

# **Protection and Accessibility of State Student Unit Record Data Systems at the Postsecondary Level**

Molly Ott  
Stephen DesJardins  
University of Michigan

*under subcontract to the  
State Higher Education Executive Officers*

## **INTRODUCTION**

As the use of unit record data to track students through their educational careers has increased over time, interest in improving both the accessibility and security of the data has grown as well. The purpose of this study, *Protection and Accessibility of State Student Unit Record Data Systems*, was to survey the practices used by the state agencies responsible for this data in ensuring the confidentiality of student unit record (SUR) information while also making data accessible to those who request it. The *Protection and Accessibility* report attends to the procedures used to make data available to stakeholders (e.g., policy analysts, agency staff, education researchers, auditors, and others) as well as identifies the stakeholders to whom data are typically made available. This study is a companion to the larger *State of State Data Systems* study sponsored by the National Center for Education Statistics (NCES) and conducted by the State Higher Education Executive Officers (SHEEO).

The *Protection and Accessibility* report begins with a description of the methodology utilized to collect relevant data and information from the SHEEO agencies. Then, a total of seven sections highlight major findings with regard to the accessibility and security of SUR data. These include considerations of who is requesting SUR data, how agencies respond to these requests, the problems and challenges surrounding data requests, and key agency, state, and federal policies that pertain to SUR access and protection. The final section discusses best practices surrounding data security and confidentiality, as well as the overall management of SUR data.

## **METHODOLOGY**

The proposal for the *Protection and Accessibility* report originally assumed that data would be collected using the same methods as the SHEEO *State of the State Data Systems* report, i.e., an internet-based survey. As the study design progressed, however, the research team decided to use telephone interviews. The main reason for this decision was the challenge in formulating and phrasing questions, as well as response options, in a manner that would be appropriate for each agency's SUR design and processes. No prior census-type studies of SUR protection and access have been conducted, so there was a limited knowledge base on which to frame the question protocol and ensure that it would be answerable for all agencies.

An interview protocol was therefore developed by the research team in consultation with project representatives from SHEEO as well as selected contacts from SHEEO agencies. Questions that were part of the original report proposal but had parallels in the online survey that was simultaneously conducted by SHEEO were removed from the protocol to prevent duplication of efforts. The final protocol for the interviews may be found in Appendix A (p. 24). In a few cases (noted in Appendix C, p. 27), the respondent requested to provide information through email rather than through a telephone conversation. The same questions used for the telephone interviews were sent to these respondents, and they typed their answers into the document and returned them to the researchers. Also, in

addition to the telephone or emailed questions, a number of agencies provided the researchers with additional policy documents or website references containing policy information via email.

Representatives from each of the 57 state agencies identified by SHEEO and provided to the Michigan research team were contacted during July and August 2009 requesting an interview. The text of the contact email may be found in Appendix B (p. 26). Also, a final set of follow-up emails were sent to non-respondents in September 2009. A total of 40 interviews were conducted with agency representatives, which represents 70 percent of those included in the sample. The full list of agencies from which information was collected may be found in Appendix C (p. 27). On average, each interview lasted approximately 30 minutes, although conversations ranged from 25 to 50 minutes. In addition, two 60 minute telephone interviews were conducted with individuals identified by the research team as having special expertise in the area of SUR data in regards to responding to challenges and best practices. Notes from each of the telephone interviews were analyzed by the research team and in addition, complementary document analyses were conducted with regard to the policy documents and website information provided by the interviewees. The remainder of this report considers in more detail the themes that emerged across the agencies' policies and practices pertaining to the management of student unit record data systems.

## **RESULTS**

### **GENERAL FINDINGS**

The diversity of SUR systems' purposes and technical structures clearly affects the access and security practices that are in place at each agency. The SHEEO agencies range widely in the length of time their postsecondary SUR systems have been in place and in the complexity of these systems. Some have had functioning postsecondary systems in place since the 1970s, while others are currently implementing or substantially upgrading systems. Similarly, the scope of data coverage, type of data collected, and technical structures vary considerably across the different agencies.

In terms of scope, a few of the agencies, such as the University of Wyoming, manage data from students at a single institution only. At the other end of the spectrum, some agencies incorporate both primary and secondary as well as postsecondary records in their single systems. Most SUR systems, however, represent more than one campus. For example, the University of Wisconsin SUR system includes all public four-year colleges and universities as well as public two-year liberal arts transfer institutions. However, neither the Wisconsin technical college system nor the private independent colleges and universities are a part of the system or the agency data. The length of time these systems have been in place similarly varies quite a bit. For example, the Florida Department of Education has had a fairly comprehensive kindergarten through postsecondary system for many years, whereas Washington, New Mexico and Pennsylvania are each in the process of implementing a P through 20 SUR system.

