Teaching Statement for Kirk Borne *(updated November 5, 2010)*

Introduction – Why I came to Mason. If I came to Mason 10 years ago, my teaching statement would have focused entirely on astronomy – *teaching the masses about the wonders of the Universe.* The students would naturally love every bit of it, because everyone is enthralled by astronomy: black holes, the big bang, the possibility of life out there, pretty pictures, and more! But, a funny thing happened about 9 years ago – I discovered the incredible science of data mining, which is the application of mathematical and computer science algorithms to the problem of discovering hidden (sometimes surprising) knowledge within large databases. I soon realized that this was an absolutely critical skill set for every future scientist and (in fact) for every future citizen. This is because all of science, government, and industry is generating massive (and exponentially) growing quantities of data. Without training in the skills of Data Science (such as data mining, visualization, statistics, domain modeling, and data management), science disciplines and societal organizations will never reap the full benefits (scientific or otherwise) from their enormous data collections.

As my fascination with the subject of Data Science sky-rocketed and my appreciation of its applicability to my astronomical research problems became crystal clear (as well as its applicability to countless other disciplines), I became convinced that the training and education of future scientists must be transformed dramatically (not an evolutionary change, but a revolutionary transformation). It was at about this time that I discovered the CSI (Computational Science and Informatics) program at Mason. I was working at NASA at the time. I decided then that Mason's CSI program was where I wanted to be, and so I began a dialogue with faculty here that ultimately led to my joining the CSI program in summer 2003.

Remarkably, I found that there was already a course in the CSI core requirements that matched my dream course – Scientific Databases (CSI 710)! I was soon co-teaching that course with Dr. Larry Kerschberg, who ultimately transferred the course to me in 2005.

At the same time, I was an adjunct associate professor at UMUC, where I taught a course in Data Mining for graduate students in the Database Technologies Program. This was interesting but not a science course. While teaching this course, I re-discovered a fundamental truth in education – the best way to learn anything is to teach it! I learned so much about data mining (how wonderful and rich it is) that I decided that my new dream course would be to teach Scientific Data Mining to science students (undergraduate and graduate). When Mason formed the Department of Computational and Data Sciences, a new undergraduate B.S. program emerged. Working with Dr. John Wallin, I helped to create the curriculum, the courses, and their syllabi. One of those courses was Scientific Data Mining (CDS 401), and another is the undergraduate version of Scientific Data and Databases (CDS 302). I am now the instructor in all of the above courses. I spend nearly 15% of the semester in my CSI 710 graduate class on Scientific Data Mining topics.

It surprises my astronomy colleagues around the country that I almost never teach an astronomy course. That's fine with me, because I teach students how to carry out scientific discovery from large databases for any discipline. It is revolutionary, exciting, and very engaging for the students.

Since astronomy is growing its own data-oriented research sub-discipline, called Astroinformatics, my next "dream course" will be to teach Astroinformatics at Mason. After discussions with the chairs of the CDS and Physics & Astronomy departments, I believe that this course is likely to become a reality within the next year. This will become one of the first such courses in the world

– a highly desirable outcome expressed by a large community of astroinformatics researchers at an international conference[1] (at Caltech) in June 2010.

In all of these cases, my goal has been to enable students to develop the skills that are needed to make discoveries from "big data", whether they do it in the future as a practicing scientist, as an industry employee, or as a government employee. The courses that I am now teaching contribute to workforce development for all of society in the Information Age of data glut.

I am now starting to catch a glimpse of the fulfillment of the goals that I set for myself when I first came to Mason – to *teach the masses about the wonders of Data Science*, to train the next generation of scientists in "next-gen science", to transform science education, to teach my colleagues on the faculty about the importance of this scientific methodology, and to export these ideas and curricula to other universities. Regarding the last point, Mason is now a recognized leader and at the forefront of change in graduate and undergraduate science education – I have received numerous invitations from other universities to visit and tell them what we are doing and how it is done – they want to emulate our CDS and/or CSI programs (at least, some of our courses).

