

IDS CASE STUDY: Silicon Valley Regional Data Trust

Supporting Students through Integrated Data
and Research-Practice Partnerships

by Rosalyn Allison-Jacobs

DECEMBER 2018



This case study is an expansion of an overview of the Silicon Valley Regional Data Trust (SVRDT) authored by the SVRDT Core Team and supplemented with information from interviews with the following SVRDT leaders: Marcy Lauck, Director; Bill Erlendson, Ph.D., Director; Richard Gold, Legal Counsel to SVRDT; Rodney Ogawa, Ph.D., Director; and Patricia Ponzini, Director.

❖ **Background**

The Silicon Valley Regional Data Trust was established as an initiative of the Santa Clara County Office of Education in partnership with the University of California, Santa Cruz and the three counties that comprise Silicon Valley—Santa Clara, San Mateo, and Santa Cruz. It is a unique initiative, distinguished by the lead role of public education, an intentional focus on trust-building and planned change, and cross-county data sharing among public schools, health and human services organizations, and juvenile justice systems.

The foundation on which the Silicon Valley Regional Data Trust (SVRDT) was established harkens back to 2001 when the San Jose Unified School District built what was, at the time, one of the only K-12 education data warehouse in California, and one of the first K-12 data warehouses in the United States. With challenge grants from the Annenberg and Hewlett Foundations, the district—under the guidance of Marcy Lauck, then Manager of Continuous Improvement Programs with the San Jose Unified School District and Bill Erlendson, Ph.D., then Assistant Superintendent for Educational Accountability and Community Development for the San Jose Unified School District—tapped into the community of stakeholder’s needs and interests (including parents, teachers, service providers) and Silicon Valley’s flourishing business intelligence brain trust to figure out how to use data to close the academic achievement gap between white and Latinx students.¹

The strictly intra-district database was used to determine the impact of specific programs and educational initiatives, ultimately enabling the district to employ data-driven decision-making on student-specific bases, and personalized learning to ensure that each child had the supports needed to be successful in school. The work was informed by a methodical “bottom up and top down” community engagement and exploration process led by Erlendson and Lauck, which incorporated massive multi-modal data gathering including a climate survey, focus groups, and community conversations that surveyed over 28,000 students and 24,000 parents, and engaged 6,000 participants. By 2012, the database contained 16 years’ worth of data on students’ K-12 educational history, 60 million records, and 3100 data elements. San Jose Unified School District was one of only a handful of districts in the state to make significant measurable progress towards addressing its racial achievement gap, closing the gap by 36 percent. Central to the success of the initiative was the process of building a culture of data literacy and shifting perceptions about the constructive use of data among educators and administrators across the district.

In 2011, Lauck and Erlendson, by then retired, were joined in exploratory conversations about a regional, systemic expansion of the database by Rodney Ogawa, Ph.D., Director of the Center for Collaborative Research for an Equitable California at University of California, Santa Cruz (UCSC). Ogawa brought over 40 years of nationally renowned research in educational reform for historically underserved and under-resourced students. The three initial planners were later joined by Patricia Ponzini, then Chief Administrative Officer for a limited liability company whose members included UC Santa Cruz and the Foothill College Community College District. Ponzini brought a wealth of cross-sector experience in Silicon Valley and acumen in successful navigation and bridge-building—both internal and external—to organizational bureaucracies. Ogawa and Ponzini, building on UCSC’s deep institutional believe in supporting research that contributes to social good, developed UCSC’s commitment to the project, which eventually became the Silicon Valley Regional Data Trust.

In 2013, under a planning grant from the National Science Foundation, Lauck, Erlendson, Ogawa, and Ponzini began building relationships with agency leaders and policy makers through a long series of face-to-face meetings across the three counties, sharing the framework for what would comprise a three-county data system. The product of these conversations formed the foundation for the “Trust Framework” described below. They also cultivated key sponsorship within the state legislature that later enabled the initiative to overcome a legal barrier to cross-county data sharing for the three counties that comprise Silicon Valley.