The specific data elements collected by each agency directly relate to their mission or purpose and often to their reporting role in their state. While many contain comprehensive student unit records, others maintain more limited data sets that are directly connected to a specific function. For instance, the Wyoming Community Colleges Commission receives data from seven community colleges. Most of the data that these institutions submit to the Commission is summarized, which the Commission then uses for its own reporting. The only SUR data that the Commission has is related to the state's Hathaway Scholarship program. Since 2006, each college has submitted student records to the Commission for reporting purposes about Scholarship applicants, amounts awarded, and student performance, which ultimately determines the financial reimbursement that the state provides to each college for Scholarship recipients. The Commission's SUR records are for the approximately 1,500 to 2,000 Hathaway-eligible students only, not the entire community college student population. So while the data are from multiple institutions, the unit record level data elements collected from each are limited to those necessary for Hathaway reporting. Most of the SHEEO agencies interviewed for this study, however, have more extensive coverage than does the Wyoming Community Colleges Commission and include all enrolled students at the institutions that are part of the given system as well as an array of student data elements, including admissions, financial aid, and academic performance.

Additionally, the technical structure of the extant systems varies considerably across states. Some unit record systems – including the University of Alaska system, Minnesota State Colleges and Universities, Vermont State Colleges – function on integrated Enterprise Resource Planning (ERP) platforms, with each feeder campus using the same database and software that synchronize together into one massive system. A few agencies maintain their unit record data on spreadsheets stored on secured servers. The majority use their own standalone database, such as Oracle, with query tools that permit authorized end users to run reports and analyze the data. In each case, the technical organization of the unit record management allows the agency to meet its primary mission (most often reporting to federal and state departments and agencies). While many respondents indicated a desire or plan to upgrade their systems in the future to more sophisticated database systems and/or to include more data elements, that they are currently quite diverse is an important factor in the way that the access and security policies and practices are implemented.

### **ACCESSIBILITY OF UNIT RECORD DATA**

Despite the diversity of SUR data systems in terms of establishment and complexity, a common practice found with regard to data access is that access to users other than internal office employees who are tasked with maintaining the data and/or generating regular reports is closely protected. When discussing “access to unit record data,” it is important to make a distinction between access that is granted directly to the unit record databases – in other words, creating a user ID and password that permits an individual to directly log into the system – versus access that is granted through data extracts downloaded from the system and provided to an individual. The former instance is rare; only three agencies mentioned giving log-in access to someone outside of their immediate

office. In two cases it was for a state auditor to conduct state sanctioned auditing-related business, and in one case (North Carolina Community Colleges), it was for a legislative researcher to conduct data matches. Extracts of unit record data that are then provided to state approved representatives occurs more frequently. In cases where auditors are concerned, full extracts of unit record data - rather than log-in access - are often provided as required by state statutes; examples include the Florida Department of Education and the Oregon University System. This section of the report, *Accessibility of Unit Record Data*, considers in detail who (outside of internal agency employees) is typically requesting and/or is given some form of unit record data, as well as the request processes in place at the different agencies.

### **Accessibility of Unit Record Data: Aggregate Data Requests**

In each agency, requests for unit record data are made in forms of either aggregate level data or unit record level data. All interviewees agreed that most requests for information about students in their systems are aggregate or reporting-type analyses that the offices themselves may conduct and then provide results to the requestor. These may be through Public Records or Freedom of Information requests or by general inquiries. Exactly who is requesting aggregated information varies from agency to agency. Almost all have formal reporting roles to the state government (i.e., governor and legislature; it is much less common to receive reporting requests from the courts) and therefore field requests from these stakeholders. Most agencies also receive requests from the state coordinating boards, K-12 district or state representatives, higher education institutional representatives, other state agencies such as departments of labor, the media, higher education faculty and researchers, and the general public. Also, many agencies are designated the responsibility to act as their state's representative for federal IPEDS reporting so aggregated information may be sent on a regular basis to the National Center for Education Statistics.

In terms of the process used when reports are requested, typically the requestor will make contact with an agency through email or telephone and pose a question. Agency staff responsible for research and/or data dissemination will then send published reports to the requestor or the requestor may be directed to a website containing the information and/or previously generated reports that can the question posed. If no such information or reports already exist, the agency will generate one to satisfy the request. The amount of time that it takes to fulfill these requests depends on the nature of the report and who is asking for the report. For instance, requests from the state capitols get high priority and are typically processed within a few days; requests from citizens take longer to process. Most of those interviewed mentioned that the current office workload (e.g., how many requests are in the "queue") is also a major factor in dictating how quickly they are able to respond to both reporting and unit record level extract requests.

Almost every agency maintains a website with static Adobe pdf or spreadsheet-type reports that are accessible to the public and updated on at least an annual basis. Several mentioned that in addition to these regular reports, they may also post reports created for requestors if they are deemed to be of interest to the broader public.

