Suggested Projects: White House Innovation Fellows
 # 11 - Big Data: Free Supercomputer Time and Analysis Software
for Discovery

In the Big Data era, the federal government should make supercomputer time and analysis software available, online, without charge when the purpose is for education or discovery! An Innovation Fellow team can work with HHS, FDA, NSF and other agencies to develop initial projects testing this new model to accelerate public + private sector cooperation and scientific discovery.

The practice of charging for discovery-oriented computer time arose decades ago - many iterations of Moore's Law! The true scarcity in science (and startup companies) is time. Why - given the extraordinary investments of years of hard work to develop their competence - should scientists spend substantial amounts of their time writing grant applications to get access to Big Data resources? These datasets should be available online, 24x7, for people worldwide to explore and test hypotheses at the speed of thought.

I have spoken with Jim Short, a Big Data/Health specialist at UCSD Supercomputing Center about free, rapid learning, global analysis of the Kaiser/NIH./UCSF Everything Included (N=500,000) Biobank. There is a basis for further discussions about the services that they could provide. One key step probably is a sub-set of Reference Databases (e.g., N=10,000, but to be determined) that are large enough for education and initial discovery of many effects. For confidentiality, data should remain on government systems and a core set of analysis software can be available. The planning team can draw lessons from the NIH-Wellcome Trust-BGI (Shenzhen) "1000 Genome" experiment on the Amazon Cloud and the Galaxy open source software suite. UCSD Reference Databases can be archived and cited in scientific publications.

Promising ideas can have access to larger samples upon application, either free or at charges determined by the government.

This national system economizes on the time of scientists, who do not need to review initial applications for discovery-oriented work.

Lloyd S. Etheredge - June 11, 2012