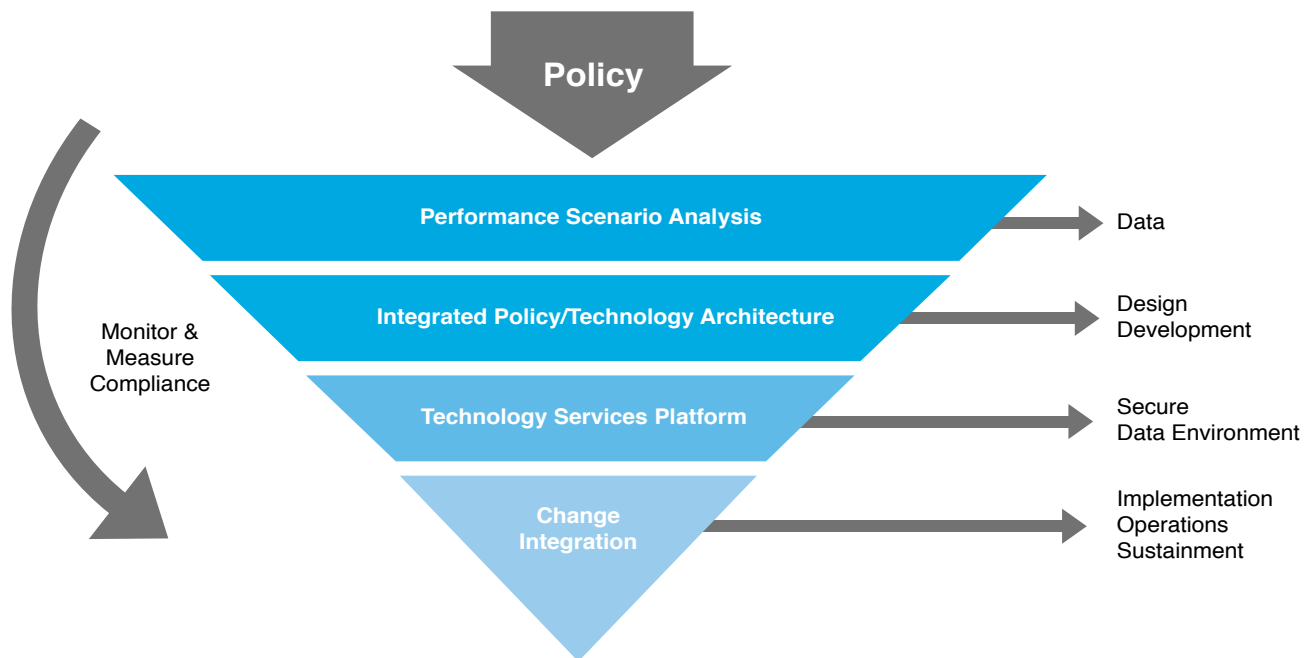
During the planning process, the team learned that the Santa Clara County Office of Education (SCCOE) was building a data warehouse, the “DataZone,” for its school districts. Discussions between the planning team and then Deputy Superintendent Mary Ann Dewan revealed that the DataZone had the capacity to serve public school districts in all three counties. Superintendents from San Mateo and Santa Cruz Counties concurred with the option of capitalizing on this resource, and it was decided that SCCOE and UCSC would enter into a partnership to develop the Silicon Valley Regional Data Trust–SVRDT.

SVRDT’s mission is to change the culture and practice of how data are responsibly and ethically used to develop actionable solutions to critical educational, health, and social problems. SVRDT provides a Secure Data Environment (SDE) that will enable K-12 public schools and county health and human service agencies to share data in order to coordinate case management, personalize and integrate services, inform public policy, and partner with UCSC faculty to conduct research in partnership with public schools and county agencies. Collectively, the partnerships across the three counties include over 60 school districts, separate Behavioral Health, Child Welfare, and Juvenile Probation agencies in each county, and three County Offices of Education that provide fiscal oversight and operational support to their respective municipal school districts.

❖ Trust Framework and Change Integration– SVRDT's Unique Foundational Components

In order to build and operate a multi-county, interagency data system, SVRDT has focused on building trust. This entailed a process of conducting many face-to-face meetings between the original four-member planning team, agency members, and policy makers across the three counties. The Trust Framework, as illustrated in Figure 1 below, uses an integrated policy and technology architecture designed to assure protections and appropriate use of data in all stages of design and development; it effectively institutionalizes the trust relationship across all stakeholders in SVRDT. The Policy layer codifies the rules for use of data in compliance with legal and regulatory requirements; the Architecture layer translates the requirements for use of data through policy compliant design; the Technology layer enables those rules for access to, and use of, data through the SVRDT portal; and the Change Integration layer guides the evolution of compliant, ongoing use and interaction with SVRDT partners.