### **Accessibility of Unit Record Data: Unit Record Level Data Requests**

Compared to aggregated data requests, much more attention is paid to external requests for unit record level data because this type of data contains identifiable individual information. A theme that emerged throughout conversations with agency representatives was that careful consideration is given to the specific data elements that are requested and released to all of these external parties. The agencies work closely with requestors to determine exactly what elements are needed to achieve the purposes of the intended analysis. In some cases, requestors will assume that merely communicating a desire to do educational research would result in a full copy of the unit record database being provided to them. This is not the practice of any agency included in the study, however, and most of those interviewed noted that careful conversations occur about the intent and needs of the analysis as well as how these translate into the appropriate amount and type of data that can be provided.

Also, most of those interviewed noted that identifiers will not be included in a unit record extract unless the researchers explicitly demonstrate a need for them (e.g., to match with another data source). As discussed in the sections below, identifiers will never be included in a unit record extract provided to an external party. In terms of who may be approved to receive extracts of unit record level data from the agencies' systems, three main groups external to agency employees were mentioned most often: 1) third party researchers, 2) representatives of the postsecondary campuses that are part of the data system, and 3) representatives of other state or federal agencies. How each group is handled by each of the agencies is discussed here in more detail.

#### ***Third Party Researchers***

Among the agencies that do release unit records, the most often cited group to whom data is provided are third party researchers. These are researchers who are not direct employees of the agencies themselves, though they may be affiliated with the campuses that contribute data to the system (most commonly, professors or doctoral students conducting educational research). As Table 1 (p. 7) suggests, the majority of SHEEO agencies will consider and, at least in some cases, approve requests for unit record extracts, both de-identified and identifiable, from third party researchers on an individual basis. In every case where unit records are provided to external researchers, the agencies require some sort of documentation or contract (shown in the last column of Table 1). The *Data Request Process* section of this report (p. 13) discusses these documents in more detail.

The regularity of such requests varies across the responding agencies. At the high volume end of the spectrum, the Florida Department of Education receives enough requests to necessitate weekly meetings of their Education Data Warehouse committee to review, consider, and approve them. Many other agencies face far fewer third party researcher requests. A number of examples illustrate the infrequency by which unit record data extracts are provided to individuals or groups outside of the agency. West Virginia recalled giving a de-identified unit record extract to a researcher from Harvard a year ago as well as participating in the Lumina project a few years ago testing the feasibility of a nation-wide unit record database. Beyond those two instances, no unit record level data has been

released in recent memory. Similarly, in South Carolina, the unit record system has been in place approximately 15 years and the Commission representative believed that unit level data has been released about six times to independent researchers. Missouri currently has two groups to which unit record extracts are provided: a longstanding research contract with economists at the University of Missouri-Columbia, who have the data including personal identifiers, and researchers associated with a National Governor's Association grant that is piloting P through 20 data linkages, who have de-identified data. Where all of these examples are concerned, the agencies typically have created a Memorandum of Understanding or Data Sharing contract that designates the researchers as an agent acting on behalf of the agency.

Table 1 also indicates those agencies that will release de-identified data only. In Ohio, for example, the Board of Regents draws a distinction between de-identified data and those with a Social Security Number attached where student unit record data requests are concerned. No extracts of student data with a social security number are given to anyone, and that prohibition even extends to the institutions that submit data to the agency. For instance, if a community college wanted to do research on its out-transfers, the Board may only release that information on an aggregate basis; they cannot release to a college unit records pertaining to individual students. De-identified unit record level data requests mostly come from academic researchers, and Ohio receives between five and six requests per year.

Finally, as Table 1 indicates, seven of the 40 agencies do not release extracts of student unit record level data to third party researchers at all. Those agencies that infrequently or never release data to researchers offered two common explanations for why this is the case. One has a legal basis, in that some agencies are concerned about violating FERPA or state laws related to privacy and find that the "course of least resistance" is to not release data. Also, several speculated that third party researchers may not be aware of the existence and/or availability of their data and this lack of public knowledge likely contributes to the limited number of requests. A few of the systems that contain fewer data elements noted that the research options for using their data are rather limited, so the datasets may not be of interest to external scholars.

**Table 1.**  
**Unit Record Extracts to Third Party Researchers.**

<b>Agency</b>	<b>Extract</b>	<b>Required Documentation</b>
Alabama Commission on Higher Education	No extracts of unit record level data	N/A
University of Alaska Statewide System	Considered and approved on an individual basis (though requests are very infrequent)	IRB documentation Data Sharing Agreement
Arkansas Department of Higher Education	No extracts of unit record level data	N/A
California Community Colleges	Considered and approved on an individual basis	Requestor must submit formal written proposal outlining the research and the data requested
The California State University and Colleges	Considered and approved on an individual basis (in the past, researchers have only been affiliated with other CA public colleges or universities)	Memorandum of Understanding
University of California System	Considered on an individual basis and may be approved for de-identified data only	FERPA Researcher Agreement
Colorado Department of Higher Education	Considered on an individual basis and may be approved for de-identified data only	Researcher Contract
District of Columbia, Office of the State Superintendent	Considered and approved on an individual basis (has occurred one time)	Memorandum of Understanding
Board of Governors, State University System of Florida	Considered and approved on an individual basis	Data Sharing Agreement
Florida Department of Education	Considered and approved on an individual basis	Unit Record Data Request packet
Kansas Board of Regents	Considered and approved on an individual basis	Standardized data request documents
Council on Postsecondary Education (Kentucky)	Considered on an individual basis and may be approved for de-identified data only	Data Request packet including cover letter, detailed description of the data requested, signed data use agreement, and for student requests, a letter of reference from a faculty member or advisor
University of Maine System	No extracts of unit record level data	N/A

