



Introduction and Context

As the use of data for strategic decision-making in education continues to expand, states and districts have been addressing multiple issues regarding the collection, housing, use, and communication of data among various individuals and entities. State education agencies (SEAs) and local education agencies (LEAs) have been focusing on their data systems through the development and implementation of their state longitudinal data systems (SLDS), data warehouses, data dashboards, assessment systems, and various other elements of the overall data process. As these components continue to expand and improve, the introduction of a comprehensive and well-designed data dictionary has emerged as a critical need for states.

Currently, the education field is addressing the issue of standardization across many levels—from districts to states to federal. Many states are focusing specific attention on the development of data dictionaries. At the federal level, the Common Education Data Standards (CEDS) project has developed voluntary, common data standards for a key set of education data elements to streamline the exchange and comparison of data across institutions and sectors. In their data dictionary efforts, many states are incorporating CEDS and using its associated tools to guide their efforts.

This paper looks at the status of state data dictionaries in selected states in order to highlight states' experiences, common challenges, and guidance for other states. The purpose of this paper is to provide a timely perspective on the development and implementation of state data dictionaries, and to offer guidance to states as they navigate the process. In addition, the paper considers how tools such as data dictionaries factor into larger data initiatives such as CEDS, *EDFacts*, and SLDS, and the relationships that can and should exist among such initiatives.

Benefits of a Data Dictionary in an Education Data System

As the use of data in educational assessment, planning, and decision-making continues to grow, so too does the sharing of data across multiple sectors. Data users must be sure that the information they are sharing and utilizing is clear, consistent, and accurate. The Data Quality Campaign states, “State policymakers and educators need a data system that not only links student records over time and across databases, but also makes it easy for users to query those databases and use up-to-date reports to adapt to the unique needs of their students” (p. 1). This type of comprehensive data system must include an effective data dictionary.

Educational organizations such as SEAs and LEAs often realize their need for an institutional data dictionary when they are faced with many of the data inconsistencies that can arise without such a standard: inconsistent definitions, inconsistent naming conventions, varying field lengths for data elements, or varied element values. A well-designed data dictionary improves the data quality of an organization by ensuring data integrity, eliminating redundancies, increasing consistency, and allowing effective communication. As SEAs continue to expand their use of data for strategic decision-making, this type of data quality is essential.

In recent years, many SEAs have chosen to share their data dictionaries with other agencies—in several cases publishing the dictionaries publicly. Doing so can provide multiple benefits, both to the agency itself and to the larger educational field. Sharing

dictionaries has provided an impetus for conversation among states and districts, as data leaders bring questions or commentary to the publishing state in order to better understand their tool and its usage. Receiving questions and feedback from other states or LEAs can allow the publishing state to recognize needed clarifications or expansions to their dictionary. These conversations can also provide the publishing state a perspective on how it is faring in its dictionary project in comparison to other agencies, highlight opportunities for collaboration with other states about design or metadata decisions, and reveal areas to draw suggestions from other states about vendors and implementation of initiatives.

States that publish their dictionaries publicly also benefit the larger education community. They provide examples and guidelines for other states and education agencies in terms of chosen data elements, metadata, system design, and various other dictionary aspects. Having this type of example can be very helpful to SEAs that are earlier in the development process, in that they can learn from publishing states' experiences and implement those elements or system designs that are most likely to work effectively for their particular state data system.

Beyond these benefits to states in the development process, open access to multiple data dictionaries also increases the alignment of data shared among different agencies and sectors. While data initiatives such as CEDS are designed to increase the communicability and portability of data across sectors, sharing of data dictionaries among SEAs and LEAs can be an early indicator of the alignment of data elements and metadata among education agencies. In fact, the CEDS Align tool (discussed in greater detail on page 9) builds upon the simple sharing of dictionaries to assess alignment, allowing states to compare their elements to CEDS and to the dictionaries uploaded by other states.

Fundamentals of Data Dictionaries

Simply put, a data dictionary provides the names, definitions, and attributes of data elements within a data system. In addition to providing varied users a collective understanding of how data will be used and expressed within a given context, it provides important metadata about the elements in the system.

What is Metadata?

In its publication *Forum Guide to Metadata: The Meaning Behind Education Data*, the National Forum on Education Statistics (The Forum) defines metadata as follows: “Metadata are structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource... a robust metadata system improves the accuracy of data use and interpretation, as well as the efficiency of data access, transfer, and storage.”

