NOTES FROM KICK OFF MEETING
PORTLAND, OREGON – OCTOBER 28-29, 2010

General Goals/“Wins”

One thing the data exchange project may be able to do of value for everyone is to establish what falls within the bounds of feasibility in terms of answering questions from the legislature, boards, leadership, and other sources.

Generation of some simple, powerful reports that enable a state to “tell a story.”

Networking and learning from other states and agencies.

Using longitudinal data to have a more comprehensive picture of how human capital is developed and how it moves around, and particularly to create and use meaningful indicators of progress and outcomes, to disaggregate information for more effective targeting of policies and practices, and to account for the non-linear way in which individuals access formal educational structures.

Deliverables

1. Aggregate reports
2. Some data back to states while protecting privacy
3. Lasting governance structure
4. Dissemination of lessons learned

2-tier option: 1. Aggregate reports for all (safe directory info) and 2. More detailed reports for others with bi-lateral agreements

General Issues

Layering complexity of the multi-state exchange on top of the existing SLDS activity, but cross-state tracking is growing in importance. E.g., in Washington, this project can assist in helping the state know whether it is meeting system design plan goals; questions about mobility are asked by the legislature all the time.

How will this effort be made to be sustainable? Also, there’s a need to appreciate how this additional burden must be carried ("project-itis").

We need to resist the urge to go too far too fast; rather, we ought to demonstrate high impact with a set of initial data elements that can answer important questions. We don’t need a massive system to answer every conceivable question – Keep It Simple. The limited, incremental approach will likely lead to more demand by demonstrating the value of sharing.

Lots of work is needed around data standardization and standards. Even within state, there are loads of issues of complexity and inconsistent definitions. It is really important for this stuff to be meaningful.
There is also a need to reconcile the various external reporting and initiatives so that they are operating in a common way. To the extent possible, this exchange should not cloud the policy picture by introducing analytical approaches that are irreconcilable with other, purportedly similar information in the public space.

FERPA – institutional constraints and UI laws – federal policy is getting in the way and there need to be fixing there. Data paranoia, Big Brother. Is it a general paranoia or are there specific fears?

We need to be conscious of the Law of Unintended Consequences. How can the data be used in ways that end up harming individual schools or possibly students? How do you put some governance around the way the data are used, i.e., when they bust some myths that are highly politicized? Capabilities vs. expressed intentions and how you put some rules/governance in place to limit what can be done with the shared data.

Need to expand metadata to address data quality, trend data, etc. How do you make the data actionable? How do you change the organizational levers using the data system? Align measures across sectors and understand the incentives that data provide. NSC has found that the most useful information was provided by derived data elements.

How public do institutions want data to be? How easy would it be to isolate a specific institution? Should the report just produce the quantitative analysis, or should it also include interpretive commentary? Who is the ultimate audience for the reports?

We ought to be thinking about a formal evaluation plan with transparency in methodology, research questions, and governance all made clear, along with a clear definition of what constitutes success.

Investigation of what is the state of data – to what extent are the data just not there? So that each state and agencies within it can point to an authoritative source that shows the data aren’t there yet.

Managing of expectations around data usage and information produced will be important.

**Policy Research Questions**

What kinds of disaggregations do we certainly want to be able to do? Limited English, race/ethnicity, income (defined how? Pell? FRPL? Geo-coding?), age, small high schools/other curricular/structural reforms, parents’ educational attainment level, special education/disability

Replication of studies like Washington’s tipping point study in Oregon and other states.

Workforce pathways – leaving education aside – reverse research design; current education production → where are the gaps? Where is there excess?