Maryland Higher Education Commission	Considered on an individual basis and may be approved for de-identified data (very few cases where identifiable data was provided)	Memorandum of Understanding
Minnesota Office of Higher Education	No extracts of unit record level data	N/A
Minnesota State Colleges and Universities	Considered and approved on an individual basis	Enter into a contract with the researcher
Missouri Department of Higher Education	Considered and approved on an individual basis	Memorandum of Understanding
Montana University System	Considered and approved on an individual basis	Memorandum of Understanding
Nevada System of Higher Education	Considered on an individual basis and may be approved for de-identified data (although this is very infrequent)	Memorandum of Understanding
New Mexico Higher Education Department	Considered and approved on an individual basis	Data Request form (in the process of being developed)
City University of New York	Considered and approved on an individual basis	Memorandum of Understanding
State University of New York	Considered on an individual basis and may be approved for de-identified data (only one case of identifiable)	Memorandum of Understanding
New Jersey Commission on Higher Education	No extracts of unit record level data	N/A
North Carolina Community College System	Considered and approved on an individual basis	Memorandum of Understanding
Ohio Board of Regents	Considered on an individual basis and may be approved for de-identified data only	Memorandum of Understanding
Oregon Department of Community Colleges and Workforce Development	Considered and approved on an individual basis	Data Sharing Agreement
Oregon University System	Considered on an individual basis from researchers affiliated with the Oregon University System and may be approved for de-identified data only	Memorandum of Understanding
Pennsylvania Department of Education Office of Postsecondary and Higher Education	Considered and approved on an individual basis	Data Request form
Pennsylvania State System of Higher Education	No extracts of unit record level data	N/A
South Carolina Commission on Higher Education	Considered and approved on an individual basis	Data Request form
South Dakota Board of Regents	Considered and approved on an individual basis	Memorandum of Understanding
Vermont State Colleges	Considered and approved on an individual basis by each college's IRB	IRB documentation

State Council of Higher Education for Virginia	Considered and approved on an individual basis	Depends on the circumstance as to what documentation is required
Washington Higher Education Coordinating Board	Considered and approved on an individual basis (though requests are very infrequent)	Depends on the circumstance as to what documentation is required
West Virginia Higher Education Policy Commission	Considered and approved on an individual basis	Data Usage Agreement
University of Wisconsin System	Considered and approved on an individual basis	Memorandum of Understanding
University of Wyoming	Considered and approved on an individual basis	Confidentiality agreement
Wyoming Community College Commission	No extracts of unit record level data	N/A
Utah State Board of Regents, Utah System of Higher Education	Considered from researchers affiliated with the agency or with one of the data providers only and approved on an individual basis	Depends on the circumstance as to what documentation is required

***Postsecondary Institutions***

In addition to third party researchers, representatives of the postsecondary campuses that are part of the data system often make a request for the unit level record data. One area around which agency practices are quite mixed is the availability of student unit records to the postsecondary institutions that contribute data to the system. As Table 2 (p. 10) indicates, a number of possibilities exist for authorized users on each campus: in four cases they may view the entire system’s data at the unit record level, in four cases they are given aggregated reports of the entire system data, in 13 cases they may view data for their own campus only at the unit record level, and in 19 cases they do not have any access (after correcting their own submissions).

Most commonly, institutions submit their data to agency offices in a one-way transmission. There is often a period of time that the submitter may view the data that were sent to ensure that the submission was successful and/or to correct errors. After the submission/correction time window closes, however, the campuses are typically not able to access the agency’s data, nor are they able to view any of the unit records in the master agency dataset, nor are they given any regular extracts of the data.

In twelve cases, authorized campus users are able to log-in to the agency’s “master” system and to view unit record data provided by their institution only. The Minnesota State Colleges and Universities is a typical example, in that while the student unit records for each of the 32 campuses are accessible through a common platform, authorized campus users are limited to view and update those records for students who are currently or in the past have been associated with their institution only. For instance, students who are cross enrolled can be viewed by all campuses that they attend or, if a student left that college and

went to another college, the campus user can view their transfer records or prior academic records.

