

**Silicon Valley Regional Data Trust:**  
Building a Brighter Future Through Data Science for Social Good

**The Need**

Our nation is threatened by the critical problem of income inequality, which is reflected in a Silicon Valley paradox: dynamic income growth for some, deeply persistent poverty for others. Education can play a key role in solving this and related social problems. However, efforts to reimagine education have been hampered by the absence of a foundation of innovation that is central to Silicon Valley's success: information technology and data science.

Fields ranging from consumer services to cancer treatment employ data science in research contributing to dramatic advances. However, educational research has not yielded conclusive answers to either fundamental educational questions or to numerous specific questions on which answers to fundamental questions can be built. A barrier to advancing research in education and related fields has been the absence of sufficiently comprehensive datasets and data infrastructures that support cutting-edge research.

**The Silicon Valley Regional Data Trust**

To fill this gap, the University of California, Santa Cruz is partnering with public schools and health and human service agencies in the region and with providers of online educational programs to form the Silicon Valley Regional Data Trust (SVRDT). SVRDT is built on a foundation of trust among stakeholders who contribute data: public schools and agencies, providers of on-line educational services and the children, families and communities they serve. The Trust will be committed to insuring that data, a public resource, will be used to serve the public good. SVRDT will provide a model of a secure data sharing environment, which supports research that informs policy, improves service effectiveness and efficiency, enhances interagency collaboration and consequently improves outcomes for children, families, and communities..

A secure data-sharing environment and analytic platform will lie at SVRDT's core. Educational datasets tend to lack necessary breadth and depth. While 70% of variance in student achievement is attributable to out-of-school factors, datasets typically are confined to in-school data. And, although students at greatest risk for poor academic performance often move from school-to-school, datasets generally are limited to a single school or district. To provide data of necessary breadth, SVRDT's data-sharing environment will combine data from public schools and health and human service agencies—including social services, public health, mental health, and juvenile justice/probation—in San Mateo, Santa Clara, and Santa Cruz Counties. Datasets generally lack depth by including data about classroom and school conditions or data about instructional activity but not both. SVRDT's secure data-sharing environment will provide depth by linking administrative data to data on teaching and learning provided by organizations that on-line educational services and resources to schools in the three counties.

- **Secure Data Sharing System:** Constructing such a comprehensive data system is

extremely challenging and requires the development of a data management infrastructure which is able to 1) integrate heterogeneous information from different schools, health and service agencies; 2) provide timely access to data by schools, agencies, and researchers; 3) provide security and privacy guarantees. The Santa Clara County Office of Education is providing the technical infrastructure through its Data Zone.

- **Data Analytics:** Developing an analytic platform that supports access and analysis to schools and health and human service agencies and to collaborating researchers. Educational research has been hindered by the lack of data infrastructure that enables researchers to bring to apply advances in data science. Faculty from UCSC's Division of Social Sciences and Baskin School of Engineering—in collaboration with educational and health and human service agencies—will build an analytic infrastructure to support cutting-edge research that informs practice and policy.
- **Data Ethics:** The types of data and inferences that are possible with this data have the potential to have great positive benefit, but there is also the possibility for misuse. A key component of this effort will be to also develop the policy and administrative structures to make informed decisions about how the data may be used, understanding the implications of using the data, and educating end-users in the short-comings and potential limitations of the algorithmic approach to education analytics. We believe by integrating this from the very beginning, our approach is likely to provide the most value.

### **Progress To Date**

Supported by a planning grant from the National Science Foundation, SVRDT has: a) garnered support of policy leaders, leaders of education and health and human service agencies and other stakeholders to establish Data Trust; b) hosted convening of representatives of county offices of education and health and human service agencies in the three counties, beginning the process for establishing inter-agency, data-sharing agreements; c) Santa Clara County Office of Education (SCCOE) built data center and warehouse, which will provide technical infrastructure for SVRDT that integrates school and agency data in a secure data-sharing environment; d) UCSC and the SCCOE signed a Memorandum of Understanding to formalize the partnership to build and operate SVRDT; e) Santa Clara, San Mateo and Santa Cruz County Boards of Education adopted resolutions supporting SVRDT; f) Santa Clara and Santa Cruz County Boards of Supervisors adopted resolutions supporting SVRDT; San Mateo will adopt a similar resolution in January, 2016; g) Assembly Member Mark Stone sponsored a joint resolution designating SVRDT as a pilot (it is in committee); h) hosted 2nd convening to organize five Work Groups and initiate the work of legal counsels from the three counties to provide input to proposed legislation; h) Established relationship with non-profit, which offers a platform providing teachers with no-cost access to online educational services, to link instructional data to SVRDT database.

### **Prototype Build Out**

To build the Data Trust will involve a) establishing the legal foundation; b) building the

core data repository, c) developing the database and analytic platforms, and d) creating the data ethics framework:

#### Establish Legal Foundation:

- Develop and approve interagency data-sharing and other legal agreements with support of a consultant from Stewards of Change, who has extensive experience and expertise in developing data-sharing agreements nationally.
- Complete memorandums of understanding, formalizing relationship of UCSC, county offices of education and health and human service agencies in San Mateo, Santa Clara and Santa Cruz Counties

#### Build Secure Data-Sharing Environment:

- Develop connectors to the Santa Clara County Office of Education (SCCOE) data warehouse to import student information system and assessment data from two school districts in Santa Clara, San Mateo County and Santa Cruz Counties.
- Develop connectors to SCCOE data warehouse to import administrative data from health and human service agencies in San Mateo, Santa Clara, and Santa Cruz Counties.
- Develop connectors to SCCOE data warehouse to import data from educational technology providers.

#### Develop Database and Analytic Platforms:

- Develop database infrastructure to address issues of database integration, data storage technology, security, and privacy with the support of a consultant from the IJIS Institute, who has extensive experience and expertise in designing secure data-sharing environments.
- Develop analytic infrastructure to answer questions that inform policy and practice

### **Summary**

Fields ranging from consumer services to cancer treatment employ data science to make dramatic advances. However, educational and equity research has not yielded answers to fundamental questions, such as the following: What causes academic achievement of youth in the U.S.? What contributes to the achievement gap between students from low-income and more affluent communities? Nor has research offered clear answers to specific questions, on which answers to fundamental questions can be built.

SVRDT will enable teams of education and health and human service practitioners, representatives of community-based organizations and university data scientists to address many critical issues, including but not limited to: a) identifying children and families who will face challenges and suffer negative outcomes, b) assessing the effectiveness of organizations, systems, programs, interventions, and practices and c) determining the particular mix of educational, health and social services and resources that will best support each child and family. Where better than Silicon Valley to marshal the power of data and information technology to address social and educational issues?