FIGURE 1. The SVRDT Trust Framework



SVRDT’s change integration concept is an intentionally inclusive and participatory process. Stakeholders engage in values-infused equity conversations and decision making around selecting the right data to drive application and uses, formulate associated security and access decisions, and help design the communications and implementation processes with team members and constituents. Thoughtful change integration is viewed by the core team as a crucial component of building buy-in towards successful implementation of an IDS.

The SVRDT change integration model facilitates a culture shift in how service agencies can access and use real-time data to advance equity, ethics, privacy, and social justice. Change integration relies upon a shared understanding of the requirements necessary to support data governance, legal and ethical compliance, information technology, research, and practice, and is defined within an assessment rubric that outlines three levels of readiness essential for engagement, piloting, and demonstration.

SVRDT’s change integration methodology is grounded in the need to build consensual communities of interest dedicated to a “Do No Harm” philosophy that addresses issues of disproportionality and systemic inequities. The methodology highlights alignment of caseworkers, county human service and education agencies, supported by responsive state legislation.

There are three SVRDT change integration levels at which sponsorship is important: 1) executive (governance), 2) primary (agency leadership), and 3) direct sponsors (educators, caseworkers, researchers). The goal is to build coalitions within each level to ensure there are adequate resources to achieve and sustain change.

The key mechanisms through which systemic transformation can occur are legislation, data governance, legal, technology, and research. SVRDT's change management tools will be used to assess current policies and practices, evaluate three phases of client preparation, and identify changes to systems and processes in each organization to support fundamental changes in how work is accomplished. Finally, change integration performance metrics measure and reinforce change integration deliverables with on-site support and consulting that are aligned with the SVRDT Trust Framework.