A few agencies have multiple layers of access that may be available to institutional users; a notably comprehensive example is the University of California System (UC). Static aggregate tables for internal and public use are available on the UC website (<http://www.ucop.edu/ucophome/uwnews/stat/>). Also, the system has developed a searchable site open to the public called Statfinder, that allows users to access a variety of pre-made and customizable tables on UC student undergraduate admissions, enrollment, financial aid, persistence and graduation. (<http://statfinder.ucop.edu/>). A private version of the Statfinder site (<https://reststatfinder.ucop.edu>), with a few additional capabilities beyond those available in the public version, is available to approved end users who must submit a signed Usage Agreement (which is accessible from the website) before receiving a user name and password that allows access. In addition, authorized campus users are able to run searches from the Corporate Student System (<https://csg.ucop.edu/dwr/dwrLogin.jsp>). The operating practice for the public Statfinder, as well as many other agencies that open such tools to the public, is that the underlying data that supports the tool is in aggregated form. It is not a query tool by which the users access the actual unit record database. None of these reporting tools allow public access to unit record level data; all information is displayed in the aggregate only.

Note that Table 2 represents common practice and not “exceptions-to-the-rule.” For instance, the California State University (CSU) system typically does not provide data back to any of the 23 campuses that are included in the system, beyond allowing for error checks and re-submissions during the data exchanges. However, in the past, CSU campuses interested in tracking student transfers have entered into agreements with the agency, which then matches data by SSN and provides de-identified data back to the colleges for this use.

**Table 2.**  
**Unit Record Extracts and/or Access to Postsecondary Institutions.**

Alabama Commission on Higher Education	No access or regular extracts
University of Alaska Statewide System	View unit record data for the entire system (if have a business need as defined by job responsibilities)
Arkansas Department of Higher Education	No access or regular extracts
California Community Colleges	No access or regular extracts, although there is a website where the colleges can download their own data only
The California State University and Colleges	No access or regular extracts
University of California System	No access or regular extracts, although authorized users have access to the Statfinder tool with aggregated reports of the entire system data
Colorado Department of Education	View unit record data for their own campus only

Board of Governors, State University System of Florida	Each campus has an "Institutional Data Administrator," (DA's) who can view the master dataset Authorized end-users at each campus receive reporting access through the DA's Also, campuses may request adhoc extracts of data
Florida Department of Education	No access, although the agency produces a set of de-identified data and distributes to the IR department on each campus for reporting/comparative purposes
Kansas Board of Regents	View unit record data for their own campus only
Kentucky Council on Postsecondary Education	View unit record data for the entire system with a few exceptions (e.g., some academic performance data)
University of Maine System	No access or regular extracts
Maryland Higher Education Commission	No access or regular extracts
Minnesota Office of Higher Education	No access or regular extracts
Minnesota State Colleges and Universities	View unit record data for students that are associated with their own campus only
Missouri Department of Higher Education	No access or regular extracts
Montana University System	Updating access for unit record data for students that are associated with their own campus only. Also, can use an Access front end where they can view individual unit records from all campuses
Nevada System of Higher Education	No access or regular extracts
New Jersey Commission on Higher Education	Aggregated reports of the entire system data
New Mexico Higher Education Department	No access or regular extracts
City University of New York	View unit record data for students that are associated with their own campus only
State University of New York	View unit record data for their own campus only
North Carolina Community College System	View unit record data for their own campus only as well as deidentified data for students from other colleges
Ohio Board of Regents	Aggregated reports of the entire system data
Oregon Department of Community Colleges and Workforce Development	No access or regular extracts
Oregon University System	View unit record data for their own campus only
Pennsylvania Department of Education Office of Postsecondary and Higher Education	View unit record data for their own campus only (including students' entire educational history from K-12 as well as PSE) and aggregated data for other institutions in the system
Pennsylvania State System of Higher Education	No access or regular extracts, although currently working on database improvements that will allow each campus to view unit record data for their own campus only and aggregated data for other institutions in the system
South Carolina Commission on Higher Education	No access or regular extracts
South Dakota Board of Regents	No access or regular extracts

Vermont State Colleges	View and update unit record data
State Council of Higher Education for Virginia	Generate pre-defined unit record data displays for their campus only
Washington Higher Education Coordinating Board	No access or regular extracts
West Virginia Higher Education Policy Commission	No access or regular extracts
University of Wisconsin System	View unit record data for their own campus only
University of Wyoming	View unit record data for their own campus only (SUR system is data for UW students only)
Wyoming Community College Commission	No access or regular extracts
Utah State Board of Regents, Utah System of Higher Education	No access or regular extracts

In some cases, agencies noted that unit record data may be shared with postsecondary institutions that are not a part of their systems. This exchange process is separate from that described above for the campuses that contribute data to the system. For instance, the University of California has on-going data sharing agreements with the California State universities as well as the community colleges. When there is a need to do matching, the records shared may be identifiable. These types of relationships are governed by agreements that explicitly define the records and data elements that are exchanged as well as security and confidentiality expectations.