Data in the Classroom – Education Research. There has been a dramatic revolution in the conduct of science in recent years, primarily triggered by the digital revolution. Online science, collaboratories, data-intensive science, digital libraries, and scientific social networks have changed the character of science practice. Despite this revolution in how scientists do their work, there has been very little change in standard science courses and curricula to match this reality. Numerous professional society working groups, national academies committees, and national agency blue ribbon panels have produced reports[2] that call for dramatic changes in science education at all levels. In one case, a panel of experts produced the special report[3] "Transform Science: Computational Education for Scientists"; and in another case, a landmark journal article[4] on "Computational Thinking" is transforming numerous NSF research programs and proposal requirements. I led an interactive presentation session[5] at Mason's ITL'2010 (Innovations in Teaching & Learning) conference where I presented some of the recommendations from the national study reports and opened a dialogue among Mason faculty on how we can transform our own science courses and curricula to match the current realities of scientific research in the digital age. Such a transformation will better prepare our students for the world into which they will be embarking, and will aid in the development of the computationally skilled and data-savvy workforce required in the 21st century workplace. These new data-oriented science programs will have the additional benefit of attracting the best and the brightest students to Mason when they see this transformative science being taught here. These programs will also prepare students in an exemplary manner for the next steps in their careers. Additional benefits include the focus on inquiry-based learning, problem-based learning, the natural inclusion of undergraduates in the scientific scholarly community (consistent with Mason's QEP initiative "Students as Scholars" Objective 2: Increase Opportunities for Scholarly Inquiry), and the proven attraction of diverse (even under-represented) student populations into science (through their pre-existing familiarity with information technology tools)[6].

[1] http://www.astro.caltech.edu/ai10/

[2] See the list of reports in Appendix A: http://www.nsf.gov/mps/dms/documents/Data-EnabledScience.pdf

[3] http://research.microsoft.com/transformscience/

[4] http://www.educause.edu/Resources/ComputationalThinking/196295

[5] http://cte.gmu.edu/events/scheduleten.html#1E

[6] http://democrats.science.house.gov/Media/file/Commdocs/hearings/2010/Research/16mar/Stassun_Testimony.pdf

As part of this effort, I have been investigating the research pertaining to using data in the classroom[7], which demonstrates the efficacy of teaching science (or any discipline) through students' hands-on experience with real data. Engaging the students in research questions using data is directly in line with Mason's QEP initiative (Exhibit 10: Potential Orientations of Research- or Inquiry-Based Curricula). This form of inquiry-based learning is highly effective. It is consistent with the constructivist learning process within the context of the 5E learning model[8]: Engage, Explore, Explain, Elaborate, Evaluate. It is not surprising that this method is so successful when we consider that every science problem can be cast as a detective story – everyone loves a detective story (*e.g.*, astronomy in particular is a forensic science, entirely evidence-based, since we cannot build a star or a galaxy in our labs, or go to a star or galaxy or quasar to see what makes it do those wondrous things that it does). This "forensic" approach is actually applicable to all disciplines. Therefore, the introduction of inquiry-based learning, through the use of real data in the classroom, across all disciplines and at all grade levels (pre-college, college, and graduate) is a tranformative, engaging, and effective approach to teaching science (and other topics). Of course, for this to be effective (*i.e.*, using data in the classroom), data science skills must be learned (by the instructors and the students), appropriate to the grade level of instruction.

In order to pursue the education objectives described above, I have obtained education research grants (and I am seeking more such grants). For example, I obtained a COS Pedagogy grant, plus I am participating in a self-study group at Mason (called SOSTC = Scholars of Studying Teaching Collaborative[9], led by Dr. Anastasia Samaras in the College of Education) to investigate my teaching practice (for consistency with my own teaching goals and philosophy), and I have an NSF CCLI (Course, Curriculum, and Laboratory Improvement) grant (called CUPIDS = Curriculum for an Undergraduate Program in Data Sciences) to develop a data science curriculum and its courses. I reported on our CUPIDS program at the 2009 International Conference on Computational Sciences[10]. In each of these education research experiences, I have been learning a new vocabulary while learning how to design courses and educational materials using the principles of "Understanding by Design" (UbD) promoted by Wiggins & McTighe[11]: in UbD, *"the teacher starts with classroom outcomes and then plans the curriculum, choosing activities and materials that ... foster student learning."* For example, the primary outcomes of the CUPIDS program are that the students will learn how to:

⊙ access large distributed data repositories,
⊙ conduct meaningful scientific inquiries into the data,
⊙ mine and analyze the data, and
⊙ make data-driven scientific discoveries.