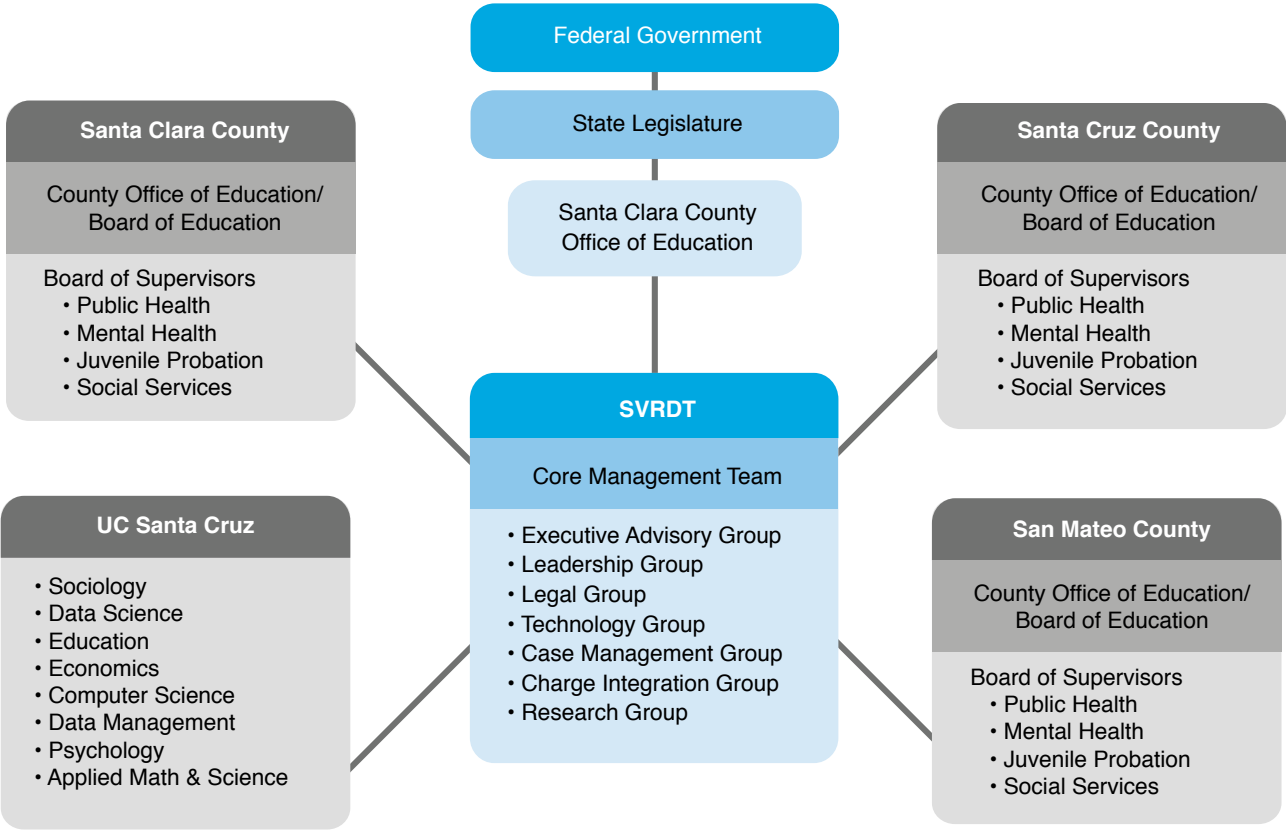
❖ SVRDT Governance Structure

While the Santa Clara County Office of Education (SCCOE) is officially the organizational home of SVRDT, the governance structure reflects parity between SCCOE and the other data trust stakeholders. The educational partners' leadership in launching the Data Trust makes SVRDT unique among Integrated Data Systems (IDS) sites; school systems are frequently among the last data-sharing partners to sign on due to privacy protection concerns.

In developing the SVRDT Trust Framework, the stakeholders used a Performance Scenario Methodology to assist agency representatives in identifying aspirations, challenges, and opportunities. This methodology is a unique approach based on the development of a client persona and real-world case scenarios that support data discovery among case managers and agency leaders, and has helped define the SVRDT data sets that guide the development of the Secure Data Environment. Most importantly, this process has established critical ownership among stakeholders and has created a shared vision among participating agencies that formed the basis of the SVRDT architecture and technology infrastructure.

The SVRDT governance structure comprises seven working groups: executive, administrative leadership, legal, technology, research, change integration, and case management. Working groups include representation from data-contributing stakeholder groups from each of the three counties: the three county offices of education, Behavioral Health, Juvenile Probation, Child Welfare, and UCSC researchers. The core management team includes eight individuals: legal architect, technology architect, Superintendent of SCOE, UCSC research partner, and the four founding directors. Together, they oversee the development and approval of all IDS core components. Figure 2 below summarizes SVRDT’s governance structure.

FIGURE 2. Silicon Valley Regional Data Organization Structure



Early on in the process, the working groups collaborated on identifying and building consensus around “up-front data,” that is, those data elements that all stakeholders wanted from each other and were already legally permissible to share, and so were least controversial. The working groups used a client persona to discuss possible applied uses of the Data Trust, and 30 different real-world case scenarios have been developed thus far to determine the legalities associated with data sharing for each. The Santa Clara County Office of Education, as part of the SVRDT Core Team, made the determination that, while each school district’s data would be partitioned with dashboards created for district use, the education data for the Data Trust could still be accessed in a central repository, while agency databases would remain separate. At the same time, the data repository, or “DataZone,” conforms the databases so that they provide standard definitions no matter the district, which enables regional, cross-county research.

❖ **Policy Architecture: Assuring Legal and Regulatory Compliance**

The legal basis of the SVRDT Secure Data Environment (SDE) is predicated on policy, trust, and ethics. All confidentiality and privacy rules and regulations, both federal and state, are honored and protected. Throughout the development process the legal architecture has guided the development of the technology architecture. Compliance is based upon policy codification through a series of authorization and consent instruments designed by appropriate officials (agency leadership; policy and case management practitioners; legal counsel from each county/agency). Furthermore, the leaders of the counties, as well as the leaders of the involved agencies, make the determination of who has access to data, and the SDE provides them with the ability to monitor that use to ensure that the rules and policies are being upheld.

The three counties and the individual agencies within the three counties participating in the SVRDT have agreed to a high-level, policy-based document entitled the Multi-Agency Agreement. The document sets forth the purpose of the SVRDT: the parties will work to ensure acceptable data use enabled by technology—whether it is for case management,

policy and program evaluation, or research purposes—and may only be used for positive purposes to improve the lives of and outcomes for the children and families served by the counties and the agencies. Data will be accessed to improve equity for persons served by the counties. All of the parties agree to implement processes to overcome institutional, structural, and individual racism that exist in our society; to enable the children, youth and families served to lead more prosperous and rewarding lives; and for the government to provide services that are the most effective possible.

