Technology, Data, and Infrastructure for Teaching, Research, and Practice

Prepared by:
Derek Brown
Brett Drake
Patrick Fowler
Susan Fowler
Jenine Harris
Kimberly Johnson
Douglas Luke
Paige Riegel
Ashby Tyler

Background

Introduction

A 2018 article in Forbes suggested that humans are currently producing 2.5 quintillion bytes of data each day and the pace of data production globally is accelerating.[1] This pace of data production is unparalleled in history and includes information that is critically important for generating knowledge that can be used to improve the health and well-being of individuals and populations. In addition, new technologies needed to harness this data for social good continue to develop at an accelerating pace.[2] The Brown School has an opportunity to build on our leadership in social work and public health by capitalizing on this new environment through improvements in our Technology, Data, and Infrastructure for teaching, research, and practice. In addition, improvements to our administrative data infrastructure are critically needed to promote efficiency of Brown School operations. To this end, we have three recommendations and suggested options and themes for each one:

1. Make a strong and continuous commitment to maintaining a modern data and computing infrastructure;
2. Develop and support a new administrative strategy to guide current efforts and develop a technology and data vision; and
3. Strengthen existing and develop new technology-related educational programs for students, staff, and faculty.

The urgency implied in the above recommendations is real. In the future, who will dominate the "data space"? Will public data be managed and interpreted by proprietary agencies driven by profit or by university-state partnerships committed to sound science and the public good? Will social work and public health workers be competent to utilize the data which increasingly inform their practice and policies? As evidence-based practice and policy are increasingly driven by live data, will we have the capacity to keep up? Will new tools such as machine learning and data visualization be something we master here, or something our students will need to find "somewhere else"? Will very nature of what we do (e.g. classroom education) transform around us and will any such wave be one which we ride, or one which falls upon us?

The current environment at the Brown School

The current status of the administrative support and capacity, capacity of students, staff, and faculty, and computing infrastructure for accessing and use data and technology at the Brown School are varied. In addition, there are financial-related data operation inefficiencies that impact our ability to achieve our mission that must be addressed (see addendum) if we are to fully achieve the goal of maintaining a modern computing infrastructure. Each is described below.

Administrative support and capacity. Institutionally, the Brown School is just beginning to recognize and commit to the new landscape of data and technology. An example of this is Dean McKay's full support of the University's application for a Federal Statistical Data Research Center. The StatLab is another successful example of administrative support for technology and data use at the Brown School. However, as *an institution*, the Brown School is behind in adapting to the shifting technological landscape including both big data and new technological tools. Washington University is ahead of the Brown School in adapting to big data, with new programs in computer science and new initiatives at the library being representative preliminary steps.

Capacity of students, staff, and faculty. Some individual faculty members have become national leaders in new technology, use of large datasets, and systems science. There have been several ad-hoc partnerships between Brown School faculty and students and their counterparts in Computer Science and Engineering (CSE). Several faculty members were involved in the conceptualization and development of the

[1] https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#6ee703e860ba

[2] https://www.amazon.com/Singularity-Near-Humans-Transcend-Biology/dp/0143037889/ref=as_li_ss_tl?_encoding=UTF8&me=&linkCode=sl1&tag=suhub05-20&linkId=1bfe28af1694717f40684a312971d87f

new data science doctoral program housed in the Division of Computational & Data Sciences (DCDS). All these efforts and many others have contributed to the technological expertise and potential within the Brown School; however, these isolated skills and temporary partnerships have not (so far) been part of any sustained or coordinated effort and therefore are limited in building school-wide capacity in technology. Because of these limitations we are not currently able to train students with sufficient skills to obtain jobs that emphasize data science knowledge and competencies.

Computing infrastructure. The two main infrastructure issues across teaching and research are computing power and software licensing. Both computing capacity of hardware and the correct and current licensing of software are necessary components for teaching quantitative and qualitative data analyses and for conducting research with quantitative and qualitative data sources. The students are well-aware of these limitations within the Brown School; computing power and software licenses are a frequent frustration for students and tutors in the StatLab where computers are old and slow with often out-of-date software licenses.

