Overall Impressions

While the discussion certainly did not follow a strictly linear path, the following are important areas where greater (though far from adequate) clarity was achieved:

- As reflected in feedback supplied by the participants via a survey conducted in December, the multistate data exchange is most likely to add value by focusing its data assembly and analysis work around linking education to workforce data (an analytical frame that several referred to as the “balance of trade” among states). That makes sense given that individuals in the workforce are generally more mobile than are students, particularly if they have achieved a postsecondary credential. But it also reflects how systematically linking education and workforce data across state lines is the linkage that states have the greatest difficulty doing on their own.

- The nature of the workforce data in the various states—how they are organized, accessed, and used by the various workforce agencies. Of particular note was that it may be feasible to use Census LED as a resource for collecting UI wage record information in one place.

- The developing extent of the National Student Clearinghouse’s (NSC) K-12 holdings. There exists the long term possibility of using state-reported high school records held by the NSC to build a viable cohort down the road (with the actual members of the cohort chosen by the states instead of the NSC).

- The specific data elements that we will need, with some progress on standard definitions for these.

- The basic logic of a longitudinal data report. This is pretty solid from college entry on with respect to when to report what. How to address secondary enrollments and graduates is still being considered. And the proposed reporting template has value regardless of the approach taken for assembling data, that is, whether we use a “built-up” approach starting with state student unit record (SUR) data or a “national” approach starting with NSC and Census-LED data.

- That any progress the exchange can make toward filling gaps in state data will be a helpful step forward (the “minimum effective dose” analogy).

- The things an MOA eventually needs to say and the kinds of language to use (see below).

- That WICHE and its partners intend to work closely with participating states and agencies in reviewing and releasing initial analytical products of the exchange. WICHE and its partners will continue to facilitate the development of a governance structure to address how analytical products are treated in the long term.
At the same time, there are still a number of basic issues that are not yet resolved, including:

- All agree that the initial analysis should be seen as a “proof of concept.” Less clear was exactly what are the concept(s) we are setting out to prove. Possible candidates include:
  - That a multi-state data resource will tell a compelling story that could not be told using each state’s data resources alone. An important part of this is whether doing this is worth the investment.
  - That state agencies in different states are capable of systematically and seamlessly exchanging actual data elements drawn from their own data files with one another to create an accessible and useful longitudinal data resource (the concept described originally in the Prescott-Ewell paper).
  - That the politics and regulatory structures (including FERPA and other privacy law interpretations) in each state can (mostly) be overcome using some common approaches, as opposed to having to do things entirely differently every time a new state is included.
  - That the approach is one that can gradually be expanded to include more states in a given region, perhaps following the current regional compact lines.
  - That the research questions have real traction with potential policy consumers.

- Whether de-identified data will really suit agency needs. WA-SBCTC is the primary case here, as they can pretty much do all of this now and yield identified data, which they can then link back to their own data resources in all kinds of useful ways. Everybody else would like to do this too. As a result, they are looking to us to protect a potential path back to identified data, even if this may not be possible in the short run.

- Whether to take a) a “built-up” data exchange approach in which each agency supplies data to populate a state longitudinal record and this, in turn, is used to construct a multi-state record or b) a “top down” national approach in which multi-state records are assembled from an existing central store (NSC for education data and Census-LED for workforce data). The former allows for greater detail, and the construction of outcome measures not currently (or anticipated) in NSC. But this approach can be seen as not adding value by states like Washington that already do all this themselves or by Oregon that already have (or will soon have) everything they need except workforce data already in the NSC. The latter is quicker to implement and will probably yield greater consistency in what is reported, but cannot easily be expanded beyond the current contents of the existing national databases, and the degree to which the data can be redisclosed back to the originating state may be more restricted. The former is more consistent with what the project originally purported to be about and will support far richer and more detailed analyses, but will take longer to implement for the states that are less fully developed.
• Who speaks for each state? For purposes of moving forward on the next immediate step (new iterations of the report formats and MOA), we made progress here with the following assignments:

  - Washington – ERDC (Melissa)
  - Oregon – Oregon State University System (Bob Kieran)
  - Hawaii – P-20 Project at the University of Hawaii (Jean)
  - Idaho – SLDS at the State Board of Education (Andy)

But this does not constitute a permanent solution to the governance and voice question at this point.

Some Specific Issues

Within this general overview, we discussed in some depth (though did not entirely resolve) a number of particular issues.