The guide goes on to state:

In general terms, a robust metadata system will include system governance arrangements that include policies and procedures for metadata management and use within the organization, and related roles and responsibilities for staff; a metadata model that links metadata items to existing data elements and data sets; a list of relevant metadata items (i.e., a metadata item inventory), including a lexicon that identifies shared vocabulary for term use and naming conventions; and a comprehensive data dictionary. (p. 11-12)



Because metadata is what drives a tool like a data dictionary, it is helpful to clarify at the outset what types of items fall under the concept “metadata.” The Forum breaks these items into three categories: data management, technical, and data reporting/use.

- *Common data management metadata items:* element name, definition, purpose/mandate, restrictions, related data elements/components, calculations/formulas, manipulation rules, ownership/stewardship, effective dates, retention period, business rule, security/confidentiality
- *Common technical metadata items:* field length, element type, permitted values, code set, translations, storage/archival destination, data source, data target
- *Common data reporting/use metadata items:* routine use, key words, quality metrics

What is a Data Dictionary?

According to the *Forum Guide*, “a data dictionary is an agreed-upon set of clearly and consistently defined elements, definitions, and attributes—and is indispensable to any information system...Data dictionaries generally contain only some of the metadata necessary for understanding and navigating data elements and databases and, thus, contain only a subset of the metadata found in a robust metadata system” (p. 15).

When asked about the purpose and expected outcomes of their state education data dictionary, a data lead from Colorado provided the following explanation:

The goal is to create and provide a comprehensive information catalog of data definitions, relationships, collection groupings, validation rules, aggregations, and generated reports. The added benefit of a Data Dictionary will be to improve the accuracy of information and standardize data definitions within a centralized repository.

As large-scale data initiatives proliferate in the American educational system, states, districts, and other organizations are recognizing the critical need for a comprehensive data dictionary—and the metadata that drive it—to facilitate communication among varied data stewards and users and to allow for the accurate transfer of data from one entity or educational level to another.

Status of Data Dictionaries in Selected States

The central goal of this paper is to provide a direct, hands-on perspective on the development, maintenance, and philosophies behind state data dictionaries. To this end, interviews were conducted with data team members in five states: Colorado, Maine, Montana, Oregon, and Washington. These states represent various stages of data dictionary development and implementation, and the information culled from the interviews offers a range of experiences and lessons learned.

Interviews addressed the current status of the data dictionary, including where and how it is housed, who “owns” it and is responsible for its maintenance, and what procedures are in place to support it. Interviewees were also asked about challenges and successes they have experienced, their relationships (if any) with data dictionary vendors, and their overall goals and expectations for the dictionary. Finally, state representatives were asked to offer guidance or recommended practices to other states as they develop, implement, and maintain their data dictionaries. (See Appendix A for Interview Guide.)



Development and Design of Data Dictionary

Interviewed data leaders described different experiences regarding the development of their data dictionaries. Each state has followed a unique path, influenced both by the initial impetus for dictionary creation and by choices made regarding internal versus external development (e.g., the involvement of a vendor). This section provides a brief summary of each state's development process, as well as the overall design of each data dictionary.

In-house Development: Oregon and Colorado

Development of **Oregon's** data dictionary began around 2000, as the data team worked with the Oregon Student Record to identify a set of elements needed for student transfer data. This process led to the state's Consolidated Student File Format, which was a technical document that described the specific layout required to exchange information with the state's Department of Education.

The dictionary initially included about 120 elements, with different parts of the data system being populated depending on which data collection a user was referencing or updating. As the data team continued to develop the data dictionary, they focused on expanding the metadata and the ways it would drive the collection system. They worked to specify standards, including field names, length, and types. Standards became tighter over time as the team created the state's data map and leveraged the information they had to make decisions about metadata.

Oregon currently uses a Microsoft SQL 2008 server to house their dictionary, and will soon be moving to the SQL 2012 system. The data dictionary is a single entity, and the interviewed data lead stresses that the true source is the metadata. Rather than seeing the dictionary as a static, stand-alone tool, he stated, they think of it as a report from the metadata. Because the metadata drives the process, the data elements are thus continually revised and updated. In sum, the state sees the dictionary not as a tool that supports the data system, but as a product of it.¹

In **Colorado**, creation of the data dictionary was spurred by legislative mandate in 2007. Colorado House Bill 07-1320, also known as the Data Reporting and Technology Act, was designed to improve and streamline the education data collection process within the state. Included in the bill was a requirement that the Colorado Department of Education develop a data dictionary:

The Department shall develop and distribute to the school districts and public schools a data dictionary to define the data the department will collect and the methods and protocols by which school districts and public schools will submit the data. At a minimum, the data dictionary shall include the following items:

1. *A map of the current data collection requirements, including the definition of each data element, when each data element is collected, identification of the external reports for which each data element is used, and identification of the method by which each data element is collected;*
2. *A description of the format for data submission, acceptable values in data submission, the available options for dealing with data fields for which the submitting school district or public school does not have information, and logical comparisons to prior reports;*
3. *Identification of data relationships;*
4. *Data element tables; and*
5. *Identification of data element locations within data access tools.*

¹ Oregon's data dictionary files are available at <http://www.ode.state.or.us/search/results/?id=349>.