The current environment at WUSTL

There are numerous entities on campus that house technology resources including:

- Washington University Information Technology
- Office of the Chief Information Officer (CIO)
- IT Research Infrastructure Services Program
- Student Technology Services
- University Library Services
- Information Security
- Institute for Informatics (I^2)

While these resources may be useful to fill some of the technology gaps at the Brown School, it was the experience of the committee that these resources are not currently well understood nor utilized. Moreover, there is little to no participation by Brown School administration and staff on these projects or with these groups. For example, the office of the CIO has a Research Storage Services program available to faculty[3] with a 5 terabyte capacity (i.e., a lot of data!) and the university library services holds regular technology workshops and events open to the entire campus community.[4] If understood and utilized, these resources could potentially improve the technological capacity of Brown School students, staff, and faculty.

External Environment

All aspects of our society are changing fundamentally in response to improvements in data and technology. Business and medicine are applying advanced methods such as data mining and machine learning as basic tools.[5] Public agencies are using similar methods in building systems and policies to help their clients (e.g., screening in public child welfare, using machine learning to monitor food safety[6]). Many of our peer schools are adapting and have engaged in efforts, including (1) investing heavily in expanding capacities for training (e.g., USC[7]/ASU[8] online learning, John's Hopkins Coursera partnership[9]), (2) developing technologies for practice,[10] (3) supporting faculty scholarship in statistics, data sciences, and methods development, (4) supporting research by making data easily available to faculty, students, staff, and the community (e.g.,

[3] https://cio.wustl.edu/projects/iris-program/research-storage/
[4] https://library.wustl.edu/about/blogs-and-events/
[5] https://jamanetwork.com/journals/jama/fullarticle/2748179?guestAccessKey=8cef0271-616d-4e8e-852a-0fddaa0e5101&utm_source=silverchair&utm_medium=email&utm_campaign=article_alert-jama&utm_content=etoc&utm_term=080819
[6] https://www.stlouis-mo.gov/government/departments/health/environmental-health/food-control/food-safety-stl.cfm
[7] https://msw.usc.edu/
[8] https://asuonline.asu.edu/online-degree-programs/graduate/master-social-work-social-work/
[9] https://www.coursera.org/courses?query=johns%20hopkins%20data%20science
[10] https://www.healthmap.org/flutrends/

University of Michigan,[11] Penn State University,[12] Rutgers[13]), (5) developing expertise in data visualization for policy change, (6) engaging globally in cross culture research, training, service and (7) using new data and technologies for participatory research. Many gaps also exist across our peer institutions and may therefore be areas to consider when setting our **technological vision** at the Brown School. For example, few social work and public health schools have transdisciplinary research capacities, few are focusing on open science, and few are focusing on systems science. All social work and public health schools and programs will continue to grapple with digitization, and issues of fairness, transparency, and community facing research.

Strategic Themes

The committee suggests five major crosscutting themes. Each of these themes applies to the "Options" suggested below the themes.

Establishing a data-aware and data-supportive environment and culture. The school must make use of emerging data and technologies a primary consideration at all levels and in all decisions, including hiring, education, research, budget, physical resources and other areas. This culture must apply to research, teaching, staff and community engagement. A new understanding of basic ethical principles as they apply to big data is an overdue necessity.

Transparency/Visibility of data. This applies to our research (e.g. engaging in open science, posting analytic programming and allowable data), our policies (e.g. not firewalling any content which need not be firewalled), educational and community engagement efforts (e.g. posting useful content, guides or data with an external face usable by students and agencies), and financial accounting systems to promote efficient business operations.

Collaboration both internally externally to optimize efficient use of resources. Collaboration should be established in the following order: Within (1) The Brown School, (2) Washington University (e.g. computer and data sciences), (3) Partner Agencies or Universities, (4) Other Existing Resources (e.g. national or UN resources). Examples might include linkages to WU library services or ICPSR or national data clearinghouses).

