Workforce Data Quality Initiative (WDQI)

Data Security as Practice: A Summary of Observations from Round 8 Technical Assistance

Resource #6
July 8, 2024

Kristin Wolff
Dallas Oberlee
Ben Mahrer
WDQI Technical Assistance Team

Acknowledgments

Many thanks to Toquir Ahmed and Greg Wilson at the US Department of Labor for their ongoing guidance and support; our SPR colleague Krystal Hong for supporting the community of practice, web convenings, and overall project; and all our WDQI grantee teams and subject matter experts and our Statewide Longitudinal Data System state and technical assistance team colleagues for making the WDQI project itself such a meaningful and rewarding experience.

This project has been funded, either wholly or in part, with federal funds from the US Department of Labor Employment and Training Administration under Task Order Number 1630DC-19F-00057. The contents of this publication do not necessarily reflect the views or policies of the Department of Labor, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the US government. More information about WDQI is available at this link: https://www.dol.gov/agencies/eta/performance/wdqi.

Introduction to WDQI

The US Department of Labor (DOL) launched the Workforce Data Quality Initiative (WDQI) in 2010 to support states in developing and improving their workforce and education longitudinal data systems (LDS) and in increasing linkages between them with a view to improving programs, evaluating new models, and integrating evidence-based policies and practices.

With the help of WDQI grants awarded in nine rounds over more than a decade, teams in 40 states have improved data quality, launched new systems, and built new tools to manage and use data more effectively. Most have leveraged grants from the US Department of Education's sister grant program, Statewide Longitudinal Data Systems (SLDS), operated by the Institute for Education Sciences' National Center for Educational Statistics. Over time, states have integrated data from various workforce education and human service agencies and programs, including the Supplemental Nutrition Assistance Program (SNAP), Temporary Assistance for Needy Families (TANF), and corrections and reentry programs.

This document is a summary of observations from Round 8 Technical Assistance (TA) on data security. It forms part of a series of deliverables under the WDQI Technical Assistance contract with the US Department of Labor. The aim is to document the progress of grantees, offering context and insights to guide the design of future grants, projects, and TA activities.

Data Security in State WDQI Grant Implementation

One of the goals of states participating in the Workforce Data Quality Initiative (WDQI) is to integrate, connect, or otherwise link wage data—unemployment insurance (UI) wage records—with Workforce Innovation and Opportunity Act (WIOA) program data comprising administrative records that include demographic information about participants and data about the specific services they received and activities in which they participated. This connectivity enhances states' ability to manage and improve programs, and to evaluate their effectiveness and that of the policies they reflect—a requirement of WIOA legislation. However, linking these data sets raises significant concerns regarding the security and confidentiality of personally identifiable information (PII) that states are required to protect. To address these concerns, states employ a range of security measures designed to protect data integrity and confidentiality while enabling the strategic use of workforce data for program improvement and policy development.

Because of the essential role it plays in states' efforts to connect data systems, data security is one of the topics that the WDQI TA team reviews with grantees during annual virtual site visits. SPR conducts these visits (with full WDQI grantee teams) on Teams, Zoom, or a similar video platform. Before COVID-19, in our TA conversations about data security, most grantees would highlight the roles of data sharing agreements (DSAs) and memoranda of understanding (MOUs) and the importance of attorneys in developing them—a process that could span years. This has changed. Today, the combination of new tools and training and the emergence of new platforms provide states with more choices for accelerating their projects while also maintaining the highest levels of data protection and security.

More Tools, Greater Awareness

Recent discussions with grantees, particularly after the shift to remote work due to COVID-19, reveal a deeper understanding of how tools and platforms can be harnessed both to advance WDQI project goals and to maintain data security, and evidence a shift toward broadly shared responsibility for data security among WDQI team members.

Data security has been and remains a priority for WDQI grantees. State government agencies have employed processes like data encryption and anonymization for decades so that analysts and performance teams could use data for research and other purposes while maintaining the privacy of individuals associated with this data. In conversations with grantees during early WDQI grant rounds, we learned that data security was typically the purview of a single data security specialist or consultant. Data security and data privacy were often conflated and identified as barriers to the development and implementation of MOUs or impediments to data sharing altogether. The process of data sharing was arduous.

