VERMONT
AGENCY OF EDUCATION

1 National Life Drive, Davis 5, Montpelier, VT 05604
(p) 802-828-1130 | (f) 802-828-6430 | education.vermont.gov

AOE Testimony: Introduction to the Education Data Program

Testimony To: House Committee on Education
Respectfully Submitted by: Wendy I. Geller, Ph.D., Division Director, Data Analysis & Management
Date: January 10, 2020

Introduction to the Education Data Program

Thank you for inviting me to come speak with you today and introduce myself as well as the work my Division is doing at the AOE. To that end, my name is Dr. Wendy Geller, and I am the first Division Director of the newly formed Data Management & Analysis Division. We call it DMAD for short.

To help bring you up to speed on what we've been doing over the last 14 months since our Division was formed, I'd like to give you a quick overview of the general conditions on the ground surrounding our legacy data and technical infrastructure to help provide context and a chance to orient.

I'm conscious that we have limited time today and you have a number of topics and testimony on your schedule, so I'd like to offer to come back and provide additional information on other dates if you'd like to know more so as not to take too much of the time allocated today.

Current State: AOE'S Historic Data/Tech Landscape

To get started, AOE's current state and its historic or "legacy" data and technical landscape has been built in a "point-to-point" fashion. This means that the vast majority of tools and processes needed to collect, manage, and then use data to report out were developed independently from one another.

This makes them brittle because they were not built with other integration or evolution in mind, so adjusting them becomes risky and hard to execute smoothly. This also means that maintenance and upgrades are cumbersome and time consuming, because each process and tool was built on its own with its own special processes, tools, and code. Each piece was independent and largely bespoke or custom built.

Additionally, the legacy processes and tools have been held together by people with special knowledge that wasn't necessarily shared with others. These conditions made the work extremely labor intensive, leaving literally no time for value-added or proactive work. Essentially this means that putting data together to explore important questions has been extremely difficult historically.

Before we move on, I want to note that Vermont is grappling with these challenges in very good company with other education agencies, branches of government, and private sector organizations across the nation and the globe. This is absolutely not unique to us.

But, if we want to be data-driven organizations, if we want to have the information we need on hand when we must take important decisions, if we want to be able to leverage data to support improved outcomes for kids, then it's time for us to pay attention to our data infrastructure so we can get to a better place. But, what does that better place look like? I'd like to show you over the next few slides.

Future State: Enterprise Infrastructure

What you're seeing on this slide is a diagram that illustrates what taking an enterprise approach to this work can look like. It illustrates what our future state can be as we move away from those legacy conditions we just talked about and towards a state where our Data and Tech infrastructure is envisioned, managed, and supported in a data-centric way. I'd like to point out how this diagram relies on integrated infrastructure here, as opposed to those siloed tools and independent people themselves holding discrete, disconnected work together like what you saw on the previous slide.

This data-centric future state means that the AOE's processes and tools will be built with view to how those formerly siloed data should live and operate together. To get there, the tools and procedures have to be examined together, not separately. This also means that the knowledge it takes to run and use them must be shared at an enterprise level.

This future state replaces those brittle qualities we talked about earlier with a condition that is fundamentally more resilient to change. It makes the tools, methods, and processes more durable because they are being built together with view to how we maintain them, upgrade them, and adjust them to new needs over time.

By identifying the siloes and actively working to break them down, bringing the work together, we can find efficiency. Whether it is in removing duplicative steps that were being taken on separate processes because we've brought them together to execute in one step, or in removing risk by sharing and documenting knowledge, this is what continuous improvement looks like in practice.

By finding those efficiencies and being oriented towards continuous improvement, we can become flexible and responsive, both to the needs of the field and advances in data science. We can begin harnessing our information to help take decisions that are forward thinking instead of reactive. We can start having the bandwidth to engage in value-added work, exploring and asking questions, important questions, of our data assets.

But how to get there?

But how are we actually going to get there? I'm sure you're thinking "This all sounds great, but how are we actually going to do it when everything seems like it's on fire." There are so very many challenges in front of us. From resource constraints to the aging infrastructure itself and the cultural fear of change – and by the way, resources are more than dollars and cents, they are people, skill sets, tools, lot of things – there are certainly no shortage of hurdles ahead of us as we strive to realize that future state. But half of the battle is just to get going, to start, and that's exactly what we're doing.