***Other State or Federal Agencies***

Apart from data requests from third party researchers or postsecondary institutions, regular or single instance data exchanges also occur between some of the SHEEO agencies and state or federal agencies. These exchanges are typically either for reporting or research purposes. In a few instances SHEEO agencies contribute postsecondary elements to another unit record data system. For example, the Minnesota State Colleges and Universities submit a subset of their data to the state Office of Higher Education; the California Community Colleges and the California State Colleges and Universities send extracts to the state governing board; and the Oregon University System conducts annual matches with the Oregon Student Assistance Commission, which then files performance measures with the legislature. Also, in states that have or are currently creating P-20 systems, such as Florida, New Mexico, and Pennsylvania, the postsecondary-specific agencies do (or are planning to) conduct exchanges with the state K-12 agencies.

In terms of data exchanges for research purposes, typically the participating agencies enter into a contract that explains the specifics of the arrangement. In Montana, for example, the University System office (MUS) conducts a regular exchange of data with the Montana Department of Labor and Industry (MDLI), which matches workforce outcomes with higher education information. According to the Memorandum of Understanding between the two offices,

this data exchange will enable research by MDLI and CHE [the Commissioner of Higher Education] on the employment of students/graduates from the MUS in Montana’s workforce for or on behalf of the MUS. Information from wage records

and other administrative data will be used to determine if and where students/graduates are employed within the state. The impetus for this exchange of data is solely for the purpose of conducting research to improve the quality of higher education programs and instruction for Montana citizens and the study of labor market trends.

Another example of an exchange for research purposes is between the City University of New York (CUNY) and the New York City public schools. Almost 70 percent of CUNY first year students come from New York City public high schools. Approximately one year ago, CUNY signed a Memorandum of Understanding (MOU) with the schools to do a two-way exchange of unit record data to enhance their understanding of students' progression through the educational pipeline. Also, CUNY had been getting data directly from the state Department of Education for at least 20 years to assist with their admissions application processing. This MOU gave access to those data as well as other information for evaluation purposes.

Both the Montana and CUNY examples illustrate a common theme across the respondents regarding unit record data exchanges. Among those who are open to sharing data with third parties, ensuring that the exchange will be a useful endeavor that generates new and informative recommendations for policy or practice is critical.

### **DATA REQUEST PROCESS**

This section covers in more detail the data request processes in place at the different agencies for managing inquiries from the external groups described in the prior section. When discussing the release of unit record data, a number of agency representatives noted that the approval process is very rigorous. As the Florida Department of Education website for unit record data requests states, "proposals are reviewed for approval based on: demonstration of clear benefits to the development of education policy within the state of Florida; demonstration of clear need for student-level data; and demonstration of sound scientific merit through the appropriate application of methodology and use of data." Many agencies deny unit record data requests for a myriad of reasons. Commonly mentioned by respondents was a concern that providing the data may be in violation of FERPA. Additionally, the agency may not have the data that the requestor desires or the requestor is vague about the purpose of the proposed project. A few of the individuals interviewed also noted examples where data requests were denied due to concerns that the proposed research may produce results that would reflect unfavorably upon the colleges and universities that contribute the data. If research is conducted using data that generates unfavorable publicity, the institutions may opt out of contributing data to the system. This may be especially true in cases where institutions participate in the system on a voluntary basis, which is often the case with respect to private institutions.

The process by which external groups, such as those described above, request unit record data is often not formally documented by agencies (exceptions include the Florida Department of Education). However, there is often a common and known chain of approval

and administration within these organizations. For example, a requestor may contact the agency office through an email or telephone call and communicate their research question(s) and attempt to establish the need for data. A common theme that emerged in the interviews was that requestors are often unsure as to exactly what data elements will enable them to meet the objectives of their project, so often times agency representatives will have several conversations to define the data specifications. A few agencies have formal documents that requestors complete and submit as part of the consideration process. The Florida Department of Education, for instance, is among the most comprehensive with its Unit Record Data Request Packet, which consists of six sections for requestors to complete (see <http://www.fldoehub.org/research>):

- (I) Project information, which includes basic facts such as the requestor name, affiliation, project title, and project funding;
- (II) Project description, which includes more details about the purpose of the research, research questions, methodology, and data needed;
- (III) Timeline, which must specify estimates of the timeframe for the Data Collection Phase, Data Analysis Phase, Report Writing Phase, and Final Report/Publication of the research;
- (IV) Statement of benefit, which asks the requestor to describe how the state of Florida might use and benefit from the research;
- (V) Data element crosswalk, which specifies the technical data elements requested; and
- (VI) Security and access agreement, which details the protection and use of the data.

After the unit record request is made and documentation submitted (where required), the size of the agency seems to dictate the layers of approval through which the request is processed. For instance, those with two or three staff members will typically field the request, discuss the issue(s) with their manager, and if the request is approved, process it themselves. Most agencies have a legal counsel available to them. However, whether the legal counsel is included as a formal step in every unit record release approval, or only in cases where the request is considered to be sensitive or (possibly) includes a request for identifiable information, varies across the agencies.