1. How are former high school students performing in postsecondary education?
2. How are former high school students performing in the workforce?
3. How are former post-secondary students performing in the workforce?
   * Mobility of students from high school to college
• Completion of college relative to high school performance/location/curriculum
• What does it mean to be college-ready?
• Initial college choices and transfer patterns?
• Students concurrently enrolled?
• Mobility of K-12 students among states
• What are the “educational footprints” of workforce participants? What are the “workforce pathways?”
• Time to employment post graduation?
• Performance/Wage gain/Time in the workplace?
• Were students employed in the field in which they were prepared? Are students earning wages at a rate at which is commensurate with their preparation?
• What is the balance of trade among states relative to educational preparation?
• Student borrowing relative to wages – gainful employment.
• State investment in disciplines vs. where they are working.
• What is the “balance of trade” in labor among states especially relative to educational preparation?
• Teacher performance relative to their training

4. How are current and former workforce participants accessing formal education systems?
• Workers returning to college?
• Students who worked concurrently?

A helpful way forward might be to develop a matrix cataloging the relative importance of the question vs. availability of information needed to answer it. How will we decide what to answer and how difficult it is to answer → governance structure.

State Capacity/Data Issues

Hawaii
• Can’t do the high school to workforce match, and also can only do the match between K-12 and UH via fname, lname, and DOB (plus R/E as a tie-breaker) – HIDOE has no SSN and UH doesn’t capture HIDOE identifier.
• A common K-12 database has existed for only about two years and big data quality issues remain.
• Recent Race to the Top award commits Hawaii to an integrated database, but does not provide the resources to create it. 2010 SLDS grant application was not funded, but the existence of it is assumed in Race to the Top.
• HI-PASS is matching 96% but may not be sustainable. Currently 85% match with UI data.
• Incorporating ACT/SAT and COMPASS data.
• Analyses so far have been done on an ad hoc basis, but they are trying to get the pieces in place for something more systematic.
• Hawaii P20 pulls the data and then hires analysts, currently trying to build connections with UH economics dept. Meanwhile, it is still sorting through what the hardware and software requirements are
to create a central repository, trying to move forward with data governance structure (and the idea is to create a parallel governance structure compatible with the multistate exchange). They hope to resolve existing data quality issues in part by using the data and surfacing the problems and how they impact the analyses informing policy.

- There is a need to educate the legislature about the value of these data.
- The WICHE project provided an impetus to convene the parties, through which they did a map of who had what data and how/where it was housed. MOUs have been developed. UH was originally intended to be the SLDS managing partner, but there may be some backing away from this in favor of a P-16 partnership.
- HIDOE received a 2009 SLDS grant to build out the K-12 piece of the SLDS. They have issues with the unique ID. But they have only 2 years with a single student information system, and there remain lots of pieces to bring in. One year ago they created a Data Governance Director. There remains the challenge of distinguishing the transactional uses of a data system and policy research uses (and that Hawaii P20 is not intended to be transactional). But that problems surfaced by the SLDS with data integrity problems, then helps them to know those problems also exist at the school level and in the transactional system.

Idaho

- Workforce data merged with universities, but on an individual basis for which workforce does the match.
- Each of the higher education institutions captures data on their own via separate and unevenly accessed MOUs – there is no state SUR and one result is inconsistent definitions.
- SLDS development in Idaho has a somewhat checkered history with a failed $40 million Albertson’s grant.
- Workforce match to K-12 is also compounded by a lack of SSN.
- Current activity in SLDS development is led by the Department of Education through $2.5 million funds committed by the legislature for a K-12 SUR system, compounded with $5.9 million from the federal SLDS program in 2009. That system was appropriated from Nebraska and is now live in all but two school districts. There is not yet consensus on how to integrate postsecondary data – whether it should be appended into the K-12 system or by way of linking in a separate system. The K-12 system has a single “Educational Identifier” that applies to students and staff. Board is moving toward using that ID for postsecondary students. Goal: by May to have course completion, transcript info.
- Currently working on a needs assessment – how to sustain. Partnerships in place with DOL for nursing programs. Participating in CCA so they need to figure out progress/outcomes metrics that each institution needs to adhere to. Once the Ed ID gets into the secondary data, then SBE will consolidate the postsecondary education down. Labor has started MOU process with transportation to get more demographic information allowing matching with the educational system.