Start Here: DataOps - Culture and People

What you see here is a visual representation of a framework and practice called DataOps. **DataOps** can be understood as "an automated, process-oriented methodology, used by analytic and data teams, to improve the quality and reduce the cycle time of data analytics. While DataOps began as a set of best practices, it has now matured to become a new and independent approach to data analytics.[1] DataOps applies to the entire data lifecycle[2] from data preparation to reporting, and recognizes the interconnected nature of the data analytics team and information technology operations.[3]" (https://en.wikipedia.org/wiki/DataOps).

We are standing up a DataOps practice at AOE as a means to rethink, reorder, and rework how we function when it comes to data. The beginning of this kind of transformation starts with people and culture.

Its genesis comes from creating an environment that prioritizes communication, collaboration, integration, automation, measurement, and cooperation between data scientists.

From 4 Teams to 1 Work Family

At AOE, this evolution started with the reorganization we underwent in October of 2018, which created our Division. DMAD represents the first time all of the data scientists, experts in measurement, analysis, and data management were brought together into one crew, with one reporting structure, with the charge to create a cohesive and coherent data program.

We went from being 4 teams scattered throughout the Agency to being one Work Family. And we're serious about that nomenclature, that's what we call ourselves, identify ourselves, and how we think about ourselves as a group.

This change united the Data Management and Analysis Team (DMAT), the Data Analysis and Reporting Team (DART), the Assessment Team, and the Assessment Data Team. These units respectively represent the full lifecycle for institution's most critical education data assets.

Partner with Extended (ADS) Work Family

Next, our new ADS IT lead, Kevin Viani, arrived, in March of 2019. As part of finding our feet in our respective new roles, he and I took a significant amount of time to examine our collective current state at AOE. We reflected deeply on what our immediate challenges and opportunities were, and we focused tightly on the work of fully defining the roles and responsibilities between our two crews, prioritizing how we document governance and processes, as well as finding ways that our folks could collaborate to problem solve together as partners.

While this work is still in its formative stages, we've successfully secured federal resources and support to continue to give structure to this effort. In fact, we are scheduled to receive additional federal technical assistance in the next two months in the form of facilitation of self-assessments surrounding the data and IT roles and responsibilities and governance as well as action planning as a result of those facilitated self-assessments. This work will shore up and solidify the good first steps we've taken in this area to date.

3-5 Year Breakthrough Objectives

Moving forward as we established ourselves as a Division, after an unflinching examination of our current state, DMAD has developed a set of strategic priorities for the three to five years ahead. They are breakthrough objectives. If we can make meaningful progress in these areas, it will be game-changing for Vermont's education data infrastructure. As we move forward with this work, we will refine them, but broadly, they are:

1) **Modernize, standardize, and fully leverage collection, management, storage, and data analysis platforms, tools, and methodologies.** Ultimately, we've got to have the right tools to do the job. Think about it: It's a real bear to try to hang drywall with a screwdriver.

2) **Move from reactive culture to proactive culture** that supports innovation and continuous improvement**.** We need to be flexible and responsive, both to the needs of the field and advances in data science. We need to be able to harness our information to help take decisions that are forward thinking instead of reactive. We can always be better. We should strive to improve how we work as well as our capacity to support the work of others.

3) **Effectively coordinate to execute cross-functional workflows.** By prioritizing a close examination of how we work, we're establishing standard operating procedures, workflow documentation, and instantiating business process management best practices. If you improve your processes, you can improve your work.

4) **Strengthen security and privacy frameworks while reducing burden of supporting secure and sound data handling.** At AOE, we take our federal and state student data privacy responsibilities very seriously. This can be a tough task with the tiny size conditions we have here. Therefore, we're taking every opportunity to create processes and automation that help us uphold the imperative to protect student data while providing as much information as possible to the public.

5) **Empower AOE and stakeholders with data to support an evidence and result-based approach to decision making.** The AOE is committed to transparency. We're in the process of creating tools that are more flexible and provide more interactivity for public reporting. This modernization will eventually allow us more opportunity to take on one-off projects and respond to changing data needs. We'd like to be able to take on the analysis efforts that we as Vermonters want and need, beyond our statutorily-mandated reporting and analysis requirements (which are hefty). Our goal is to answer this call while balancing our constrained resources.

Now go here: Supporting Processes – DataOps are Lean

After we established where North is by charting our three to five-year Breakthrough Objectives, in order to make some meaningful progress in that direction, as part of our DataOps evolution, we knew we needed to pay attention to our processes. That meant we needed to get Lean.

What is Lean? Lean is:

- A methodology and culture dedicated to continuous improvement.
- The practice of constantly reviewing processes to identify, measure, and learn about where there is waste and how it can be mitigated.
- An environment of continuous feedback loops that drive adjustments and improvements to yield more value.