The above philosophy, goals, and guiding principles apply to informal education research as well, such as Citizen Science (one of my major areas of research) through the use of public-facing science interfaces to educate the general public about science and the scientific process of discovery. I am working with the education team on the LSST (Large Synoptic Survey Telescope) project specifically in these areas, including pre-college and undergraduate education, Citizen Science, and public outreach. We have even submitted an unsuccessful (but very good) proposal to the US

[7] Using Data in the Classroom: http://serc.carleton.edu/usingdata/
[8] http://faculty.mwsu.edu/west/maryann.coe/coe/inquire/inquiry.htm
[9] http://mason.gmu.edu/∼asamaras/selfstudy.htm
[10] Borne, K., Wallin, J., & Weigel, R., "The New Undergraduate Program in Computational and Data Sciences at GMU," Lecture Notes in Computer Science, 5545, 74-83 (2009)
[11] http://en.wikipedia.org/wiki/Understanding_by_Design

Department of Education, to teach data mining to school kids – the proposal was titled "CSI The Cosmos" (a spinoff on the very popular CSI forensic science television shows).

Some Innovations. I list here a few of the teaching innovations that I have implemented:

- Transforming science education – recognizing that the future of science is computational and data-driven, and in line with Mason's QEP "students as scholars" initiative, I attempt to increase awareness of scholarship among students and then have them participate.
- Crowd-sourcing the exam – I ask the students to suggest their own questions for the exam. This has been marginally successful, since some of their questions are quite shallow or trivial, but not all of them.
- Teaching through distance education – this is not particularly novel or innovative, but it is new for Mason and especially new for the CDS department.
- Engaging the students in research questions using data (inquiry-based learning) – this is hard to implement, but it works, since every science problem can be cast as a detective story and everyone loves a detective story. This actually is applicable to all disciplines.
- Teaching Data Ethics – teaching students how to make objective, evidence-based conclusions based on the data. This is an essential skill in the modern world in which we are deluged with information from countless sources (reliable and unreliable).
- Extending inquiry-based learning to the informal education arena, including public lectures and Citizen Science.
- Adding a module on grant-writing in my graduate courses – this is a form of professional development for future researchers. I have three decades of experience in grant-writing. This is offered as a resource to Mason students, who are embarking on their own careers in science, research, and proposal-writing. (I have not formally added this module to my classes yet, but I have conducted informal class discussions.)

Concluding Remarks. In 2009 I summarized my vision for transforming Astronomy education in the current era of "Big Data" in a national study report that I co-authored[12]: *"The Revolution in Astronomy Education: Data Science for the Masses".* The rationale and recommendations in that report are applicable to other science disciplines as well, and as such the report could serve as a blueprint for transforming science education. I have had some success so far in applying the above teaching philosophy to my undergraduate and graduate courses, to undergraduate research experiences, and to graduate student mentoring. For example, I have applied this philosophy during my term as CDS Graduate Program Coordinator, and I am now applying this as CDS Undergraduate Program Coordinator. I tell the students all about the revolution in Science (*i.e.*, the 3rd Paradigm = Computational Science, and the 4th Paradigm = Data-intensive Science = Discovery Informatics), I encourage them to participate, and I tell them how Mason is at the leading edge of these developments in both research and education. In addition to these activities, I am collaborating with faculty across Mason (*e.g.*, in the College of Education and science departments) to turn these ideas into reality. I am excited to be here at this time doing these things.

[12] http://arxiv.org/abs/0909.3895