DS4PS Symposium Narrative

Faculty from 15 schools of public affairs were invited to the 2.5 day symposium. In the end, 13 faculty from 9 institutions attended the working sessions in Atlanta. (See list of symposium invitees and attendees.) These faculty were already teaching data science courses in their MPA/MPP programs or teach quantitative courses adjacent to or extensions of data science.

The first day of the symposium posed three questions to attendees:
- What is unique about data science in the public sector?
- What are working assumptions regarding the student body and career paths distinct to the public sector?
- How do these issues inform the unique curriculum needs in schools of public affairs.

The day started with an overview of the landscape of data science in public affairs, which sparked rich discussion about the compliments and tensions between the evolving field of data science and public affairs applications. While data science applies to all sectors in society, its most robust uses to date have been for private/for-profit entities. Its principles of open source, transparency, and efficiency serve businesses well, where consumers and organizations are the units of analysis for prediction and automation. In contrast, the public and nonprofit sectors value open data, individual privacy, and equity to drive data-driven solutions primarily for communities. The data used in the public sector are largely administrative and survey-based, and its data science applications aim to monitor population behavior and inform/evaluate policy interventions to social problems. Many data science degree programs and educational resources do not use government-generated data in their courses, nor do they grapple with problems/research questions that are trying to optimize social benefit. These domain-specific needs make the case for developing curricula specific to the "public interest" where it is lacking in traditional data science programs. Schools of public affairs slowly are building these resources, but the capacity to do it quickly enough to meet demand is lacking. The case for the consortium is that these resources can be built more quickly and with higher quality when institutions work together to create/share materials or entire courses.

The day one overview was followed by an online industry panel and alumni panel attended by faculty from approximately 30 schools of public affairs. The industry panel included representatives from large open data/smart city initiatives (New York City and the Urban Institute) and a faculty member from Carnegie Mellon University with extensive applied experience in local and federal governments identifying technology solutions to improve public service delivery. These experts discussed what they see as the essential skills and mindsets for students who want to pursue data science in the public and nonprofit sectors. All agreed that the "hard skills" of data science are important for those who may work on the technical side, but they emphasized the importance of domain knowledge about the public sector. They see tech savvy contractors hired by the public sector to implement data science solutions repeatedly fail to build sustainable programs/processes because they do not understand the bureaucratic structure of government and fail to take the time to listen to the end user's needs. Panelists agreed that students graduating from schools of public affairs can develop sufficient technical data science skills to direct the design of or effectively manage teams of data scientists, even if they do not become highly skilled in programming/engineering themselves. For students who do want to develop the technical skills, they agreed that concentrations in data science/data analytics within MPA/MPP programs can serve students well enough to get an entry level job in a public/nonprofit organization's data-focused units, and on the job training and intentional upskilling can occur to gain more technical expertise.

The second panel included four young alumni from MPA/MPP programs who are working in data-intensive agencies in/for the public and private sectors. They shared their opinions about the value of their data training in their programs and how it prepared them to be successful in their current positions. All agreed that the public sector focus of their degrees was essential to understanding the social problems they were addressing and the institutional challenges of collecting, managing, and analyzing data to address them. They shared that although their programming and analysis skills were basic at graduation, they often were the most "skilled" among their longer-tenure colleagues and tasked with figuring out solutions quickly from open sources on their own (having learned how to ask the right questions in their programs). Their advice to MPA/MPP directors designing curricula to meet the needs of the public and nonprofit sectors was to introduce more "messy" data into coursework and to teach students how to "fetch" data in SQL or other calls from large databases. They also expressed a need to learn multiple applications quickly and how to teach data analytics to their teams/direct reports after setting up systems/protocols in their agencies, especially to those who have limited quantitative skills. In addition, most emphasized that higher level statistics and machine learning skills were not widely used in their work and could have been replaced in the curricula with data wrangling/ management skills, effective communication techniques, and applied cases to practice in an environment where it is "safe to fail" instead of having these experiences on the job with promotion and/or public consequences. Lastly, they wished they had been exposed to the different paths/careers in data science while still in their programs and how to communicate their skills as they applied for jobs in technical units. They did not understand their best fit on data teams or even that they had value to offer (e.g., design, communication, project management, problem translation) outside of the traditional roles in data science (e.g., programmer, engineer, etc.) Interestingly, they did not recommend that universities develop general training modules to upskill alumni, since so many resources are free online. Instead, they advised offering micro-credentials/certificates that address specific public sector data needs (e.g., emergency management tools, sustainable administrative data tracking in Excel, etc.) to enhance data use and value to agencies.

