Research Seminar :Extra Credit Paper

As a proof of participation, you should write a half-a-page summary. It should contain
a) name, title and time of the talk
b)The name of the seminar series if it belongs to any
c) what did the authors do and what did they find and
d) your own reflections (did you understand? was it interesting? Was it related to the class topics? Was it relevant in your opinion? ...)

Data Then and Now : Stephen Molldrem - HIV, Meet Data Science: On the Arrival of Big Data Methods, the Open Source Ethos, and Predictive Analytics in U.S. HIV Prevention.

The Data Then and Now seminar series explores the social and organizational history of data and data practices in order to better understand the current data-intensive moment through its antecedents and continuities.

Following confirmation that antiretroviral treatment for HIV is a highly effective way to prevent transmission, the U.S. HIV strategy has reoriented around a “treatment as prevention” paradigm. The rationales supporting treatment as prevention are widely grasped among stakeholders. However, the infrastructural and epistemological transformations that have facilitated the implementation of programs to support treatment as prevention are poorly understood. This talk describes the arrival of approaches from data science, and from the “open data” and “open source” movements, to HIV prevention in three cases.
The first case focuses on how HIV data, formerly a class of “exceptionally” sensitive information with many restrictions on its exchange, were re-articulated during the 2010s. Mandates from CDC for state departments of public health to collect and use routine HIV care data in prevention have made HIV doubly-exceptional: exceptionally sensitive, but also exceptionally important to exchange, share, and use in prevention. Programs in this area often involve partnerships between health departments and correctional agencies, a concern in jurisdictions where HIV nondisclosure is criminalized.

The second case considers open source tools for conducting HIV research and prevention. Software packages like HyPhy, PHYLOSCANNER, and HIV-TRACE are maintained by researchers on GitHub. They facilitate new forms of “cluster tracing” to identify people living with HIV in “transmission risk networks.”

The third case describes a project in South Carolina which will use “[electronic health records], claims and data from private institutions, housing, prisons, mental health, Medicare, Medicaid, State Health Plan and the department of health and human services” in the hope of developing

“machine learning” and “predictive model development” tools for use in prevention. Public health re-uses of HIV data are done without consent. I close by considering bioethical issues, recommending the development of affordances for people living with HIV to assert controls over some re-uses of their data.

The speaker also discusses how before 2013, this wide database of HIV was only used for prediction and logistical purposes such as tracking new cases, people living with HIV moving to another state or cases that had worsened over time or progression of HIV to AIDS. Post 2013 it was seen that the HIV surveillance system was collecting real time data of any medical care or assistance received by a patient to get the latest data into the system. This data was then used to study trends, collect data about when virals occured. This information was then an input for prediction and prevention using data analysis and machine learning models that help in better management of HIV.

The data to care program allowed data to be sent out of surveillance systems to centers for prevention that could then reach out to people in the program via phone calls or mail to update them on any changes in their medical care. Previously in contrast, this HIV surveillance system only received data that was used for epidemiology and analysis. This re-use of data has been a major breakthrough in dealing with HIV and providing care since the epidemic in the 1980s. This re-use of data has set up a feedback loop that has helped set up a better management and care of patients.

The explanation of the data to care program’s emphasis on data collection and cleaning was something I found extremely interesting. It was also very interesting to learn about how the feedback loop in a data system that deals with such sensitive data such as healthcare data is maintained. The topic was definitely relevant, related to class topics such as the ones learnt during the beginning of the quarter - data analysis, data cleaning, exploratory data analysis. Data Science and machine learning in health care is a very emerging field, the power of data when fed into machine learning models that train on complex algorithms, coupled with medicine can provide major breakthroughs in the healthcare system in the country. It was an interesting talk and I learned a lot about the different applications of our classroom learning in a very different domain than what we generally talk about.