Social Impact and Social Justice are core missions of the school. Both are applicable across the domains suggested below. Two specific social justice goals have been identified: (1) the representation of women, LGBTQ+, black/African-American, and Hispanic people in data science is extremely low; recruiting and supporting diverse students, staff, and faculty in data science could contribute to the diversity in the field and set us apart from other institutions; (2) adopting an *open science* culture where we share our data, code, and publications publicly as much as possible would allow researchers and practitioners in lower resource communities like parts of the global south and rural and inner-city US communities to access our work and use it to improve health and well-being. In general, we must be sensitive to the full range of social justice implications in the information age (e.g. service needs of displaced workers, better communication opportunities, situational awareness tools…)

Infusion is perhaps the least tangible but most necessary crosscutting theme. Data access, procurement, management and use must be covered in all aspects of Brown School efforts at a level similar to that evidenced in our commitment to diversity and social justice. Researchers, students, staff and agencies cannot function optimally without data literacy and use. For example, recent mass shootings should be covered in class with live access to underlying data (e.g. UCR homicide tables). As another example, there is a qualitative difference between discussing "the Delmar Divide" in class and showing it visually using Census Data (which takes, literally, seconds if you are data literate). As an institution, it is our responsibility to equip students with a data literate toolkit.

[11] https://www.icpsr.umich.edu/index.html
[12] http://www.datacommons.psu.edu/
[13] https://rucore.libraries.rutgers.edu/

Options

We make the following recommendations for advancing data and technology at the Brown School. These recommendations are presented in three domains with a set of options for each:

DOMAIN 1: Create and support a new *administrative* strategy to guide current efforts and develop a technology and data vision. Big data and sophisticated analytic processes have changed the world outside our walls. They need to change inside our walls too. We suggest the following:

- **A Brown School Data Center** with the goal of helping (1) researchers, (2) students, (3) staff and (4) agencies to (A) understand availability of data (including live data), (B) understand the potential methods by which data can be employed, (C) procure data, (D) manage data and (E) apply data. This might include (I) an online site facilitating use of data and technology, (II) a "bullpen" of skilled data locators/managers/analysts, (III) other functions such as collaboration, presentations and training. This center would **not** follow the lead of the Evaluation center in being largely outward focused and externally financially supported, although grants in the school could provide some support (e.g. contracting for database management or linkage to tricky external public sources).
- **DataLab** – Expand the current StatLab to include more data-related resources and services. A DataLab would provide additional training through seminars and individual support beyond course work to help students identify, access, manage, and use data in a methodologically sound and ethical manner.
- **Enhanced School Resources.** These may or may not directly be part of the Data Center. For example, while the StatLab has been a tremendous success story, we have no similar "DataLab" helping students identify, access, manage and use data in a methodologically sound manner. These tasks can be quite daunting and specialized. For example, many publicly available datasets can be very difficult to load into statistical programs. We will need to decide how many resources at the school could be "outward facing" as consistent with our school's community support mission.
- **Associate Dean level position** to head data and technology efforts in support of the Dean. This person must be able to integrate research, data, practice and teaching, as their primary goal will be to oversee a system which works for researchers, students, staff, and community partners. Key responsibilities would be areas such as visioning, prioritizing, resource allocation, short and long-term planning, gap analysis, collaboration, monitoring data and technology issues in social justice, inclusion and infusion, etc. Tenure track or Professor or Practice faculty lines may be appropriate to this position, which could have intensive responsibilities for several years. Perhaps high-profile events such as "Hackathons" focused on social issues could be coordinated.

DOMAIN 2: Strengthen existing and develop new technology-related *educational programs* for students, staff, faculty, and community partners.