Oregon

- K-12 does not include SSN, but they are trying to get it from Drivers’ License or ACT Work Centers to link to workforce.
- Department of Community Colleges and Workforce Development doing UI matches for the CCs
• SLDS development to date: there are unique state IDs for students and staff in K-12. The Network for Educational Research in Oregon (NERO) is a common network for postsecondary and K-12 education, preserving data in and “education agency” for FERPA purposes, which is integrating data in distributed 6 regional K-12 data warehouses throughout the state. OR launched a statewide student ID launched in 2002, a staff ID in 2005.
• OR has had a total of three IES grants: 1. Data project (2007) – direct access to achievement: primarily a professional development effort to build a culture of data use through data teams throughout the state. 2. OFAR (2009) – OR Formative Assessment Resources – statewide rollout of the formative assessments and integrate it into pre-service curriculum; also move the data center’s physical infrastructure to Corvallis. 3. ALDER (2010) – create an educator-student data link via course information and other specific data pieces; scaling up and integrating early childhood and kindergarten; same with postsecondary; design and implementation of statewide data quality plan with governance. Develop a list of possible sources to link with workforce – “pretty weak language in the grant.”
• Integrating data with data governance for all three sectors is part of current SLDS project, but it will be three years before all is linked. Shared data for this exchange will have to be obtained from three separate sources in the near term.
• OUS is getting 70+% matching just fname, Iname, DOB. Oregon Identity Theft Protection Act 2008 – removing SSNs in K-12.
• OUS will have an identity resolution system. Annual data exchanges within OR via MOUs, with established procedures, also ad hoc.

Washington
• There is a “system design plan” for P-20 generally, and it should be integrated with the SLDS effort – tracking across state lines is an important component.
• Workforce data has been shared with Oregon (and Idaho?) for years.
• Education Research and Data Center (ERDC), housed in the state’s Office of Financial Management (OFM) is housing integrated data. ERDC is an “education agency” for FERPA, intended to focus on transitions among sectors. Their governance arrangements could serve as a model for what the multi-state exchange is trying to do.
• Currently ERDC is developing conventions on who can see what data under what circumstances. ERDC wants to take a preliminary look at the data on the statewide level, but then offer local folks to look at the data they are allowed to see? How do report recipients learn how to gain value out of the reports and/or the datasets available? ERDC will be traveling via grant funding to try to bring together stakeholders in all 3 sectors to get what they need out of them and then to train them on the use.
• ERDC has its own unique identifier. K-12 (OSPI) does not carry SSN, but ERDC is getting SSNs from ACT/SAT. Getting about 50%, used to be getting a higher rate by linking to drivers’ licenses and they’re trying to do it again.

Other Longitudinal Data
• NSC is matching with combinations of DOB and name and getting 98%. Wants to leverage its “agent” status, extending it to DMV, LMI to import records to do the matches and create an enhanced dataset.
• Census’s LED combines tax and workforce records to track all employment. Gets the four-digit industry data by county. Census gets it, replaces the SSN with its own identifier. All 50 states have an agreement with the Census Bureau to exchange data these ways. LED – you don’t get identified data back – you only get summary data. You might be able to send a list of SSNs to the Census to get those files populated, but it takes a lot of resources on the Census Bureau’s part and the returned data aren’t fine-grained enough for answering many questions of interest.

Are there research questions that can be addressed without an SSN?

Architecture

Washington is the only state that is ready to share integrated data (with possible exception of workforce data) in the manner WICHE first envisioned. But each of the other states is moving in this direction with respect to K-12 and HE data, to varying degrees. Oregon is closest, Idaho farther away (largely because HE institutions have independent data holdings), and Hawaii somewhere in between.

Big problem appears to be workforce data (UI). Agencies maintain that they cannot share this data in “identifiable” form, which means any combination of elements that might identify an individual. One interpretation maintains that this means that demographics cannot be passed through the system for analysis. They are also concerned about the number of data elements being matched, as the more there are, the more there is a possibility of identifying an individual. Also, they do not like algorithmic assignment of project ID numbers and prefer a centrally-generated random ID.