But in WDQI Rounds 7 and 8—and especially since 2020—team members approach the issue of data security in increasingly nuanced ways. They have become more (and more universally) fluent in data-security language and tools. Both within and across WDQI project teams, participants in TA discuss strategies like data masking (obscuring PII while allowing access to data needed for research or other purposes), role-based access controls (enabling different levels of data access for different data users based on their roles in relation to the data), and multifactor authentication with greater familiarity and appreciation. This shift signals a higher level of data maturity in line with data maturity models we have used during previous WDQI rounds, as well as with DOL's own data management model.

On the related topic of data matching—linking anonymized records or records without a common unique identifier once data has been shared—teams have also built capacity, again aided by the universally urgent need to work in a remote environment, imposed by COVID-19. Probabilistic matching, long cited as a popular method by grantees, has given way to a greater range of options and choices. Tokenization (replacing sensitive information with symbols that can be matched across data sets) and hashing (replacing data with a character string to allow matching without revealing the underlying data) are two examples.

Toward a Golden Record in New Mexico

New Mexico, a WDQI Round 8 grantee, built a data lake in the Google Cloud Platform (GCP) that included data from all the P20W (i.e. pre-kindergarten through college and into the workforce) agencies and partners. To link data so the partners could use it for analysis, the state used a tool called TAMR (on GCP) to create a master unique identifier, train a data model based on the data in the P20W, and link records from across data sources to their respective unique identifiers, creating a golden record for each individual represented in the data lake. The unique identifier—not the PII—is then used in research and analysis. While the team maintains "humans-in-the-loop" to assure quality and consistency, the TAMR's machine learning approach to records-matching outperformed human matchmaking within months after implementation. And, as the New Mexico team emphasized, TAMR can scale in ways the team cannot.

More Robust and More Frequent Training

Round 8 WDQI teams also prioritized training and capacity building—and not just within their own programs and agencies. Many states are launching statewide information and/or data offices or shared service models, creating new statewide data policies and procedures, or updating job descriptions to better reflect the roles of data managers, stewards, analysts, scientists, security experts, and other workforce data personnel. Some states have adopted government-wide policies that require ongoing data security training and/or certification protocols within each agency. As a part of these shifts, WDQI teams regularly engage in training and capacity building to enhance data security—often in conjunction with new technology platforms that themselves require training to use.

Perhaps more importantly, a wide range of national workforce data training and peer learning opportunities have also emerged that have both increased the demand for agile and

secure data sharing and revealed the value of such exchange. The Coleridge Initiative (and the National Association of State Workforce Agencies, NASWA), Georgetown University's Beeck Center (and the National Governor's Association), Results 4 America, GovLab, the Federal Reserve Bank of Philadelphia, and others have all supported initiatives aimed at boosting performance, increasing data capacity, and/or building cross-agency data capacity, evidence, or learning agendas in ways that require sharing data—while adhering to proper security protocols—and to conduct collaborative research and analysis. These efforts have catalyzed new data collaboratives at all levels, which typically require secure data facilities or secure transfer protocols to support their work. Such projects and partnerships help increase state interest and capacity and shape field-level expectations around data security and use. Alignment around these expectations helps teams to partner more effectively (within and across states).

As one example, the Coleridge Initiative Data Analytics training uses real data shared by states through its Administrative Data Research Facility (ADRF). Most of this training is aimed at developing research, learning agendas, and data products, or requires these as part of an evidence building plan on a particular topic. This focus turns members of WDQI teams (and/or their agency and partner colleagues) into data users rather than data holders and makes the benefits of data sharing tangible.