This legislation thus drove and provided funding for the development of the dictionary. It also allowed the department flexibility in the development process and the actual dictionary structure. Interviewed data leaders stated that they began with a more technical tool, and later shifted the focus more to the business side. They focused on the data collection process, evaluating what needed to be collected and whether the actual collections were effective. Because the development team felt that none of the available vendor products suited their needs, they chose to keep the dictionary development in house.

Colorado currently uses an Oracle application server, and expects to phase into Web Logic in the near future. Interviewed data leads noted that the tool they developed was “over-engineered” in the beginning, as they attempted to accommodate a wide range of elements and changes within an overcomplicated schema. Realizing that this model was too difficult to program against, the team created a significantly simpler model that is now a mere five tables—yet is flexible enough to accommodate their needs.

The state does, however, maintain three versions of its data dictionary: a public version maintained by the Data Dictionary Analyst, a version for “power users” of the data warehouse, and a third version currently being created as part of the SLDS project. At this time, there is no synchronization between the business and the technical dictionaries, and they are maintained by different teams.

Vendor Development: Maine, Montana, and Washington

In **Maine**, development of the data dictionary began in 2006. The state contracted a vendor to complete a data sources inventory, which focused on two main areas:

1. What data is being collected, for what purpose, and where is it then housed?
2. What specific data elements are within the collections?

The inventory revealed 236 separate collections and 745 unique metadata elements. (It should be noted that they defined “collection” as any request for data from the LEAs, and thus included federal reporting, varied state collections, surveys, and any other medium through which the LEA was asked to submit data to the state level.) They also found that they had 133 data repositories, which was equivalent to one repository for each person in the agency. “We realized we had a problem,” stated the interviewed data lead. The state developed a data management team, which worked with the vendor to put all of this information into its metadata inventory tool. They evaluated the timing of collections, related data stewards, element definitions and metadata, and data mapping. They created a data dictionary that was initially considered useful, but the interviewee acknowledged that the state has since struggled to keep it up-to-date—there has not been enough impetus to force data stewards to keep it current. More recently, the state’s new education commissioner has placed particular focus on the data dictionary and the data collection process, working with district superintendents to solicit their input and improve the tool.

In terms of design, Maine’s data team made the decision to keep the data dictionary as a business tool. As the interviewee stated, “only a SQL programmer could have understood the early ‘techie’ version that we had.” They have since focused on maintaining a dictionary targeted towards the business side that also includes enough technical information and business rules such that users, vendors, and managers can all understand system protocols.

Montana began its path to a comprehensive data dictionary with the basic dictionary attached to the state’s student information system. Interviewed data team members stated that the original system was mainly simple spreadsheets and basic elements. More recently, the state included a requirement for a data dictionary in their RFP for a data warehouse and state longitudinal data system (SLDS). Their vendor has done the initial data element and metadata load, and the dictionary is expected to be completed in the upcoming months.

Montana maintains its dictionary as a single entity, with the ‘techie stuff’ included in the design. Its reach is fairly broad, and the state plans to include all of the data categories featured in the SLDS. The focus is currently K-12, but will be expanding to K-20.



When **Washington** applied for its SLDS grant (which was awarded in 2009), they included a data dictionary tool in their proposed plan. They also included a requirement for enterprise architecture, and they were able to find a vendor that could provide a tool to handle both issues. It provides a framework onto which other modules can be added, and also allows a user choices in data management and reporting.

Washington's data dictionary is a single entity, but is designed in such a way that different users and program offices can choose to view only the sections relevant to their work (thus making it seem as though there are multiple, targeted dictionaries). The data team worked within the state's context to develop a user interface that is as simple as possible. The metadata underlying the system, however, was described by the interviewees as complex. They are building the data model through a combination of taxonomies, linking the elements to the taxonomies, and then defining each element. Mapping the data in this manner has allowed them to find issues with the element definitions (e.g., lack of consistency or alignment) as they occur, and then mark them in the proprietary tool for later adjustment. At this point, the dictionary is not as big as the team would like it to be: it includes the primary student data collection, some detailed special education data, and some other limited data points, but the tool allows for a great deal more complexity.