Notably, the Multi-Agency Agreement has been intentionally drafted to avoid legal jargon. Rather, it reflects the shared values and goals that emerged from the inclusive planning process, and the “do no harm” intent underlying the systems goal of improving the lives of all children and families.

The second document that comprises the legal and regulatory basis for SVRDT is the Enterprise Memorandum of Understanding (E-MOU). In this document, the parties set forth what data will be shared, the role-based designation of who will be credentialed to access the data, the process for credentialing the users of the data, the process for identifying the clients served by the different systems and possibly by the different counties, and the technology criteria and processes. The document describes the process of accounting for, identifying, and correcting any breaches that may occur. The E-MOU is designed to explain how additional counties and additional service systems can participate in the SVRDT, and how this type of expansion can occur without negotiating a new data sharing agreement. When expansion occurs, the parties will sign an addendum, adding the new participating agencies and the new data elements that will be made available. This E-MOU format can be replicated in other jurisdictions and can ultimately become a national model for data sharing among different agencies and jurisdictions while utilizing different technological platforms.

The third document comprising the legal and regulatory foundation of the SVRDT is a universal consent. Currently, systems operate separately with clients’ or participants’ consents for data use in effect only within and not across systems of service. Since the

Data Trust is dealing with data regarding minors, normally the parent/guardian signs the consent and therefore the parent/guardian is the person referenced as “individual,” “client,” or “participant.” To date, an individual may have needed to sign numerous and duplicative consents in order for service providers and systems to access information on the same individual. Representatives from education, behavioral health, child welfare, and juvenile probation in the three Silicon Valley-based counties have agreed to work together to create one consent that will be applicable for all four systems. A client or participant will be able to sign one consent in any of the four systems and that consent, clearly written and easily understood, will apply if that individual is involved with more than one system. The individual will know what data are being shared and with whom, and it will be a self-determination by the client that permits the data to be utilized. The universal consent format will be able to be replicated elsewhere in California (when dealing with the same four systems) and in other jurisdictions (different federal and state laws must be reviewed and incorporated).

As the academic research partner, UCSC has instituted a global Internal Review Board (IRB) policy and process that covers SVRDT.

❖ **Technology Architecture: Building Secure Data Environments**

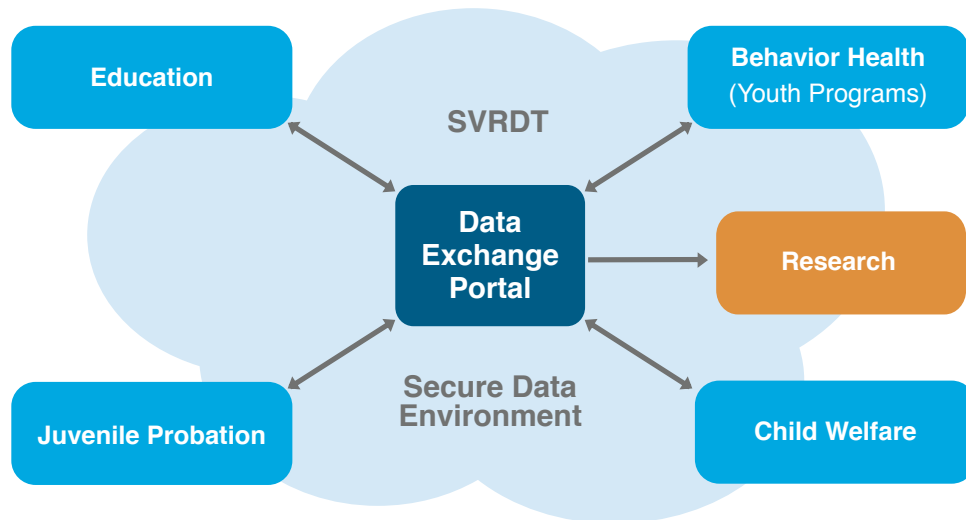
SVRDT has also designed a secure data environment (SDE) that could serve as a transportable model for other jurisdictions working to provide increased data access for approved uses. Improved access to cross-agency data provides streamlined coordination of case management, enables more personalized and integrated services, informs public policy, and supports the integration of research—all of interest to the SVRDT stakeholders and participating agencies.

