State of Washington

P20W Database and Data Governance Creation

Category: Cross-Boundary Collaboration and Partnerships

Agency Contact:

Jim Schmidt, Office of Financial Management

(360) 902-0595

OCIO Contact: Ben Vaught, Special Assistant

Project Initial and Completion Dates: January 2011 – June 2013

Section 2: Executive Summary

The state of Washington's Education Research and Data Center (ERDC) was created to bring together education data from amongst multiple organizations and spanning preschool, K-12, higher education and the workforce (collectively referred to as P20W data) into a statewide longitudinal data system (SLDS) with a focus on the transitions between these sectors. The goal was to make this information available, where it wasn't before, to state agencies, policy makers, researchers and other data requestors to help answer P20W questions.

This effort is required by law (RCW 43.41.400), and will provide the state with the unified data platform to fuel the next wave of education research and data-driven policy decisions. To achieve this, the ERDC created and implemented a successful, strong, open-data focused governance structure prior to embarking on the creation of the P20W SLDS. As a result, the ERDC has been asked to perform multiple data governance webinars and presentations for various education-related organizations across the nation. Elements of our model have been adopted by other states.

The ERDC, through the P20W Data Sharing Work Group, brought together P20W organizations and other subject matter experts to collectively establish P20W data governance that would allow for the sharing of data with the ERDC, storing it securely in a P20W SLDS to be implemented and operated by the ERDC and further sharing it with data requestors on behalf of the data contributors. ERDC leadership laid out the important foundational philosophies to guide the Group, then worked diligently for over a year to communicate, dispel false understandings and bring the Work Group members together on a set of shared P20W data sharing principles and a data sharing agreement template.

Upon establishment, P20W data governance included 12 state and non-state partner organizations representing approximately 350 organizations, including state agencies, universities, colleges and school districts, enabled the ERDC to receive over 40 data feeds for inclusion in its P20W SLDS and immediately begin releasing P20W data sets to requestors, even before the completion of the system.

Already, partner organizations are using this new data resource. For the 13 months ending March 31, 2014, 197 P20W data requests from multiple types of requestors had been received by the ERDC and 130 fulfilled. Data partner organizations have been able to focus on data uses, methodologies and research rather than pulling and preparing data for requests and figuring out how to share it properly. Data requestors have only had to go to one location, the ERDC, for P20W data and policies, rather than to multiple organizations with various rules and request processes. More and better P20W data has been made available through a simpler, straight-forward process.

Section 3: Business Problem & Solution

Washington's Education Research and Data Center (ERDC) was created because policy makers did not have the data they needed to make effective education and policy decisions. It solved this problem through an innovative data governance model that led to the creation of one of the most comprehensive collections of disparate education data in the nation, from preschool through college and into the workforce.

In 2005, the Washington Legislature created the Washington Learns Steering Committee chaired by then Governor Chris Gregoire and individual advisory committees on early learning, K-12 education and higher education to identify recommendations for building an education system that prepares all Washington students to succeed in a global economy. The steering and advisory committees conducted an intensive study and top-to-bottom review of Washington's entire education system. While other groups had looked at education before, Washington Learns was different in that it was focused on looking at the transitions between the education sectors. When the Committee had questions that crossed the sectors, answers were difficult to find because data was not linked in a comprehensive way. Examples of these questions include:

- What are the demographic, mobility, program, class, grade, and course-taking profiles of students who do and do not achieve and what are their outcomes?
- What high school characteristics are the best predictors of students' success in higher education?
- What are the education and workforce outcomes of high school students with various characteristics? Of post-secondary students with various characteristics?

The state of Washington was incapable of providing any of this information in a reliable way. Historically, education data in the state of Washington was collected and managed at the level of individual agencies and institutions for their own needs. Each of the education organizations has multiple data systems, including Washington's K-12 agency which has twenty-eight separate data systems to collect student, teacher, facilities and financial data. If school districts and public higher education institutions who perform the actual data collection are added, then the numbers of data systems and data sources reach into the several hundred.

In 2007, as a result of the Washington Learns findings, the ERDC was established by the state legislature. Its mission, in part, was to provide data to Washington Learns. The ERDC was later directed to identify P20W research and policy questions, govern the P20W space and pursue funding to create a P20W statewide longitudinal data system (SLDS).

In July 2010, the ERDC embarked on a federally-funded grant program to implement a P20W SLDS system to formally collect, cleanse, match, link and store more than 40 P20W data feeds from various education, workforce and other organizations. The goal of the ERDC's P20W SLDS is to simplify, automate and improve the ability to provide requestors with more and better P20W data and information faster and more efficiently.

