

**Academic Statement of Purpose | Sean Kross | PhD in Information | UM ID: # 98264382**

The event that precipitated my interest in Information Science occurred when I was an undergraduate working at the NYU Center for Genomics and Systems Biology performing “wet” lab bench experiments and processing their results. Our lab produced data at a faster rate than it could be analyzed so I taught myself how to use the R programming language and started hacking on the data. At the time I was untrained in data analysis but I quickly realized that with R I could automate a task that I usually had to perform by hand using Excel. This R script saved me a significant amount of time and I gave the script to a colleague in the lab who was used to doing the same task by hand. Within a week the code was being used across two floors of the Center for Genomics. This was the first time in my life I had ever created something that spread so quickly and seemed to have so much utility. After speaking with each of my colleagues who were using the script I was interested to discover the small modifications they made to the code to fit their needs and how those tweaks propagated to other users. This experience made me feel that I had uncovered a talent and passion that I didn’t know I possessed.

I had no formal training in computer science and I was in my senior year at NYU, but my newfound passion was so strong that I was determined to learn more about writing software. I immediately enrolled for computer science courses at The University of Maryland which started just a few weeks after I graduated from NYU. While at the University of Maryland I was introduced to Jeff Leek, Roger Peng, and Brian Caffo, all professors in the Department of Biostatistics at the Johns Hopkins Bloomberg School of Public Health. They were building a series of massively open online courses in partnership with a company called Coursera. They were looking for folks who were interested in helping them build nine courses in data science which would run every single month. By the time I met them I had completed enough statistics and computer science courses to be able to meaningfully contribute to the courses they were

building. While working with them I was introduced to Nick Carchedi, then a graduate student at Hopkins who was looking for collaborators on a software library for R called “swirl” which interactively taught students how to use the R programming language within R’s programming console. I joined the swirl development effort and I have since become the package’s maintainer and lead developer. Since that series of data science courses launched on Coursera over two million people have enrolled in at least one of the courses, and over one million people have downloaded “swirl.” Both of those experiences have been extremely valuable for learning about how students want to consume educational content online, and what can be expected in terms of the computational resources available to students all over the world. It has also been an incredible exercise in developing workflows for creating and sharing open source educational content.

It is my aspiration that my primary focus in the School of Information would be in the field of Human Computer Interaction, specifically studying and designing systems to distribute online education. Data, information, and new knowledge are being generated at their fastest pace in human history, therefore methods must be developed in order to help the mind comprehend this data deluge in order to take full advantage of the current era of discovery. At the same time personal and mobile computing has become ubiquitous to the extent that even a majority of the global poor use mobile phones every day. Meanwhile the cost of computing and data storage continues to approach zero thanks to platforms like Amazon Web Services. It is my highest priority and career goal to see that these three factors (abundance of new knowledge, ubiquity of devices, and negligible cost of computing) converge, creating new platforms for learning that are accessible to everyone on Earth. I hope that before the end of my life it will be seen as unethical to hoard knowledge and educational opportunities due to the ease with which they can be created and distributed.

With regard to the present day, the original excitement about massively open online courses has shrunk significantly. Online courses are not taken as seriously compared to traditional in-person courses. One of the reasons for this discontinuity is that most online assignments resemble assignments from before the advent of online education (multiple choice questions, fill in the blank, short answers, etc). These types of questions take no advantage of the online medium which users expect to be interactive, dynamic, and colorful. Part of the reason that online assessments haven't evolved is because there aren't good tools for creating the new kinds of assessments that the digital medium begs. Creating new tools for the assessments of the future is one group of projects I would like to explore in graduate school. Another serious issue is the ability to verify the identity of a student who is submitting work online, and being sure that their submission is their original work. I would also be interested in developing technologies and methods for securing student's integrity in online education.

A secondary interest of mine is the role of technology in the lives of healthcare providers and their patients. The modern floor of a hospital is a cacophony of alerts and alarms of which few indicate actual medical emergencies or even convey any actionable information to healthcare workers. This phenomenon induces "alarm fatigue" among healthcare workers, desensitizing them to alarms and making them more prone to miss the occurrence of a medically relevant alert.

I'm interested in working on methods for prioritizing these notifications and organizing them so that whatever information is contained in these alarms can be interpreted meaningfully.