Experiences with Vendors

As highlighted above, three of the states interviewed worked with vendors to establish their data dictionaries. These arrangements range in complexity. In Maine, for example, the vendor did an extensive data sources inventory prior to recommending and implementing a proprietary solution. Because the state was one of the first to adopt this tool, they have maintained a relationship with the vendor, who has provided continued support and updated trainings, as well as hosting the data solution.

Montana and Washington each approached the vendor question hoping for an effective “out-of-the-box” solution, after doing an internal needs assessment (coincidentally, Montana also ultimately chose the same product used in Maine). Because Washington sought a solution that provided both a data dictionary tool and enterprise architecture, they chose a different vendor solution.

Overall, these states describe their relationships with their vendors as positive and are mostly satisfied with the products they have provided. Because Maine has maintained an ongoing relationship with their vendor, they have found that the vendor has met their needs better than expected: they created a symbiotic relationship wherein the vendor has been able to improve upon the product by working with the state's data, while providing the state ongoing technical support and training.

Montana uses the same vendor solution as Maine, but they have found that it has some limitations. They note that limited financial resources meant that their choice of metadata tools had to be a bit more basic than some of the “Cadillac” tools that are available. Because of this, they have found their chosen solution to be labor intensive, requiring them to still do some of the tasks they would like the technology to handle. For example, they cannot execute business rules from their data dictionary solution, instead having to manually document where these exist. Though the organization's vision included a product that would incorporate this function, their selected product's limitations have required them to develop it as a separate protocol.

Washington states that its vendor solution met 95 percent of its needs out of the box, with customization needed for the final 5 percent. They note that the product is both affordable and flexible, so making adjustments and updates is relatively easy. The product has thus been functional for their needs, though they state that they have not yet done all they would like with it: the product has additional capabilities that they plan to employ in later stages of the project. Thus far, they have mainly worked with the vendor on the development side, but state that the vendor has been pretty supportive of their needs as they have implemented the product.



Ownership

Interviewed states vary somewhat in their perspectives on, and plans for, ownership of the data dictionary, but all agree that keeping the data updated and relevant is a responsibility shared by data stewards collectively. For example, Colorado’s legislative mandate created a Data Dictionary Analyst position, whose role specifically includes maintenance of the metadata. However, interviewed data leads stressed that the data dictionary should be “owned” by all data collectors: coordinators, stewards, program officers, etc.

In Montana, an assigned member of the Measurement and Accounting Division has specific responsibilities for maintaining the data, but this individual is also a part of the larger core data stewards committee, which makes final decisions on definitions and other data element questions. This committee includes members from each business area.

In Washington, the two data dictionary leaders see their roles—respectively—as “custodian” and “salesman” for the dictionary. The custodian maintains the ultimate quality and accuracy of the dictionary, while the salesman works as a liaison with the program areas, advocating the need for the dictionary and the importance of data stewards actively keeping their particular groups of elements up-to-date and accurate. Thus, these data leads are providers of the tool, while data stewards are also expected to own their data.

Maintenance and Updating of the Data Dictionary

Each state interviewed has procedures in place for the updating and maintenance of the data dictionary. These vary based on the states’ data collection schedules, perspectives on ownership, and number of data stewards involved.

In **Oregon**, data elements and specifications are reviewed by the data governance committee on an annual basis. As each data collection has a data owner assigned, this annual process allows those involved to provide feedback on the associated metadata before it is discussed and—if necessary—updated by the committee. Beyond this annual review, changes may be made during the year if there are critical changes to federal requirements. Overall, however, the state tries to avoid these individual changes, as there is concern that districts would not have adequate time to make system changes and send needed data up to the state level.

Several of the other interviewed states address changes on a monthly basis. In **Montana**, the data governance group meets monthly to approve or deny changes requested by stewards of various data collections. Similarly, the Educational Data Advisory Committee (EDAC) in **Colorado** holds meetings on a monthly basis, and these meetings are seen as an opportunity to address any issues or updates to different collections (each data collection has an official annual review). Data stewards—all of whom are invited to these monthly meetings—are encouraged to visit the data dictionary site prior to the meetings and identify potential changes or updates.

Washington also holds monthly meetings. Interviewed data team members noted that the state sees the maintenance of the data dictionary as a very collaborative process. The timing of actual changes is dependent on the collection in question: for example, the data manual for the annual student collection (which is generated from the data dictionary tool) is updated and released in late February, while changes to fiscal collections—because they are dependent on legislative mandates—happen in August or September. The state’s representatives stressed that what is critical is to have a regular, predictable schedule for updates throughout the year.