At that time, the ERDC did have data-sharing agreements with the various P20W organizations that allowed for the ERDC to request, receive and link data from numerous sectors. However, this was a piecemeal and ad hoc process that was not conducive for supporting a comprehensive P20W SLDS and its outputs. As a solution, the ERDC brought together the P20W education and workforce organizations to collaborate in creating the state's P20W data governance. Before work on the P20W SLDS even began, the ERDC established the P20W Data Sharing Work Group to try to ensure data would be available to populate the system and, just as important, be made available for use by data requestors.

The ERDC acted as the P20W Data Sharing Work Group Coordinator and facilitated the meetings. The Work Group was comprised of a representative from each of the following organizations, including data partner organizations with P20W data to share and other organizations with expertise or customer experience to draw upon:
- Department of Early Learning (state agency and data partner)
- Office of the Superintendent of Public Instruction (K-12 organization under an elected official & data partner representing the state's 295 school districts)
- State Board for Community and Technical Colleges (data partner reporting to an appointed board, representing the state's 34 Community and Technical Colleges)
- Council of Presidents (representing the state's 6 public universities/data partners and reporting to the university presidents)
- Washington Student Achievement Council (state financial aid organization and data partner reporting to an appointed board)
- Employment Security Division (state agency and data partner)
- Office of the Chief Information Officer (state IT policy office)
- State Attorney General's Office (education data sharing legal expertise)
- Legislative Evaluation & Accountability Program (legislative support organization and data user)
- Department of Social and Health Services (state agency and data user)

The ERDC requested that each data partner organization's representative to the P20W Data Sharing Work Group be someone who:
- is assigned to focus on data governance
- is preferably from business; IT representatives were welcome to attend and listen
- is a problem-solver
- believes the data should be shared

At the kick-off, the P20W Data Sharing Work Group's underlying philosophies for creating the data governance were laid out by the ERDC leadership to the group members:
- Data is a key asset that belongs to the state, not to program managers
- Organizations with data are "Data Partners" and "Data Contributors" instead of "Data Owners"
- We are here to figure out how to make P20W data available; Our mantra is "Free the data"

The P20W Data Sharing Work Group met together monthly for over a year to draft a memorandum of understanding around sharing P20W data with the ERDC and P20W data requesters and a data sharing template that the ERDC would require all P20W data requesters to use. Between meetings, the Coordinator would meet individually with the representatives from each organization in a sort of data sharing "shuttle diplomacy" to discuss the concerns and issues members had but did not necessarily want to voice in the Work Group. The Coordinator employed several strategies to arrive at consensus on key P20W data sharing policies:

- Data partners were not allowed to just say "no" to a policy but were required to explain why they thought they couldn't follow it
- The Associate Attorney General (Education) was included as a member of the work group to quickly clear up data partners' misinterpretations about not being able to share data
- Positive peer pressure in meetings from those organizations who supported the goals of P20W Data Governance was leveraged to move all organizations forward toward agreement, making it tougher for others to say no
- Communication with data partners was key

The Data Sharing Work Group created two important outputs:

- Memorandum of Understanding. Over the Responsibilities and Principles for Sharing and Using P-20 and Workforce Data [1]
  The MOU memorializes and details principals agreed-upon by the partner agencies concerning how data will be contributed to the ERDC's P20W SLDS and shared with data requestors and sets forth some responsibilities of the data recipients, whether they are partner agencies contributing data or other requestors. The five principles are:

  1. Education Research and Data Center provides cross-sector, linked data to all data consumers in a consistent, transparent way
  2. Education Research and Data Center maintains the P-20 and workforce data warehouse
  3. Protecting the privacy of individuals is a priority
  4. Partner agency data contributors (at the state and local levels) are experts at understanding and explaining the data
  5. Common understanding and use of data increases its value

- Data Sharing Template. For the ERDC to use with P20W data requestors [2]
  The P20W Data Sharing Template details the policies the Work Group collaboratively identified for data requestors to adhere to upon receiving P20W data from the ERDC. These policies include:
  - Data Transmission
  - Data Security
  - Data Confidentiality
  - Use of Data
  - Disposition of Data
  - On-Sight Oversight and Records Maintenance

---

[1] http://www.erdc.wa.gov/P20W_DG/research/mou_final_201109.pdf
[2] http://www.erdc.wa.gov/data/docs/dataouttemplate.pdf

In addition, the Data Sharing Template requires the data recipient to sign
- o Statement of Confidentiality and Non-Disclosure Agreement
- o Certificate of Data Disposition

While the MOU was finalized in September 2011, it took until January 2012 to obtain 12 data contributors to sign it. As of May 2014, 12 data partners representing nearly 350 data collecting organizations signed data sharing agreements with the ERDC to contribute data for the P20W SLDS and allow the ERDC to give their P20W data to requestors on their behalf.

