Data&Society

Exercise 2: Data Clinic Model

Goal:	To design a Data Clinic that helps students store, protect, clean and prepare data.
Background:	Many universities have "Statistics Clinics." While the structures vary, they tend to allow students and faculty to drop in (or schedule a meeting) with someone with advanced statistics knowledge who can help them address their statistics problems. Often, faculty and students play a role in the statistics clinics, but there are also models where full-time statisticians (often master's or PhD level) are hired to support researchers. The funding model for these clinics varies. Some are paid for centrally; others charge a fee to departments. In most cases, those using the clinic can do so free of charge, but some institutions charge for extensive use. (Often, those who are going to need statistics help add this as a line item in their grants.)
Ideas:	One idea that has emerged in our research is the notion of a "Data Clinic." Often, this is an extension of the Statistics Clinic, since many of them are already helping students and faculty with some data-related issues. Yet, there are other issues that people are encountering that could benefit from a data-centric perspective. Examples include: • How to store, secure, encrypt, de-identify, and otherwise protect data. • The trade-offs of using different private vendors to store or process data. • How to clean data and assess its limitations and biases. • How to prepare data to make it share-able without violating privacy. Stanford's "Data Science Drop-In" describes its consulting services this way: *Trouble wrangling your data? We're here to help with all aspects of data collection, cleaning, analysis, and visualization. During our open consulting hours, we provide assistance with any and all data science problems, including: APIs and web crawling; parsing unstructured data; online experiments; data storage and querying; distributed computing and*

continued	*scalable algorithms; large-scale regression and classification; natural language processing; and visual exploration and communication. We also host mini-tutorials on various topics throughout the academic year. Even if you don't have a specific technical question, feel free to stop by our open office hours to chat. We'd love to hear what you're working on, and we're happy to help you brainstorm about your project.*
Instructions:	In small groups, discuss the questions below that consider the audience served, services offered, and support required to set up a Data Clinic at your organization. Use a large sticky note or whiteboard to list the main features of your Data Clinic. Consider personnel, equipment and services.
Exercise Questions:	• Who would the Data Clinic support? Students? Faculty? What disciplines? What would be needed to support computer science faculty vs. humanities students? • What range of problems could the Data Clinic help with? Imagine technical, logistical, and security-related support services. *If your organization has a Stats Clinic, where would there be overlaps?* • Who would staff the Data Clinic? What is the role of faculty, students, and support services? What kind of hires would be needed? • What kind of funding support would be needed for this to work at your organization? Who would pay for the services? What are the economic hurdles for this? • What kind of organizational support would be needed to make this work? Bottom up versus top town? Who would need to buy-in? What are the organizational barriers to making this happen? • What kind of technical and structural support would be needed? (e.g. secure data storage, centralized company contracts) Who currently provides those services? • What would this model not help with? • What will get in the way of a Data Clinic working at your organization?