The Maryland Higher Education Commission is among those with well documented and multiple levels of unit record approval. There are six layers through which each request is processed. Requests come into the agency's research group, and they examine each one and decide whether or not it is possible to execute. The request is then passed on to the owner(s) of the data; the agency is considered to be the "custodian" of the data but it is "owned" by the institution that collects it. The request is next sent to the Commission's legal counsel. Then, the Secretary of Higher Education is presented the request for approval. The request is then given to the "Segment" advisory committee, which is composed of Segment CEOS (e.g., Chancellor for Universities, community college representative, independent colleges' representatives), for their approval. Finally, the request is given to the Commission's Chief Information Officer for approval.

The chain of approval for the Kansas Board of Regents is different from the norm, in that the agency has developed a relationship with the Kansas Research Consortium, a group of

representatives from the state's research universities, and plans to incorporate the Consortium into the management of its data request process. Although the specific details and policies related to the request process are under development, the agency envisions that the consortium will handle and approve third party researcher requests for educational data. Because the agency does not plan to release personally identifiable unit records, privacy and protection concerns are not as critical of an issue for the types of de-identified data requests that the consortium will manage. The Florida Board of Governors (FLBOG) also has a slightly different way of processing requests, in that each member campus' president has identified an "Institutional Data Administrator" (DA) who may field requests for data. Although these requests are approved by the FLBOG, the DA may actually administer the requests and provide the data to approved requestors. The FLBOG is an exception in this regard, as almost every SHEEO agency administers the requests in-house only.

Commonly, after the appropriate approvals have been given, the agency will establish a formal Memorandum of Understanding or Data Sharing Agreement with the requestor. Table 1 (p. 7) lists the types of agreements that each agency creates with third party researchers. Several agencies provided sample agreements to the *Protection and Accessibility* researchers, and these were analyzed for common characteristics. These agreements typically include stipulations around the following:

- The researcher will use the data for research purposes only and not release it to any other person or organization beyond those approved by the agency.
- The data will be physically secured (e.g., behind locked doors) and electronically secured (e.g., password protected, behind firewalls) from access by unauthorized persons.
- The researcher must follow all federal, state, and local statutes, regulations, and other requirements pertaining to the security, confidentiality, and privacy of data including FERPA and the Privacy Act of 1974.
- All reports, publications, or other documents that are created from the data will not include identifiable information, pursuant to FERPA, such as name, social security number, student identification number(s), and date of birth.
- An expiration date by which the research should be concluded is specified, and the agency asks the researcher to destroy or return the data at that point if the agreement is not extended.
- Any violation of the terms will result in the termination of the agreement and the imposition of penalties on the researcher. In some cases, the agreement may specify the penalties. For example, the Kentucky Postsecondary Education Commission's user agreement states that, "I acknowledge that I have been informed that revealing personally identifiable information can be considered a Class B misdemeanor under Kentucky law, KRS 164.991 punishable by a fine of not less than \$25 nor more than \$100 and by incarceration in a county jail for no more than thirty (30) days."
- The agency retains all rights to and ownership of any and all data (and copies of data) provided to researcher pursuant to the Agreement. Sometimes, the agency will also ask that the researcher submit any materials intended for publication or dissemination. For example, the University of California Office of the President has

an optional provision that states, “Researcher agrees that, at least 30 days prior to publishing, circulating, or otherwise disseminating research based on the data provided pursuant to this agreement, he or she will submit to UCOP any materials intended for such dissemination along with a thorough accounting of how the results of said research were derived. Such materials include, but are not limited to, preliminary or final drafts of research papers or presentations. Researcher may provide such materials only to [Administrator] or to UCOP staff or agents designated by him.”

When asked to estimate the typical turnaround of a unit record request, most stated that the time varies considerably. A frequently identified factor that influences turnaround is internal workload – many of the agencies have reporting relationships with their state legislatures, and when these bodies are in session their requests take priority over all others. As discussed in other parts of this report, the number of external requests for data that each agency receives varies considerably as well, so those who handle perhaps one per year clearly face different turnaround issues than agencies facing many requests per month. Communication with the requestor about the nature of the request and technical specifications in terms of data elements to include is another factor. In many cases, requestors are unfamiliar with the data and/or exactly how and what they plan to do translates into a request. When asked in the interviews about turnaround of unit record requests, most respondents stated that there is a great deal of variation. Some of the smaller agencies that handle less than one request per year estimated a week or a few weeks for turnaround, but others with heavier request loads estimated up to six months for some of the more complex and/or lower priority extracts. For instance, the Florida Department of Education informs all requestors: “researchers should expect a minimum of 3-4 weeks to receive notification regarding whether the request has been approved. Once a request is approved, researchers should expect an average of 5-6 months to receive the data requested. This duration can vary greatly by individual proposal depending upon data permissions required, datasets requested, and the number of proposals currently approved.” In some states such as California, public record laws require the agency to respond within a given period of time, although these apply much more frequently to reporting and aggregate data inquiries than unit record level data.