Overall, interviewed data leads agreed that having a clear, continually evolving data governance process has improved the quality and accuracy of their data dictionaries, reducing redundancies and increasing the portability of data. They also noted, however, that maintenance of the data dictionary can be a continuing struggle for data stewards due to competing time and resource priorities. In addition, several of those interviewed emphasized the need for careful consideration of any changes, as a single



change can affect multiple collections and users. As one Colorado official stated, “we have to realize that any changes to collections have a ripple effect” and can have significant impact on LEAs.

Designing a Data Dictionary: Key Components

Beyond the design of their own state’s tools, interviewed state data leads were asked to discuss what they would consider key elements in an educational data dictionary. In addition to those items common to all dictionaries—element definitions, code sets, business rules, etc.—respondents offered several other specific suggestions:

- *Traceability*: knowledge of every place (e.g., collections, reports) that particular data appears
- *Archive functions*: access to old definitions and evolution of changes
- *Automatic synchronization with the data warehouse*: consistent alignment between the elements of the data dictionary and the other parts of the larger data system
- *Intentional, thoughtful design*: interface and presentation that is based on the state’s context and the needs of the end users
- *Data quality checks*: automatic, regular checks of data accuracy and completeness built into the maintenance processes for the dictionary (with clarity regarding the timing of these checks and the parties responsible for them)

Challenges

Interviewed data leads were asked to summarize the challenges that they have experienced regarding their own data dictionaries, as well as those challenges that may be inherent in the development and maintenance of such tools.

Keeping the Dictionary Current

States at various stages of the data dictionary process have experienced similar challenges in consistent maintenance and updating of data elements and metadata. In some states, this has been due to unclear expectations regarding roles—that is, whose job it is to keep things updated. Even in states where program offices are specifically tasked with maintaining their data elements, there is often no impetus for stewards to stay on top of the data. In addition, these stewards are often faced with many competing priorities, and updates to the dictionary may be put on the back burner. When the dictionary is not up-to-date, the tool is not useful, and people do not use it.

Oregon combats this issue by using the metadata to drive the data dictionary. Their data lead stated, “It is critical that the metadata is used transactionally. If we didn’t actually use it, we could not know things are accurate. By using the metadata throughout the system, we know if something is wrong or the system is not working.” Rather than seeing the data dictionary as a static tool that supports data initiatives (as is the case in some states), the team in Oregon perceives their dictionary as an evolving tool that is driven by the metadata and the actual processes undertaken by data stewards. “We see the data dictionary as a report from the metadata.”

Beyond this, interviewed data team members agree that making the business case for the use of the data dictionary increases the buy-in of data stewards and program officers. Simply put, users must be able to see how the dictionary will benefit them and make their jobs easier. When they do, they will use it and be invested in its accuracy.

Lack of Resources

Many states struggle with a lack of adequate resources to develop and maintain the data dictionary. This is especially true in smaller states, which often have fewer SEA staff available to focus on the dictionary. In Maine, for example, there is not a specific



education technical team—their IT staff is the overall state IT. This limits their availability to the Department of Education, and the SEA thus typically purchases applications rather than developing them in-house.

Resources can also come into question when the data dictionary is introduced as part of a particular grant or initiative. When that funding is no longer available, will the state be able to maintain the dictionary? Overall funding cuts in many states may exacerbate this problem. When funding is reduced, staff is sometimes phased out.

Existing Frameworks Cannot Support Expanding Needs

In some states, the initial design and functionality of the data dictionary no longer support the needs of the department as the number of end users increases or the complexity of the data grows. This may be due to lack of long-term vision at the beginning of the project (that is, the state did not realize how large their needs would turn out to be). Limited financial resources are often a factor as well—states may buy the vendor solution that fits their budget, even if it does not meet all of their needs. In these cases, states must decide whether to find other ways to solve data problems (or find a “work-around”) or to seek greater financial support in order to implement a more appropriate tool.

Hesitation to Align with Early Versions of Federal Standards

Previous investment of time and resources into data initiatives has caused some data stewards and end users to be cautious of large-scale data dictionary efforts and the work that must be invested in them. For example, Maine’s data lead described having aligned to early versions of various federal standards, followed by the need to invest additional efforts each time those standards are updated. While standards must continually evolve to meet the needs of federal and state agencies in order to remain effective, states with limited resources may feel hesitant to commit those resources to early versions. These states may instead choose to wait to align their dictionaries until later, more expanded versions become available. Though this may be understandable in light of states’ concerns about where to devote staff time, it can also preclude a more gradual integration of dictionaries and metadata (thus creating a larger, more onerous project in the end). In addition, it keeps those states from being involved in the federal versioning process, when their input might have been an asset to the developing standards.