The process through which the specifications for the technology architecture emerged was preceded and guided by the development of the legal and policy architecture. To honor the trust relationships so essential in a data sharing environment in which personally identifiable information (PII) is intentionally retained, the seven working groups painstakingly formulated multiple possible data-sharing scenarios utilizing a fictitious persona, and articulated the required security and access requirements associated with each scenario. On this basis, the technology architecture required to support this framework was designed.

The Santa Clara County Office of Education (SCCOE) initially established the DataZone (DZ) as an educational data warehouse intended to serve the County's school districts. In 2015, SCCOE offered the DZ as the repository for education data from all 3 counties involved in SVRDT, and a Memorandum of Understanding formalized the partnership between SCCOE and UCSC in August 2015 to jointly build and operate SVRDT. The SCCOE now hosts data from public schools in all 3 counties. Thus, the DZ serves as the anchor node in the SVRDT Secure Data Environment (SDE) and is a comprehensive data source, holding 25 domains of PK-12 student data. The DZ is now receiving data from 30 school districts on over 280,000 students (more than 60% of the three-county student population).

The SVRDT SDE will be deployed as a trusted application platform, initially connecting the agencies included, as depicted in the Figure 3 below. SVRDT application services will allow authorized and certified users the ability to query available case information that is resident in each agency's case management system. Capabilities to connect and access data will be available to these authorized individuals through a series of centrally hosted data messaging services, managed by the SVRDT Data Portal. The SVRDT Data Portal is the data exchange platform that will route requests for information from authorized users to selected Agency Data Environments. The response to these requests will be returned to the authorized user via the SVRDT Data Portal.

FIGURE 3. SVRDT Secure Data Environment



As the SVRDT SDE matures, expansion of data availability, as well as the addition of new participating organizations, is expected. A substantial dimension of SVRDT will be the engagement of the University of California, Santa Cruz (UCSC) as a research partner. SVRDT has convened a Research Working Group comprising UCSC faculty and graduate students and representatives of participating agencies in three Silicon Valley counties. The research partnership and functions are discussed in the paragraphs under the “SVRDT Data Environment Partners” section below.

❖ Core SVRDT Data Environment Partners

Santa Clara County Office Education DataZone

The SCCOE’s DataZone data warehouse houses the core education data set for the SVRDT and is powered by SCCOE’s high capacity data center. The DataZone’s architecture aggregates, normalizes, and conforms data from districts’ disparate source systems using automated extraction, transformation and loading (ETL) processes, and then creates dashboards of metrics that are based on district needs. Over 90 PreK-12 dashboards and 300 actionable

metrics in the DataZone include early warning indicators, state accountability measures and granular data on interventions and dosage which can play a critical role in the exchange of information between agencies about which interventions and at what dosage are most effective in changing outcomes for students. For an example of SCCOE data analysis that led to policy change and improved outcomes for students, see page 16 of this case study. Continuing recruitment of districts in the tri-county region will expand the core education repository and SVRDT stakeholders' capacity to provide more aligned services policies and practices.

University of California, Santa Cruz (UCSC) Research-Practice Partnership

SVRDT will anchor a Research-Practice Partnership that engages faculty and graduate students from UCSC with research and evaluation staff from the three County Offices of Education and participating health and human service agencies. UCSC's research team is transdisciplinary, engaging faculty and graduate students from the Social Sciences Division—including economics, education, psychology, and sociology—and from the Baskin School of Engineering, which includes computer science, data science, and applied mathematics and statistics. The user-driven SVRDT Research-Practice Partnership will conduct three types of research:

- ▶ To analyze data accessible through the SDE to answer questions formulated by school and agency representatives in collaboration with university-based researchers to inform practices (services), programs, and public policy.
- ▶ To study the extent that staff in participating schools and agencies access data through the SDE, how they use data (which will include analyses of the ethics of data use in SVRDT), and the outcomes of data use for children and families.
- ▶ To examine aspects of the methods employed to facilitate researcher access to and use of data available through the SDE, including entity resolution and bias in data from multiple sources.

Through this partnership, research and practice will be linked through ongoing feedback. Issues and problems of practice identified by school and agency staff will provide questions to which research seeks answers, and results of research will inform the practices, programs, and policies of schools and agencies. Additionally, the operation of the SDE itself will be the subject of research, and the findings and conclusions of that research will inform the ongoing growth of the SVRDT, in accordance with rules of responsible and ethical use of data. The source data will be provided in compliance with all legal and regulatory requirements.