After learning from industry experts and alumni, day 2 of the symposium focused on what a data science curriculum should look like in public affairs and how programs are designing them with current resources. The guiding questions for the day were:
- What are programs doing currently?
- What should a core curriculum for data science in public affairs include?
- How can a consortium of schools of public affairs support data science curricula?

The day began with a presentation of the data project life cycle and a map of existing courses from schools of public affairs using the cycle, as adapted for the public sector (Figure 1). The map includes a series of "on ramp" courses that expose students who come from non-technical undergraduate majors to coding in R and Python using public sector data and/or use cases. These courses often are pre-requisites and/or requirements in a four-five course sequence for a concentration in data science/analytics. The next set of courses mapped were identified as "core" to the project life cycle. They include introductions to computation thinking, big data sources/uses, data ethics, database design/management, data security, data wrangling, machine learning, data visualization, and practicums/capstones. The last set of courses were grouped as "extensions" and include applications that are specific to the data use case. These include GIS, social network analysis, and text analysis.

Figure 1
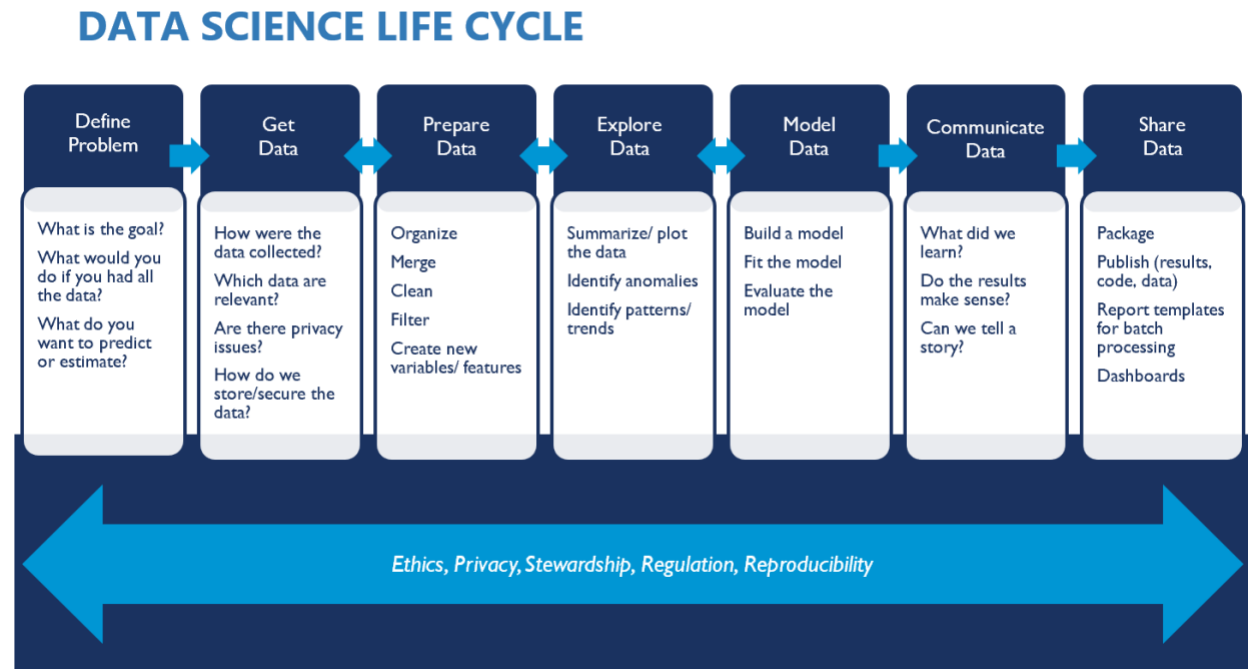
Faculty attending the symposium gave lightning talks that described how their MPA/MPP programs currently offer data science training. These presentations identify related curricula for Georgia Tech – Ivan Allen School, Georgia State University – Andrew Young School, North Carolina State University, Syracuse University – Maxwell School, University of Illinois Chicago, University of Minnesota – Humphrey School, University of Southern California – Price School, and the University of Texas Austin – LBJ School. (Links are the PPTs for each school.) Programs at institutions with engineering schools have had some success partnering with them to offer data science courses, but students struggle to meet pre-requisites or get seats in those courses. Others have designed their own sets of courses using "unicorn" faculty or part-time instructors who work in public technology positions. These programs largely still are making the case to their faculty that data science skills need to be taught in public affairs despite classes having enough students creating demand for them (at least as much as any other concentration). Lastly, a few presentations noted the value of having adopted courses from other schools to test student interest before adopting an entire concentration/certificate.