Data Dictionaries, ED*Facts*, and CEDS

States have found greater need for data dictionaries as their data initiatives and requirements increase in number and complexity. For example, the data collected and reported for ED*Facts* requires a clear understanding by all users of element definitions, reporting conventions, etc. Having a state-wide data dictionary employed by the LEAs providing information to the SEA, as well as the data team at the SEA itself, means that varied users will be speaking the same language and the data reported to ED*Facts* will be more accurate and of higher quality.

As more states begin using the CEDS tools (i.e., Align and Connect), consistency and comparability of data across states can increase as well. Thus, having a state data dictionary that can be aligned with CEDS allows a state to assess how in line its data definitions and elements are with other states (and CEDS itself). If all states have dictionaries that can be aligned with CEDS, this will only improve the accuracy and consistency of data reported to ED*Facts*, thus making ED*Facts* more useful and relevant overall.

Interviews with the state data team leaders addressed the possible connections between CEDS and ED*Facts* and how data dictionaries relate to these initiatives. Questions addressed the states’ usage, if any, of the CEDS tools, as well as where their ED*Facts* data elements fall in their data dictionary projects. (See Appendix A for Interview Guide.)



Common Education Data Standards (CEDS)

Version 3 of the Common Education Data Standards was released in January 2013. In addition to the elements, definitions, option sets, etc. that are included in the standard, CEDS includes tools designed to help LEAs, SEAs, and other organizations align their data dictionaries to CEDS and see the connections between their initiatives and those of other entities.

CEDS Align is a web-based tool that enables users to import or input their data dictionaries, align their elements to CEDS, compare their data dictionaries with those of other users, and analyze their data in relation to various other CEDS alignments. CEDS Align is designed to assist education stakeholders in sharing data among educational sectors and across state lines.

CEDS Connect builds upon the CEDS Align tool and allows stakeholders to apply specific and relevant maps to a growing pool of CEDS “connections.” Stakeholders from varied educational organizations can use the tool to identify policy questions and related data elements, define analytic approaches, calculate metrics and indicators, address reporting requirements, etc. CEDS Connect enables users at different levels to consider the metric definitions of data points such as graduation rate, program enrollment, or academic outcomes. By establishing the data elements necessary to answer a given question, as well as recommended logic and routines for analysis, CEDS Connect is designed to help the education data community work together towards standard definitions and methodologies that will provide common, comparable data measurements and reporting that can cross districts, states, and educational agencies.

Interviewed state representatives discussed their states’ involvement with CEDS and the level to which their agencies have worked with the CEDS tools. Oregon, Colorado, and Washington all described specific efforts to incorporate CEDS into their data discussions and dictionary initiatives. Oregon’s data team has submitted comments to the CEDS team as the different versions have been developed, and several members of the team have access to Align and Connect. Their data governance group has officially adopted CEDS, and their P-20 Operational Data Store is based on the standard. They are currently doing a crosswalk of option sets, and hope to fully migrate their system to the standard within the year. Thus far, the state has found a significant difference in naming conventions—their system was designed with shorter field names than CEDS, which does not allow them to be very descriptive.

Colorado’s data team leaders describe being “completely on board with the idea of national standards.” They initially benchmarked their data definitions based on existing state statutory, current legislation, and state board rules, and then referred to Ed-Fi. They have now aligned with CEDS, and are part of the Multi-State Data Exchange, which includes CEDS as part of its strategy.

Washington reports using both Align and Connect. They use their vendor solution as an interface, which allows them to take their data, export it, and load it into the CEDS tools. Their Align map includes some of their directory and student identifying information, and they are working to import the elements from their student information system (SIS). They note that their SLDS and data warehouse were each created in alignment with CEDS Version 2.

Montana and Maine have not yet done as much work with CEDS as the other three interviewed states, though both report plans to expand their usage of the standard. Montana’s business analysts are currently evaluating the CEDS tools, and the state plans to incorporate CEDS in their upcoming new high school transcript standard. In addition, their data dictionary lead has submitted comments to the CEDS team as the different versions have been released.



Maine used CEDS as a standard when developing their state RFPs for vendors: they required a crosswalk, rather than asking vendors to change existing products (they feared the latter could be cost-prohibitive). The state’s data lead notes that their vendor is planning to build a CEDS crosswalk into its metadata inventory tool.

EDFacts

Each of the state representatives interviewed agreed that their data dictionary efforts have been integrated into their *EDFacts* collections. In some cases, working with these in conjunction has allowed them to identify problems in their *EDFacts* processes that might otherwise have not emerged. In Montana, for example, the team has entered all *EDFacts* elements into the data dictionary—and has since identified misalignment in numbering conventions on birthdates. Similarly, as Washington’s data team aligned their data dictionary with *EDFacts*, they found that the state had not been reporting magnet schools in File 129 (CCD), though some local districts do have them.

