

Rhode Island State-Led Topical Conversation

Topic: Rhode Island's DataHUB and Public Dashboard

Speakers: Ariel Neumann and Joel Stewart, [DataSpark](#)

Date: January 9, 2019

Overview

[DataSpark](#), which was launched in 2009 and is currently located at the University of Rhode Island, houses Rhode Island's [dataHUB](#), the state longitudinal data system. DataSpark links data across multiple state agencies, including the [Rhode Island Department of Education \(RIDE\)](#), the [Rhode Island Office of the Postsecondary Commissioner \(RIOPC\)](#), the [Department of Children, Youth and Families \(DCYF\)](#), the [Rhode Island Department of Health](#), and the [Rhode Island Department of Labor and Training](#). The mission of DataSpark is to (a) connect data across sectors to support research and analysis, (b) inform policymaking and program evaluation, and (c) improve the well-being of all Rhode Islanders.

The Rhode Island dataHUB currently houses data from several state agencies within education, civic engagement, justice, health and human services, and the economy. The data can be used in a variety of ways to explore trends across the Rhode Island population from infancy to the workforce. For example, DataSpark used data provided by RIDE and RIOPC to explore and create a report on the relationship between higher rates of high school absenteeism and low college persistence.

Funding

DataSpark and Rhode Island's dataHUB were originally funded by U.S. Department of Education State Longitudinal Data Systems (SLDS) grants, which allowed initial development and the forging of relationships between state agencies. DataSpark continues to be funded and developed through a series of Workforce Data Quality Initiative grants.

Online Resources

DataSpark produces many materials, such as **data stories**¹ and **data reports**,² which are [publicly available online](#). DataSpark is working on a statewide Public Dashboard, which is projected to launch in February 2019. The dashboard will allow the public to track the data and efforts of participating agencies, including Rhode Island PK–12, higher education, job training, and workforce metrics. The dashboard was driven by the [Governor's Workforce Board \(GWB\)](#). It serves as a website where citizens can review multiple cross-factor indicators. The dashboard visualizes comparisons across the Rhode Island population and will

¹ Data stories are guided, interactive tours through selected and curated data. They introduce Rhode Island dataHUB users to issues, policy questions, and the current datascape.

² Data reports are shorter than data stories and allow the user to examine a specific topic.

allow one to disaggregate data by indicators, such as race/ethnicity, age, gender, and income status. It will also be possible to compare data between districts and schools within Rhode Island.

One comparison the dashboard visualizes allows users to understand student wage earnings over several years after graduation. The dashboard displays these data in two graphs. The first graph shows the progression of average wage earnings by number of years after high school graduation. The second graph shows this same data, but the data are broken down by the different graduation years organized by color-coded lines. Because the dashboard allows for users to look across multiple indicators, users can compare student earnings for those who participated in career and technical education courses with those who did not.

Q&A With DataSpark

QUESTION: How does DataSpark address K–12 data privacy concerns? Are Social Security Numbers (SSNs) available? Does DataSpark use probabilistic matching?

ANSWER: DataSpark never receives SSNs associated with K–12 data. DataSpark uses probabilistic matching, and there is a two-phase matching process based on other available data sets within the system. If there are features of the individual's data that match with other data sets, it is factored into the score and the data are sent into a second phase of probabilistic matching. Privacy is a top priority; DataSpark works in intricate sharing agreements and must obtain permissions from all agencies they work with to release any data, even aggregate data. When the dashboard goes live, suppression issues must be solved where $N > 10$ to protect the identity of those individuals. DataSpark works with the Department of Health, and it must comply with FERPA and HIPAA. Data involving children are deemed a security risk and are handled with care. There are many different privacy and security protocols, including private swipes into the building that houses DataSpark at the University of Rhode Island. Also, DataSpark analysts do not see individuals' names; the individuals are assigned random numbers and there is no personally identifiable information.

QUESTION: What types of pushback has DataSpark received about data collection, matching, or analysis? How has DataSpark addressed that pushback?

ANSWER: DataSpark is fortunate in that it was initially a nonprofit and a neutral partner. Many Rhode Island agencies wanted to share data with each other, which provided an opportunity for DataSpark to build relationships and initiate the sharing of data across agencies. DataSpark is beginning to be more formal in many policies about governance and data sharing. It is important that each state agency is involved in conversations and involved in the data governance meeting. DataSpark takes its feedback from partners very seriously. For example, an individual from Portland State University aided in the improvement of the probabilistic matching methods at DataSpark through their provided feedback.

QUESTION: Are users able to download and print reports or charts from the data sets?

ANSWER: With some aggregate data, data sets can be downloaded from Power BI. However, any individual data sets, even if deidentified, must come through DataSpark due to security. Many of the data stories allow for the download of aggregate data. A tool called [Community Profile](#) allows you to examine aggregate data by geography that you may also download.

QUESTION: Regarding the analyses that DataSpark conducts, how does DataSpark determine what questions are of priority? Do you receive input from Rhode Island stakeholders about questions they want DataSpark to investigate and report on? Through what channels does DataSpark collect that feedback?

ANSWER: A lot of the data storage is grouped by agency and comes about through working groups with various stakeholders. It is a long process to create a data story. Reoccurring meetings take about six months. The community-driven process, including DataSpark and the DataHUB itself, is funded entirely by grants, not by the state. All projects are at the request of state agencies or community nonprofits, and there is constant communication between these stakeholders and DataSpark about the indicators and the overall narrative that will be portrayed to the public through the data. Furthermore, projects, requests, and questions come from the governor as well and often reflect gubernatorial priorities.

QUESTION: How do you envision the Rhode Island state agencies (RIDE, RIOPC, GWB) using the new dashboard?

ANSWER: DataSpark's dashboard will be used to inform policy, to provide technical assistance support, and to develop guidance and resources. DataSpark's dashboard is intended to be open and useful to the public as well as for state agencies, specifically for reporting purposes. The dashboard is intended to build capacity among agencies. For example, RIDE will get questions from the community, and the dashboard will be a useful tool to point these individuals to. The dashboard can also be used to include statistics in grants or at meetings. The dashboard can also be used for accountability. For example, the governor has a goal for 70% of the Rhode Island adult population to have a postsecondary certificate or degree. With the dashboard, it will be possible to see and account for progress toward that goal.

QUESTION: What role does comparing schools and districts have within the dashboard? Will stakeholders use this to address equity gaps?

ANSWER: Yes, one of the main priorities of this tool was to point out the equity gaps within our society, especially with Department of Education data. DataSpark believes this will be a huge part of how agencies and the public use the tool. DataSpark worked on a tool called [InfoWorks!](#), which also compares public education data between districts and schools to highlight equity gaps.

QUESTION: Are you aware of state efforts that are like what DataSpark is doing?

ANSWER: Many states are moving toward a statewide data hub. Two models are being implemented: federated and centralized. DataSpark is centralized, meaning all the data are housed in one location. Federated models bring in data on a case-by-case basis and the data are not housed in the same location. This federated type of model works better and faster if there is distrust among state agencies. States that have or are working toward statewide data hubs are Mississippi (centralized), Connecticut (federated), and Minnesota (K–12 and higher education data are in the same place, and it is working toward having workforce data in the same place). You can go the [U.S. Department of Education website](#) to see which states received the SLDS grant. Rhode Island has no legislation requiring a centralized data hub; other statewide data hubs have been implemented top-down via governor or legislative decisions. DataSpark goes to events with other states or substate areas working on building an integrated system. Rhode Island is fortunate to be a small state with one Department of Health, one Department of Education, and so on, which makes the process of developing a centralized dataHUB somewhat easier than the process may be in a larger state.