Workshop Title	Coleridge Initiative: A Model for Training Data Scientists to Use Longitudinal Data for the Public Sector
Workshop Length	Half Day
	Universities are increasingly working with public data in structures to improve the use of data in decision making by government (Hawley, 2019). In recent decades, government has focused heavily on improving new technical systems. There are two significant barriers to transforming the government's use of data. The first is a lack of workforce capacity. Government employees have the requisite skill. The second is that many policy problems require analyzing data that cross agency lines, and confidentiality rules prohibit data-sharing (Reamer, Lane, Foster, & Ellwood, 2018). We argue it is possible to create a virtuous cycle of change within agencies whereby employees can share data, share knowledge, and apply modern technology to transform the way problems are solved across agency and jurisdictional lines. To accelerate the development of the current workforce new executive education style courses for public sector employees are also needed. Our half day workshop will provide a hands on introduction to the Coleridge Initiative (coleridgeinitiative.org/). Over the last five years we have developed a training program which demonstrates how a data science program for public policy might be developed at scale to meet the quickly rising demand. The interest and investment into the program has been overwhelming (300 +/- participants). The key elements of the program include (1) access to confidential agency microdata in a secure computing environment, (2) developing new skills, and (3) creating products that have value for government agencies. Participants work in a secure cloud-based environment and are introduced to new ways of collecting data and making use of new methods and tools. The overarching approach includes training government staff in how to keep fundamental statistical concepts like population frames, sampling, and valid inference but expand their skillset to include modern computational data analysis tools as well as using new types of data.