development
- 597S – Seminar in Applied Project Evaluation
- 604 – Using Data to Analyze and Evaluate Public Policy
- 605 -  International Trade and Policy.
- 607 – Cost-Benefit Analysis for Health and Environmental Policy
- 633 – Topics in Population, Health and Policy (cross-listed)
- 637 – Health Policy Analysis

SOC
- 333 – Quantitative Analysis of Sociological Data
- 534 – Topics in Population, Health and Policy
- 643 – Global Inequality Research Seminar
- 690 – Special Topics in Sociology – Computational Sociology
- 720 – Survey Research Methods
- 726S – Advanced Methods of Demographic Analysis
- 728 – Advanced Methods: Introduction to Social Networks
- 790 – Graduate Seminar in Visualization

STA 521L - Modern Regression and Predictive Modeling
523L - Programming for Statistical Science
561D - Probabilistic Machine Learning
571 - Advanced Probabilistic Machine Learning
601 – Bayesian and Modern Statistics
640 – Causal Inference

VMS VMS 242.01 – History of Art and Markets (cross-listed with ARTHIST 231.01)
567.01 – Art and Markets (cross-listed with ARTHIST 508S.01)