We currently lack advanced (or sometimes, even basic) programming in a range of key areas (e.g. machine learning, data visualization, data mining, advanced computer-assisted interviewing…)

- For our educational programs, this would involve reevaluating all programs (MPH/MSP/MSW/doctoral) to assure **infusion** of data and technology content, support in filling gaps and/or integrating new content. It is also suggested that a data science specialization might be considered at the master's level, although overlap with current specializations (e.g. the research specialization on the MSW side and the epidemiology and biostatistics specialization on the MPH side) would have to be evaluated. Students would benefit from clear specifications for their laptops before the come here, from "genius bar" resources to support use of their personal computers, and perhaps through standardized free or low cost software packages.
- For faculty, post-docs and doctoral students, increased resources should be applied to promote advanced data and analytic skills (e.g. conferences, invited experts). Hiring priorities should be shifted to emphasize on onboarding tenure track, research and practice professors with a data focus

being desirable. At the staff level, those who directly support the use of data would benefit from continuing education that would cover basic principles of data literacy, open data, and open science.

- At the community level, an external face to our efforts (perhaps through the School Data Center) should be made available supporting outside use of data. For example, a user-friendly (visual or tabular) Zip code-based data file reflecting crime, census, health, environmental (e.g. food desert) and other data could be established to support agency work.

DOMAIN 3: Make a strong and continuous commitment to maintaining a *modern computing infrastructure*. The school must make a fundamental commitment and *coherent* (see administrative suggestion above) plan to support high functioning technology in the school. Improvements must be made in (1) hardware, (2) networking, (3) software and a range of other areas (e.g. availability of particular systems to satisfy specific grant requirements). We need a responsive and flexible infrastructure that keeps pace with rapidly changing technologies. Acquisition and fundamental support of necessary resources to support a flexible infrastructure that meets the data processing needs of researchers, students and agencies requires staff capacity in, but not limited to, Information Technology, Human Resources, and Facilities. A range of specific issues must be addressed including transitioning from a "build and fix" to a "build and assure" IT paradigm, solving data problems including storage and the "skinny neck" (data movement speed) problem. Better collaboration between the school and the Washington University IRB is needed around data security.

Short- and Long-term Plans

Short Term (beginning Fall 2019):

1. Execute a current and future needs/gap analysis specifically in relation to technology and data infrastructure.
2. Run an environmental scan on currently available technology and data related resources and services as well as departments and schools with whom we can partner and share such services and resources.

Mid-term (beginning Spring 2020):

1. Expand the StatLab into a DataLab, combining broader data expertise with statistical expertise.
2. Based on the previous steps (gap analysis and environmental scans) implement new curriculum and training opportunities for students, staff, and faculty.
3. Facilitate 2-5 continuing education sessions a year for faculty and staff. Faculty trainings include data related software, coding languages, and analytic strategies. Staff may access faculty trainings or will receive training in basic principles in data literacy, open data, open science, and resource awareness.

Long term (beginning Fall 2020):

1. Use information gathered from needs analysis, environmental scan, and DataLab experience to provide support and defining frameworks for a data science center and a potential Associate Dean position.

Estimated Costs

The world has already changed and beyond the efforts of individual faculty, the Brown School has not. The cost of not making an immediate, general and sincere effort to upgrade our use of emerging data and technology will be the future irrelevance of our institution. Compared to this, the costs associated with the short and long-term plans listed above are minor. We anticipate the key costs will be:

1. Human, informational, and financial resources needed for an environmental scan of existing technological resources and expertise;
2. Resources to facilitate and sustain the expansion of the StatLab into a DataLab to support students;
3. A fund available for faculty publishing manuscripts that are not grant-supported as open-access in all venues where this is available;
4. Funding to support ongoing on-site or off-site training of staff and faculty on new technological tools and data resources;
5. Funding to support an Associate Dean position;
6. Human, informational, and financial resources to support and sustain expanded infrastructure capacity including staff, hardware, and software.

Addendum

The committee discussed impacts of our financial tracking and accounting systems. The committee felt that although an argument can be made that this area of operations falls under the scope of data and technology because it impacts our ability to efficiently conduct our core mission activities of research, teaching, and service, we ultimately decided to include it as addendum with the recommendation that a separate committee be convened to address this topic. Specific examples of areas in critical need of updating that were discussed are:

- Systems for electronic tracking and reimbursement for business expenses
- Transparency in financial health of the school
- Systems for invoicing and timely payment of contracts