Though integrating the data dictionaries and the *EDFacts* collection has been useful, interviewed states note that the process can still be improved. Oregon reports that the *EDFacts* collection is still a fairly manual process. Similarly, Colorado’s data leaders explained that *EDFacts* requires them to locate and report particular data, but the elements pulled from the data warehouse and reported to *EDFacts* are not automatically connected to the dictionary’s metadata, such as specific tables where those data appear in other reports or other uses of those elements within the data system. They state that they still need a way to connect the *EDFacts* collection and reporting process back to the data dictionary’s technical side. Though the state is shifting its focus from a more tech-based data dictionary to one aimed at the business user, the interviewees note that the *EDFacts* collection requires both of these sides. They stated that they need to add a field that provides the physical location of each data point in order to link the different systems together and have a better picture of their integration. At present, they do not have means to automatically make these connections.

These selected states are optimistic about the possibilities for *EDFacts*, however. When an *EDFacts* team visited Maine last year, they advised state officials to use their SLDS to do their *EDFacts* collection and reporting. By automating these processes, they will take pressure off the program managers assigned to the data. In addition, they are able to track multiple years of submissions to get a fuller picture of the data.

Recommended Practices/Guidance for States

The states highlighted in this paper are at different points in their development, implementation, and maintenance of their data dictionaries. Similarly, the size and complexity of their dictionaries is dependent upon each state’s unique needs. Overall, however, the state representatives were in consensus regarding much of the guidance they would offer other states as they embark upon their data dictionary initiatives, or continue those efforts already underway. This section summarizes the key guidance points provided by the interviewed states.

Focus on Data Governance and Data Ownership

There must be a clear data governance process that everyone understands, and that includes representation from LEAs. A mature governance process is predictable and standard, allowing all involved to play their necessary roles.² Some states interviewed, in fact, have found that working on the data dictionary actually catalyzed the governance process, highlighting weaknesses in the data system and allowing them to reduce the number of collections, redundancies, and data repositories. Washington’s representatives noted that moving from a one-off aggregate collection (in their federal reporting) to using student data to populate reports has

² An example of a comprehensive data collection schedule is available on Oregon’s website at <https://district.ode.state.or.us/apps/info/>.



made the process and the reported data more consistent: they are all pulling from one source, and working in this way has allowed the agency to solidify business rules surrounding the data.

The governance process should include clearly defined ownership of collections, elements, due dates, etc. Though the process must be collaborative and an interactive effort between IT and business, there should not be questions as to which individuals are responsible parties.

Seek Support from Above

The states that have seen the most success appear to be the ones that have support from higher up—whether in the Department of Education or the larger state government. For example, Colorado’s data dictionary emerged as part of a legislative mandate. Because there were individuals at the legislative level who understood the importance of these data initiatives, the SEA received needed funding and flexibility to create a sustainable solution. In Maine, the new commissioner has been particularly supportive of data efforts (including the dictionary), working with district superintendents to gain their insights and improve the overall process. States that do not already have this kind of support should consider bringing their case to departmental or legislative leaders, to impress upon them the need for tools (such as data dictionaries) that increase data accuracy, quality, and portability. It is likely that many of these leaders would support these data efforts if they had a greater understanding of the critical need for them, as well as the informational improvements possible under a robust data system.

Do Your Due Diligence

A comprehensive needs assessment—prior to any steps forward in development—allows a state to consider its data needs fully and create a plan that will not fall short of the particular requirements of the agency. Due diligence regarding vendors is critical as well. States need to consider the customizability of any potential solution, especially when considering the long term. Will this product suffice not just now, but as the needs of the project and the number of end users continue to grow? States need to consider both the basic needs of the initiative *and* the different phases of the project. One interviewee stresses that data stewards need to “do their homework”—know that there will be a need for flexibility, but nail down as much as you can before moving forward.

See the Data Dictionary as a Process, Not a Product

States whose dictionaries can expand and adjust based on changing needs and context tend to view the dictionary not as a static product that is created and thus “finished,” but as a continually evolving process from which a malleable tool can emerge. Similarly, they see data collection as a life cycle, not as a series of independent events. Users who are on the ground, using the data day-to-day, will have an understanding of why the metadata is critical and will inherently see the value in keeping the dictionary up-to-date. In short, “the data dictionary cannot be a reference document.”³

Know That It Will Take Time to Do It Right

Beyond the initial needs assessment and due diligence, states should recognize that the development process overall must be handled carefully. Interviewed states found that the process took longer than they had originally expected, due to the sheer amount of data involved. One state representative noted that the tech side was much easier—the harder part was sifting through all of the data, agreeing on common definitions, and creating new elements when necessary. He noted, “It’s easy to throw something together quickly. It has to be a priority to get it done right.”