Once these outputs were created the Data Sharing Work Group sunset. In its place, the ERDC established on-going P20W data governance with the following groups:
- Executive Sponsors-advise the ERDC about P20W research topics needing data
- Research Coordination Committee - focuses on using P20W data
- Data Stewards – advise the ERDC about what data to store in the warehouse
- Data Custodians – communicate with ERDC about data feed changes

Section 4: Significance of the Project
P20W Data Governance helped shift the culture of data management in Washington education, from one of silo'ed custodianship to one of collaboration and sharing, while still respecting the unique requirements of each data set and data provider.

P20W Data Governance improved operations in the P20W sectors by:
- providing a venue for P20W organizations to openly communicate and share data issues and concerns in a constructive environment and facilitate the integration and synchronization of data from different systems
- changing the business model of the data partners, moving all organizations to rely more on integrated P20W data, rather than on individual data sets from each entity
- taking the focus off of data constraints and placing it on improving methodologies for using P20W data
- creating a single location from which to obtain P20W data with ERDC's added value of improved data quality and longitudinalization
- ensuring the flow of data to the ERDC's P20W SLDS through formal agreements

The ERDC's approach to establishing P20W Data Governance was innovative in
- its stance that data is a key asset of the state and does not belong to individual program managers
- establishing it well-before the creation of the P20W SLDS that would house the data
- the role of the business as the lead on data governance rather than IT
- the scope of the number of organizations represented – approximately 350 individual organizations' data is covered

P20W Data Governance supports a key priority of state government, to build an education system that prepares all Washington students to succeed in a global economy, by facilitating the "freeing" of data to inform education and workforce

decisions. P20W Data Governance will enable the ERDC to provide data later this year toward the development of evidence-based approaches for increasing learning opportunities in science, technology, engineering and math (STEM), one of current Washington Governor Jay Inslee's priorities and Governor-sponsored legislation.

Section 5: Benefit of the Project
Washington's P20W Data Governance:

- facilitates and supports a robust universe of data available in the P20W SLDS
    - 13 organizations contributing data
    - 49 data feeds
    - 1.1 billion records in the system
    - 5.6 million de-identified lives linked longitudinally
- enables more and better quality P20W data to be available faster
- decreases duplicative staff effort amongst data partner organizations to generate and prepare the same data for internal and external use
- lead to two of the data partners establishing their own internal data governance
- will lead to better-informed decision making for P20W issues
    - school districts now receive information about the outcomes of their students[3]
- facilitates ongoing collaboration amongst the P20W sectors and other sectors
    - research was performed for the first time on the educational outcomes of students served by social programs under the state's DSHS[4]

The ERDC estimates that the cost of the P20W Data Governance activity was approximately $50,000. This reflects 50% of a year's salary and benefits of the P20W Data Sharing Work Group Coordinator. Costs saved have not been estimated but are represented by the time no longer needed to be spent by data requestors on pursuing multiple data sets through various processes.

Since March 2013, when tracking began, the ERDC has received 197 requests for P20W data sets, analyses, reports or research briefs from customers using the P20W data governance process. Of these 197 requests, 130 have been fulfilled to date, and 67 remain open. Table 1 below depicts the break-out of requests by customer type.[5]

Table 1. P20W Data Requests By Customer – March 1, 2013 – March 31, 2014

---

[3] http://www.erdcdata.wa.gov/
[4] http://publications.rda.dshs.wa.gov/category/19/online/order-by-date/show-20/
[5] Publicly available: ERDC Visual Management Dashboard http://www.erdc.wa.gov/dashboard/management_info/management.html

NASCIO 2014 State IT Recognition Awards
P20W Data Governance Nomination

Media — 5
Community-based Organization — 9
University Researcher — 10
Private Organization/Citizen — 11
Legislator/Leg Staff — 21
Other Gov Entity — 24
School Dist/College/University — 33
OFM/Gov Staff — 36
Partner State Agency — 48