❖ Data Trust Application and Use

The application of SVRDT’s integrated data system is dedicated, first and foremost, to enabling educational, health and human services, and juvenile justice professionals to reimagine and innovate in providing services to at-risk children and families by employing information technology and data. Consistent with the model pioneered by San Jose Unified School District, SVRDT embraces aggregating and matching data at the individual student level as well as in support of systems performance dashboards and high-level actionable metrics. Case managers and school-based and agency staff are able—with tightly controlled access authorization and as part of a multi-disciplinary team—to monitor student-specific interventions, performance, and outcomes in order to “go upstream” to determine which interventions have been effective, and how services might need to be customized to ensure student success.

The Core Management Team has identified three major hurdles associated with piloting the multi-systems data-sharing initiative. First, getting the data organized, “cleaned,” and matched on eight data fields is complex when importing from multiple agencies. Second, building consensus around what to do with the data, once matched, will be the next level of challenge. Finally, the change integration process, which entails practitioners adopting different practices and behaviors based on learnings from the data, is perhaps the most challenging hurdle of all. The first multi-agency data-sharing effort will be piloted with San Mateo County, supported by other counties.

As a partner and co-operator of SCRDT, University of California, Santa Cruz will engage faculty and graduate students in multi-disciplinary research teams to conduct the aforementioned three categories of research, all of which are directed at informing practices, program design, data use optimization in the interest of outcomes optimization, and facilitation of researcher data access and use.

❖ **Financial Sustainability**

A 2013 National Science Foundation (NSF) planning grant to the Santa Clara Office of Education was the initial big financial catalyst for the establishment of the databank. This grant enabled the core team to identify the opportunities and challenges associated with developing a three-county, multi-agency data system.

A subsequent (2016) grant from the Chan Zuckerberg Initiative is enabling SVRDT to actively expand the participation of the Silicon Valley school districts in the DataZone, and design and develop the SVRDT Secure Data Environment for the three Silicon Valley Counties. It will do so by utilizing an internet-based technology infrastructure. Long term, the costs of operating SVRDT and the DataZone will be covered through grants and funding from new and/or participating partners.

❖ **EXAMPLE: The Santa Clara County Office of Education DataZone Data Warehouse Supports Santa Clara Unified School District's Efforts to Better Understand and Address Chronic Absenteeism**

As SVRDT completes development of the Secure Data Environment to enable cross county and cross-sector integration, Santa Clara County Office of Education's DataZone, the repository of education data for the SVRDT, has already achieved important insights and impacts simply by making school data more actionable for participating districts.

For example, the Santa Clara Unified School District used DataZone's dashboards on chronic absenteeism to unpack the root causes and disparate rates of absenteeism and to design new, more targeted solutions.

Challenge:

- To understand how different subgroups of students were suffering from chronic absenteeism
- To go beyond the symptom (absenteeism) to identify and address root causes

Findings:

- Data dashboard analysis revealed absenteeism (including suspension) was disproportionate by ethnicity, gender and income.

Actions:

(Beginning 2015-2016 school year)

- The district shifted focus from truancy enforcement to personal outreach to families and to efforts aimed at addressing the needs of economically disadvantaged students
- The district invested new resources (\$3M over 3 years) in student health & wellness, better linking services to schools, additional counselors and professional development for administrators and teachers on trauma-informed care
- Chronic Absenteeism Teams of teachers and administrators now meet regularly to build new processes for re-engaging disengaged students and helping them to get back on track.

Results:

- The percentage of economically disadvantaged students who were chronically absent decreased from 13.5% during the 2015-2016 school year to 12.3% during the 2017-2018 school year.

❖ End Notes

- ¹ The National Laboratory for Educational Transformation (NLET), a private non-profit organization, was involved in early discussions about how to extend the reach of the data warehouse that had been developed and operated by the San Jose Unified School District. UCSC partnered with NLET to pursue an initiative toward this end. The NSF planning grant awarded to UCSC included NLET as a collaborating organization, whose leadership served as consultants to organize, manage, and engage in field-related activities. Before grant activities were completed, NLET and UCSC ended the partnership to pursue differing strategic opportunities.