This focus on training, capacity building, and data use reflects a change in the emphasis of WDQI-supported activities. It also provides incentives and enables the know-how to solve

Evaluating Enrollment to Employment Pathways Advanced Data Analytics (ADA) Training

Originally a two-course pilot that has since expanded to five courses, the Evaluating Enrollment to Employment Pathways ADA training was supported by a DOL grant awarded to the National Association of Workforce Agencies (NASWA). NASWA partnered with the Coleridge Initiative (which offered the training and provided the secure data facility, ADRF) and with Arkansas (a WDQI Round 8 grantee) to provide data. More than 25 states participated—each comprising teams of at least three agencies. The teams used data sets (from Arkansas) that spanned WIOA, TANF, SNAP, RAPIDS (Registered Apprenticeship), and more to launch some 27 projects, providing each team direct experience evaluating outcomes across programs in a secure environment.

security challenges that vexed WDQI teams in earlier rounds.

More External (and Cloud-based) Platforms

The increased use of external and cloud-based platforms and tools is also more common among Round 8 grantee teams than in previous rounds—and has facilitated the training discussed in the previous section.

Again, COVID-19 vastly increased states' interest in these platforms and incentivized new market entrants. WDQI states are embracing these developments. Among previous grantee cohorts, on-premises servers and Oracle Cloud were the primary options for servers and hubs and shaped the ways states were able to secure, link, and share data. A few states—like Indiana, New Jersey, and Mississippi—maintained university partnerships that resulted in hybrid approaches. Today, Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) are cited more frequently. Cost savings is one reason. Scale—the ability to right-size the level of compute capacity, storage, and service as needs change—is another. But states also cite data security as a driver: the platforms are maintained by platform

ADRF: An Accelerator for Arkansas

Arkansas was introduced to the Coleridge Initiative and ADRF in spring 2020—as COVID-19 took hold. In 2022, they began putting their own data—RAPIDS, UI, Quarterly Census of Employment and Wages (QCEW), and postsecondary outcomes data—into ADRF to support research on work-based learning. In 2023, they added SNAP, TANF, and WIOA—using their WDQI grant—and laying the foundation for the multi-state Enrollment to Employment ADA training. The state began connecting Arkansas researchers to its data for program studies and evaluation and is expanding its data pool even further—to include Medicaid, juvenile justice, foster care, and early learning programs. In the state's view, Arkansas' embrace of ADRF has meaningfully increased security:

- Fewer "copies" of data are distributed.
- Authorizations, disclosures, and other tools are built into the ADRF
- The facility provides an audit trail, and the ability to control access, use, and data movement in near-real time.

Robert McGough, the chief data officer with Arkansas Department of Transformation and Shared Services and the WDQI team lead, referencing the use of ADRF for evaluation and research that employs wage data from more than one state, observed, "We've proven it out so many times. For multi-state work, it's the only game in town."

providers relieving states from the need to update hardware or software. They are also located off-site, often in purpose-built complexes not subject to the same physical threats (e.g., natural disasters) as data owners might face locally. Further, many of the data analytics and machine learning tools states want to use are either native to these platforms or offered by integrated third-party applications, reducing the need for specialized expertise able to install custom modules locally. Relying on cloud-based platforms has helped some grantees focus less on keeping data safe and secure and more on using it to create value to customers.

WDQI states are also increasing their use of third-party secure data facilities for sharing data with research partners—such as the Coleridge Initiative's ADRF. RIPL offers a similar service, as do Mathematica and other large research and evaluation organizations, including universities. These services adhere to common federal National Institute of Standards and Technology (NIST) security standards and comply with the Federal Authorization and Risk Management Program (FedRAMP) while also offering the flexibility to adapt to state-level standards where they differ. The experience and domain expertise these platform providers bring inspires confidence among states in efficient, cost-effective, flexible solutions to data sharing challenges.

For recent WDQI state grantees, security remains a key concern but is increasingly understood not as a single, process, task, or compliance issue but as a set of competencies and tools they can embrace to help their states and partners advance shared priorities and achieve shared goals.

Conclusion

Prompted by COVID-19, new tech platforms, and a desire to use the data systems states are building through WDQI, Round 8 TA revealed grantees' new appetite for tools, training, and shared platforms to accelerate state paths to data use.