• The Role of the NSC. There was healthy and helpful discussion of this matter, but it remains a point of discomfort for many participants. Dimensions of this topic include, at least, the following:

  o Whether the NSC is used as a primary data source or largely as a matchmaking utility. We made progress in collectively recognizing that these two roles are functionally different from one another and need to be visibly distinguished, although there was no final consensus about how this should be done. Peter offered the suggestion that each state be responsible for using NSC data to fill out their own state’s data record, through a separate arrangement with the NSC, before moving to the merger stage. Some states more or less already do this but there appeared to be uncertainty remaining about the resource consequences of such an approach. If so, we need to reinforce the message that the project can cover such up-front costs. The role of the NSC in this regard, of course, also depends on a decision about the basic architecture question—“built-up” or “national.”

  o Sustainability. Using the NSC as the matchmaking entity within the three-year duration of the project was not unacceptable to participants. But many are asking the inevitable question of how a central NSC role (whether as matchmaker or data supplier) can be supported in the long run. Although not directly expressed, there is also the fact that there are limits to how far down this “national data resource” path that the NSC can go with state agencies that govern or coordinate institutions without risking institutions cutting them off with respect to supplying the data in the first place.

  o Perceptions of losing control of the data. There remain concerns among participants about the level of safeguards that will be placed on data access and use and some about the proprietary nature of NSC’s matchmaking process. As one participant put it, there are “transparency” questions about the “reliability of the match.” This array of perceptions suggests that we
include much more explicit language in the MOA about the NSC’s role in the project. They also suggest that a good deal more mutual understanding about NSC’s intentions and procedures must still be achieved among state participants. Finally, it suggests that an architecture based on the states selecting the cohorts is imperative if the “national” approach to building longitudinal data files is selected.

- **Cohort Architecture.** Two alternative cohort designs were considered, one based on a high school graduating class tracked forward and one based on an entering postsecondary cohort to be tracked forward and backward. It revealed that the group’s preference was for the former, at least for the initial look at the data. Doing so is conceptually clean, easily understandable to policymakers and the public, and raises the ultimate possibility of knowing more about those who drop out of the pipeline entirely. There are some issues that must be taken into account with such an approach:

  - A probable delay in getting longitudinal data over the desired time period. Some states are still at an early stage in developing their SLDS resource on the K-12 side. For instance, Oregon only has high school graduate data from 2004 forward, which will make it difficult to fit a seven year tracking window (year of high school degree plus six subsequent post-secondary potential enrollment years) within the parameters of the current Gates project. The only fix here would be to relax the constraint of seven years and see where we are after only X years.

  - Leaving out of the tracking picture a large number of entering college students. This includes returning adults coming forward after delaying college entry, those exiting from GED, ABE, and workforce training programs, and incoming transfer students from other states. While beginning with a high school exit cohort is conceptually appealing, it does not allow investigation of many policy issues of interest. Knowing this, we may want to find some ways outside the originally agreed-upon high school exit cohort structure to allow a parallel college-entering adult pipeline to be constructed. Washington SBCTC has already done this, and could be used as a model.

- **Report Content and Design.** Regardless of cohort architecture, the group took a major step forward in fleshing out the longitudinal report design. For example, the tentative decision to include annual tracking snapshots of highest degree attainment and current participation in college were major gains. The initial proposal of capturing workforce data at the beginning and end of the tracking period was suggested by several workforce people on day one for simplicity but, like postsecondary educational data, we may want annual snapshots instead. This would enable an immediate look at value-added with respect to workforce outcomes for shorter term credentials like certificates and associate degrees whenever they are earned in the tracking period. Use of the Census-LED approach probably makes it practicable. It might also be possible to propose additional annual snapshots of all outcomes (postsecondary and workforce) beyond six years, although that would be beyond the duration of Gates support. The only thing we cannot immediately fix is
the inability to link workforce data to high school graduates who do not go on to postsecondary study because of the lack of SSNs.

- **The MOA.** We also made some good progress in modifying the draft MOA—or, at least, in identifying the major issues to be addressed. Among the more important “take-aways” from this discussion were:
  
  o Constructing MOAs with as similar a set of contents as possible across states. A good suggestion was to include common understandings and agreements at the beginning of the document, with necessary state-specific modifying language in paragraphs toward the end.

  o Including a good deal more explanation of each step regarding what specifically is going to be done by whom. This may be especially needed in the language about measures to ensure data integrity and access, where specific language from FERPA and other applicable privacy laws might be helpful.

  o Including a whole section dedicated to the NSC and its role in handling data. The language of this will, of course, depend on the decision of which data assembly approach (“built up” or “national”) is eventually decided upon, but there was consensus that the legal folks would need to see exactly what the contractor’s responsibility would be and the limits on the contractor’s potential actions.

  o Crafting the language of this original MOA as governing a “study” or “project” of limited duration, with the resulting data destroyed at the end. During the project, a different MOA to govern a more enduring structure of “data exchange” can gradually be evolved. The annual renewal of the MOA seems like a good idea and allows for this kind of flexibility to be built in.

  o The language throughout should be crafted to refer to “records containing data on individual students” to emphasize the need to protect the rights of real people, not just administrative records. This is also consistent with FERPA.

  o The language might put explicit limits on the eventual use of the resulting data. We did not elaborate this, but there are several examples in Florida’s policies and ERDC may have something.

As a follow up before the next meeting, a subcommittee of the primary state contacts (Melissa, Bob, Jean, and Andy) will work with Brian to revise the language, operating largely by conference call.

- **Next Meeting.** All agreed that the next meeting of the full group should not take place during the legislative session. Meeting this condition for all states except Oregon seems very doable, suggesting a meeting in early to mid-May. Oregon could be partially accommodated by again meeting in Portland.