What does that look like for us at AOE?

It looks like the identification and or creation of standard operating procedures for how we conduct our work, and it looks like a Division-wide effort to visualize our work together.

Doing these things helps us:

- Identify the work
- Define the work in standard ways
- Define who does each part of the work (this includes having conversations with our ADS Extended Work Family)
- Visualize the work (if you can see it, you can work on it and you won't lose it)
- Visualize the process of doing the work
- Identify waste in the process
- Address the waste in the process (if you can see it, you can address it)
- Find better ways to perform the work
- REPEAT

Now I'm sure you're asking, but what does this actually look like? I'm going to show you on the next couple of slides.

Started with Kanban, Graduated to Azure DevOps *Together*

What I'm showing you here on this slide are two real screen shots of live environments that we've built. They contain the standard business processes and workflow documentation that we've created through our practice.

The one on the left is the first environment we used to begin our practice, which was limited to one team that joined our new Division. It was a very lightweight, intuitive tool but didn't quite offer all the functionality we needed once we had "cut our teeth" so to speak so we have graduated into a more sophisticated environment called Azure DevOps.

The screenshot on the right shows you what some of this looks like as of yesterday. It got cut off a bit on this slide as I was trying to fit it into the presentation so folks could still read the text, but at the top the full title actually reads "AOE and ADS Teams." We've begun to be able to work there in collaboration. This is a first.

This platform is designed to help develop and support our agile, lean DataOps practice that can be shared and collaboratively managed across our entire Work Family and in conjunction with ADS partners as they stand up their DevOps practices.

DataOps and DevOps are sister practices that support and propel one another forward through both being founded in a dedication to continuous improvement, one of our core values.

So now that we've talked about what we're doing to support processes, but what's next?

Then go here: DataOps – Intelligent Metadata

Now, many people define "metadata" as "data about data" but because that can mean a lot of things to a lot of people, for our conversation today, I'd like you to try to think about metadata as information that tells us something about some data that we have. In this case, the data we're talking about is data on how we were working and what those data were going to be good for helping us know.

So, once we had many of our standard operating procedures documented in a centralized platform and we're beginning to use shared language and standards, we were able to not only begin supporting one another in sharing and executing work, finding and eliminating waste together, as well as creating automation where we could to free up time, reduce risk, and improve quality, but we were able to measure important characteristics of our work.

We were able to collect data about how we were working, like how long certain work took, how many iterations we had to take before we could call a task complete, how many steps or tools were needed to complete a product, etc. By knowing what those data could do to help us learn about our work, we could use those data well to inform us about how we could work *better* together.

Collecting those data allowed us to intelligently measure our work, report out, and then drive process improvement. For example, what you see on the right-hand side of this slide is another real screenshot of a dashboard that I built in Power BI. Power BI is a data visualization tool. I made the dashboard you see screenshotted here to communicate about workflow to my supervisor, Deputy Secretary Bouchey.

What you see here are data about the type of work that was carried out across each quarter for a year and a visualization of the volume of hours it took to complete certain types of work across that timeframe. Leveraging these data this way allowed me to communicate my crew's work allocation to Dep. Sec. Bouchey. By doing this, I could make enough time and space for us to complete certain work within the time we had available by reallocating resources and deprioritizing other work that was a competing for our time.

It also informed me how I could allocate additional time and space for my crew ahead of high flow time periods so we could focus on the most pressing workflows. This helped me and my

team balance our loads and execute on the highest priority tasks while planning ahead for when and how we would tackle the rest of the work that remained. And let me tell you, there's *a lot* of work that remains!

This culture and practice positions us to go here…which leads to here: DataOps – Enabling Technologies

So, now that we've talked about the concrete pieces, we'll zoom back out to tie this together by highlighting that it is this culture and practice of DataOps that will ultimately position us to be able to achieve that future state that we saw and talked about earlier.

I've included that diagram on the right-hand side of the slide here so we can look at it again together. This future state is one where we've got transformative technologies and practices that help us and our infrastructure be data centric, resilient to change, supporting our continuous improvement efforts, and continuing to free up resources for value-added work that explores the data we steward at AOE and puts it to work to try to answer the questions we all have.

It's early days yet, there's a whole lot to do ahead, but this is what my Work Family has been able to do to date and I'm proud to have been invited here today to share it with you.

Thank you for your time, and again, I'm happy to come back another day if you'd like to discuss more. For now, with the remaining time I may have left, what questions or conversation would you like to have?