The time spent performing clerical work by the most junior nurse all the way up the medical hierarchy to the most senior physician is beginning to eclipse the amount of time that healthcare providers actually spend with patients. Documentation and compliance are of course important tools for ensuring a patient's well being, however the computer interfaces that

healthcare providers use for documentation are not currently designed for taking advantage of a healthcare provider's expertise. I often find myself saying to healthcare providers: "Imagine if Apple designed your iPhone like your electronic health record system. They'd quickly go out of business." I'm interested in developing open source alternatives to current EHR systems which are consciously designed for healthcare providers in order to ensure that they can spend the bulk of their time seeing patients.

Another major issue in healthcare is the void between a patient's behavior in a clinical setting versus their behavior outside of clinical setting. One of the most fascinating things I've learned while working at a school of public health is how the best treatment for asthma patients is to enroll them in observational studies where their inhaler use is monitored. This might be due to the lack of communication between healthcare providers and patients in between visits and hospital stays. I'm interested in designing systems in order to sustainably strengthen the provider-patient relationship so that patients can better understand and monitor their own health, and providers can intervene in the management of a patient's health before the patient has to visit the hospital.

In terms of faculty I'm interesting in working with, Dr. Brooks would be a natural fit considering his interests in massively open online courses and data science. I'd be interested in working with him on feedback loops where analysis of online course strengths and weaknesses inform future decisions in updating course content and creating new courses. I'd also like to explore how students can help each other learn by commenting on each other's work. I'm also interested in working with Dr. Adler-Milstein in the Lab for Health IT Impact. Her work with health information exchanges and the challenges they present in terms of data quality, and the difficulties that come with sharing those data, are areas where I would be interested in designing solutions. I recently discovered the work of Dr. Ceren Budak via social media, and her

experience researching media bias is now more important than ever considering the proliferation of “fake news” in the recent Presidential election. An accurate understanding of how news is shared in social networks is going to become vital for the public’s understanding of our democracy.

One research project I’m interesting in pursuing is the development of an online education platform that focuses primarily on teacher’s workflows and emphasizes the quality of online assessments. Major open online course platforms like Coursera and edX try to serve as a home for a course, packaging together all of the materials for a course with the assignments, projects, and a discussion board to facilitate class communication. Each of these platforms implements their own version of a video player, a discussion board, and the assignments are limited to formats that are digital translations of traditional paper based activities. These education specific implementations of online tools for video and discussion are often hobbled versions of products that are being perfected elsewhere online. YouTube and Vimeo have mastered online video and Piazza, Discourse, and Reddit are constantly trying to improve the quality of our conversations. Why should students be forced to use subpar products, and why should resources be spent on their maintenance when others are incentivized to improve these services? Another issue with content on these platforms is that they force course creators to enter readings, assignments, and other content in a rich text editor in a web browser. Although this method may be user-friendly for some it prevents programmatic and scalable updating of materials and links, raising barriers for course creators to keep their content up to date.

In order to address these concerns I propose building an open source online education platform that focuses mainly on giving course creators tools to make quality assignments. This focus would allow materials and discussions to be hosted on services that specialize in those spaces so that course creators can have larger creative possibilities for teaching their students.

The platform would feature all of the current canonical question types (multiple choice, fill in the blank, etc) but it would also feature a grading API that course authors could use if they wanted to implement their own platforms for delivering assignments. The platform I propose would also serve as a certification system for whatever topic a course creator wants to cover. In other words, if a student wants to immediately start completing assignments in order to show that they have mastery in a field before looking at any of the material they will be encouraged to do so. I envision the platform awarding certificates for completion of courses, however these certificates will have little value at first. In order to compensate for this, the bank of assignments created by a course author will be made available to employers so that they can verify the expertise of a particular student. If employers pay for access to the platform then it could be used as a vehicle for employer-subsidized education. This platform would also assuage the interviewing process by providing relevant questions for employers to ask candidates and it would provide candidates with a framework of expectations about what skills they need to have in order to get hired.

The interdisciplinary nature of Information Science and Human Computer Interaction has drawn me to both fields. I feel that my best work has been the result of combining seemingly disparate parts of my background in statistics, computer science, biology, and even the arts. Today's biggest problems could be solved more quickly if the technology we used to learn and to work was designed to be more humane - technology that is able to take advantage of all human abilities, senses, and talents. The most important work of this century is not going to be performed sitting for hours in front of a glowing screen. I'm determined to be one of the many people who will shape the future of work and education, and it would be a privilege to be able to launch that journey from the University of Michigan.