A possible interim solution is aggregate reporting from NSC. The minimal version of this would just include education data, though aggregate labor data might be a possibility if very simplified (e.g. employment status by state). On the other hand, three of the four states (OR, WA, ID) have agreements to share UI data amongst themselves and do this regularly. This may raise the possibility of adding UI data (collective) after the educational merged files have been created.

NSC has the capability to generate a random ID. They do this now for internal purposes but cannot use it to make a match. So they could do it for the consortium.

What is the role of the AGs in interpreting FERPA? Should we get a ruling on this in general? Should legal counsel be involved in governance in each state? Also, need to be award that this is not just FERPA, as there are legal/privacy considerations that are specific to labor data.

Governance

What does governance need to do?
• Give appropriate voice to all parties
There’s a difference between voice and vote in the long run. In order to make decisions, you need a manageable number of people in the group with authority, who have established trust within the stakeholder groups they represent.

Ideally, there would be a governance structure “expected” within each state that then rolls up to the multi-state level. But first we need to generate consensus among all the stakeholder agencies as part of the “construction” process, as distinct from governance. But we need to define when it is logical to move. And shrinking the group is understood at the outset and eventually we need to move to a “tiered” or “team” approach.

Eventually: one idea is that representation may be one person per state, but perhaps also one person – maybe rotating – representing the labor perspective, also one from each of the three levels. The rep from each state is a consensus-builder for his/her state.

It may be necessary to develop a subcommittee structure for each sector to which the larger representative group can refer questions.

It may be worthwhile to engage an individual from each sector in sort of an ombudsperson role.

It must be recognized that each state is in a different place with respect to their internal governance process. So when do we ask the question of moving to a more streamlined governance? One idea: at the point at which we have a report ready for delivery.

Need to get a handle on the value-added data products already being produced without formalized data governance.

A separate body of policymakers? When do we need it? Who should be on it? What should it do? Perhaps a WICHE commissioner? Some kind of communication and consultation pathway to governors’ offices?

Policy committee charged with oversight over MOUs, research direction, access to data.

Technical committee charged with data audits, definition alignment, quality control, review, etc.

- Set the rules of engagement
- Provide a mechanism for conflict resolution
- Determine who gets to make the decisions and what accountability is there?

Issues for governance

- Overall operating agreements and MOUs
  - Establish boundary conditions
  - Transparent and accessible to all parties
  - Require formal consent and amendment procedures
  - An idea: a single MOU with each state in which each data owner signs off on it for use of their data.
- How permanent is the resulting data file?
- What’s the communication plan when reports are produced?
- Who gets access to the data? Who gets to say and under what authority?
- Who “owns” the resulting data?
  - Who is responsible for its integrity?
  - What conditions govern reporting and use?
- How are the provisions of a data-sharing agreement enforced?
**Going Forward Discussion**

Thoughts on sustainability after Gates grant: Will WICHE continue to support this effort? This might be a subscription run by WICHE. Might be a third party created for the purpose.

Need to construct multiple scenarios of what this might look like, together with their data structure and governance implications. Might include:

- Original vision (“super ERDC”)
- Common core of data within the above
- Series of bilateral agreements (coordinated in some way)
- Consortium of common procedures and definitions (like the Common Data Set)
- What does governance “mean”? Who gets to make decisions? What accountability is there?

All agreed in principle to the goal of [eventually] creating a single point of contact for supplying data for each state. But they acknowledge that some states have not yet created their own governance structure for data exchange and they have all are committed to moving in this direction.

The actual role of NSC is still unclear to many participants. They originally appeared as a candidate matchmaker, then moved to being the actual matchmaker, and the last session made it appear as though they were a main player. Some states that already do a lot are concerned about this. This role needs to be clarified.

Are we making this too complicated? What if we just focus on trying to replicate the NSC reporting but at the state level and with labor data?

Project staff assignments that arose from the meeting for follow up include:

- Selecting research questions and having participants rate them in terms of priority and difficulty. Cross with needed data elements and availability.
- Collecting and analyzing (and perhaps distributing) all interagency MOUs.
- Coordinating a process of auditing data elements for each participating agency.
- Refine matching proposal (e.g. Andy Mehl’s proposal).