³ Oregon Department of Education Information Services official, personal communication, February 20, 2013.



Suggested Resources

Interviewed data team members were asked to suggest any resources that they found helpful during their development process.

- *Data dictionaries from other industries*: as they developed their education data dictionaries, several states found it helpful to consult those from other industries, such as telecom and health care. These dictionaries allowed them to consider what they might want theirs to look like, how complex their definitions should be, etc.
- *The Forum Guide to Metadata*: this guide, referenced earlier in this paper, was also a helpful source for states as they worked to create data models, make decisions about business rules, and design their dictionaries.
- *Gartner Magic Quadrant*: this tool uses a set of evaluative criteria and a graphical treatment (the quadrant) to assist users in choosing technology providers and products for different projects.
(see http://www.gartner.com/technology/research/methodologies/research_mq.jsp for more information)

White Paper Sources

AHIMA. "Managing a Data Dictionary." *Journal of AHIMA* 83, no.1 (January 2012): 48-52.

Data Quality Campaign. (2007). *How Can My State Benefit from an Educational Data Warehouse?* Washington, DC: Data Quality Campaign.

National Forum on Education Statistics. (2009). *Forum Guide to Metadata: The Meaning Behind Education Data* (NFES 2009—805). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

This project has been funded at least in part with Federal funds from the U.S. Department of Education under contract number ED-CFO-10-0084/0001. The content of this publication does not necessarily reflect the views or policies of the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.



Initial Development of the Data Dictionary(ies)

1. When did you begin development of the data dictionary (DD)? What were the circumstances that brought about its creation? (i.e., was it tied to a particular project or initiative, or did the data team decide to develop it separately?)
2. Was the DD developed in-house, or did you use a vendor? (If it was a vendor, who, and what was the process?)

Maintenance/Updating of Data Dictionary(ies)

3. Where is the DD housed?
4. Who “owns” the DD? Who is responsible for updating it? (Are these seen as the same or separate issues?)
5. What is the role of the data stewards in defining items in the data dictionaries (e.g., migrant program staff defining migrant items)? How are they engaged in the process of developing the DD?
6. What procedures or protocols are in place to support the DD?
7. How many people are involved in the maintenance of the system?
8. What technology underlies the system?
9. What strategies do you employ to support an increasing number of end users?

Overall Nature of the Data Dictionary(ies)

10. How complex is your DD?
11. Some states have described maintaining different versions for different audiences (e.g., a more “techie” version vs. a public version), or simply having multiple dictionaries for different uses. Do you have multiple versions of your DD? If so, how has this affected the process? What are the differences in their content and use?
12. What components/elements are included in the DD?
13. What features do you consider essential, not just in your DD but in an effective DD in general?
14. How would you describe the actual use of the DD? What outcomes are expected?

Experiences with Vendors (inclusion based on answer to #2)

1. How would you describe your overall experience with your vendor? (How responsive have they been to your needs? Has the relationship been collaborative, or did they come to you with an “out of the box” solution?)
2. Are there particular challenges that come with using a vendor product? (If so, is it still preferable to developing the DD in house?)
3. If you are using a vendor’s proprietary product, how does that affect your DD when the contract is up? What happens at that point?

Common Education Data Standards (CEDS) and Data Dictionaries

4. How does CEDS fit into your overall work at this time?
5. Do you use the various CEDS tools (Align, Connect), and if so, how? Please describe your experiences (or how you plan to use them).
6. Does CEDS help you present the public face of your DD?



EDFacts and Data Dictionaries

7. How has the use of a data dictionary affected your EDFacts collection and reporting? Has incorporating a DD across multiple users been beneficial?
8. How does your SEA handle unit record to aggregate reporting metrics development?

Challenges/Successes/Guidance

9. What have the main challenges been in developing or maintaining your DD?
10. Where do you feel you had the greatest success?
11. If you were to offer guidance to other SEAs based on your experiences, what would that be?
12. Are there particular resources regarding DD that have been helpful to you, and that you would recommend to other states?
13. What other comments would you like to add in regards to data dictionaries?



Appendix B: Publicly Available State Data Dictionaries

The list below provides links to a selection of publicly available state education data dictionaries. Please note that this is a sample of states and is not comprehensive.

Colorado: <https://cdeapps.cde.state.co.us/DataDictionary/>

Maine: http://dw.education.maine.gov/DirectoryManager/Web/Maine_report/MaineLanding.aspx

New Hampshire: <https://my.doe.nh.gov/datadictionary/collections.aspx>

Oregon: <http://www.ode.state.or.us/search/results/?id=349>