Quite a bit of discussion about the data science life cycle and what components should make up the core of a data science curriculum in schools of public affairs circled back to the previous day's conversation about student aspirations and skillsets. While some MPA/MPP students might aspire to have highly technical positions, most will be directing or working in teams with those who have these skillsets. The directive then becomes to make MPA/MPP students fluent but not experts in the uses, systems, and skills needed to leverage data science for public value. A set of key terms and methods were developed that students should, at a minimum, be conversant in as they begin their careers. Programs also need to be designed so that students can continue their learning curves in data science after graduation (regardless of pathway), especially those who are underrepresented in data science.

Day 2 of the symposium ended with an ideation session during which small groups of faculty discussed what needs a data science consortium could address based on what curricula exists already. Although

the group noted that students attending top ranked programs in public affairs had opportunities to pursue data science knowledge and skills now, a vast majority of students who attend smaller, regional institutions do not. These students largely work in state and local governments outside of large urban areas. To the extent data-driven decision-making and data science applications benefit public service delivery for any sized government, there is a need to "democratize" these curricula and make courses accessible to smaller, lower-resourced MPA/MPP programs. The group agreed that a priority of the consortium should be curating and sharing existing courses with these programs, either through online courses or supporting faculty to teach a course that is already designed.

The final half day of the symposium focused on two guiding questions:
- How can we become stronger together through a consortium?
- What are concrete first steps we can take?

Most faculty attending the symposium teach a data science or methods course already. The discussion on the last day of the conference centered around how to make these courses accessible through the consortium for other schools of public affairs. A template course built in GitHub was demonstrated to the group by Jesse Lecy who led the project effort to standardize how content can be organized and displayed for future courses shared through the consortium. The format includes a ReadMe file that describes how to copy the course and change key elements for any new faculty who wishes to teach it (e.g., title, course number, syllabus language, etc.). The goal is for faculty from multiple institutions to teach it under the support of Dr. Lecy, then augment its content over time to keep it responsive to the changing needs of teaching data science (e.g., updated problem sets, new readings, etc.). Faculty agreed that the format was accessible and an ideal way to share additional courses that they hoped to prep for the consortium in the coming academic year.

In addition to the template course, a draft charter/MOU for the consortium was shared with symposium attendees. The charter identifies the governance structure of a consortium, its purpose, terms of membership and participation, and dates in effect until renewal. The draft is modeled after other consortiums in higher education that aim to share teaching resources to broaden the reach of faculty expertise without incurring extensive costs at any one institution. Faculty at the symposium thought the draft was a good step toward formalizing the efforts/goals of the group and agreed to share it with their department/college leadership to gage its feasibility.

The symposium ended with faculty committing to next steps during the 2022-23 academic year to continue the work discussed. All agreed that annual professional conferences in public affairs should include information sessions and meet-ups to promote the consortium and what it can offer to programs that want to introduce or deepen their data science curricula. Several faculty also agreed to share their syllabi and collect course materials from other faculty in their programs to distribute through the consortium. A handful offered to build the course(s) they teach in GitHub using the template designed for the consortium or potentially teach one of the existing courses under the advisement of the course designer.