Also, the turnaround of requests may be influenced by the amount of work required to construct the agreed-upon data extract. In some cases, complex data matches and merges must be done in-house, and these require several hours of dedicated programming. In Ohio, it typically takes one to two weeks of a full-time trained programmer’s time to do a unit record extract from the Board of Regents’ system. So the approximately five to six requests per year that the agency receives can be a substantial drain on staff time, and the agency has begun to deny requests if the requestor is unable to pay for the staffing. Most agencies did not report charging for requests on a regular basis, but several noted that researchers have occasionally written programming time into their grant proposals and in these cases financial compensation will be provided to the agencies. This does not occur on a regular basis at any of the agencies interviewed for this study, however.

In terms of the actual data exchange process with third parties, agencies use several

methods. Some rely upon secure FTP of files, some have secured websites, and some use physical medium like CD's. In a few cases, if FTP is not an option or if the data recipient is physically located nearby, data exchanges may occur via CD or disk. At SUNY, when data is exchanged using CDs, the encryption key to access the data is provided separately from the actual CD where the data is stored. Although not explicitly asked about the use of email, when discussing the exchange of data, many agencies noted that their data security practices prohibit the exchange of identifiable student records through email and have employee data security or confidentiality policies that reiterate the importance of this.

In some situations, agencies follow the same procedures described above when exchanging data with other agencies such as the state department of labor or K-12 agencies. In other cases, regular interfaces are in place that allow for the transmission of files back and forth. For instance, the University of Wyoming has an operational data server that is connected to the unit record server, which feeds data to the National Student Clearinghouse (a non-profit group that confirms the enrollment status of financial aid recipients to lenders and the Department of Education) as well as sends files to the Wyoming Department of Education, with which the agency has a shared scholarship program that necessitates data exchanges.

## **DATA RETRIEVAL SYSTEMS**

Regarding non-restricted aggregated data that do not need special requests and/or approval processes, there are several ways that agencies make data publicly accessible to stakeholders. As mentioned in the *Accessibility of Unit Record Data: Aggregate Data Requests* section (p. 4), most maintain websites with static aggregated tables designed to answer often asked questions. In addition, several agencies have built interactive reporting tools that enable public and/or internal users to generate aggregated, de-identified data displays.

The Florida Board of Governors (FLBOG) is an example of one agency that recently revamped its reporting to better meet end users needs. While the FLBOG has maintained a system for the state's 11 public four-year colleges and universities since 1972, until approximately four years ago, the system was mainframe-based and lacked the capabilities to meet all of the FLBOG's reporting requirements. After a two year re-design to a relational database, with security upgrades and a web interface used for data submission, the FLBOG has developed a system that provides reporting capabilities to institutional researchers and other university representatives. Implemented in Fall 2009, this system includes de-identified unit record level data. It is longitudinal in that it includes data for the past five years, although the goal is to build a system that will contain the prior 30 years of data. The system has a set of pre-existing reports that attend to the topics around which users most frequently need information, and the system also permits user-defined queries. Additionally, the FLBOG maintains an Interactive University Data website, which allows the general public to create reports from enrollment, degrees, credit hours, education preparation programs, instructional expenditures (see Table 3 for website link and other examples of public reporting interfaces).

**Table 3.**  
**Examples of publicly accessible online reporting systems.**

Florida Board of Governors	<a href="http://www.flbog.org/resources/iud">http://www.flbog.org/resources/iud</a>
Minnesota State Colleges and Universities	<a href="http://www.mnscu.edu/board/accountability/index.html">http://www.mnscu.edu/board/accountability/index.html</a>
West Virginia Higher Education Policy Commission	<a href="http://wvhepcnew.wvnet.edu/index.php?option=com_content&amp;task=view&amp;id=104&amp;Itemid=0">http://wvhepcnew.wvnet.edu/index.php?option=com_content&amp;task=view&amp;id=104&amp;Itemid=0</a>

## **POLICIES**

In processing data requests, states and agencies vary a great deal regarding whether and what types of policies are in place pertaining to student unit record data. However, FERPA is the common standard under which all operate. When asked about the policies that guide the way the agency manages its data, FERPA was invoked by every single agency without prompting. All expressed an intention to do their best to comply with FERPA regulations, although interpretation often presents a challenge (see *Challenges* section, p. 21). As noted in the Data Request Process section, most agencies rely upon in-house legal counsel to assist in determining whether their data sharing and security practices comply with FERPA. The advice of legal counsel leads to consistency within each agency’s data release practices. This section further examines policies across the different agencies; specifically, how they interpret and comply with FERPA and institutionalize policies that ensure data accessibility and protection at both agency and state level.

### **Organization Level Policies**

Whether the agency itself has any sort of internal policy in place pertaining to postsecondary unit record data – as well as policies specific to the access of the data – varies quite a bit. As Table 1 demonstrates (p. 7), the majority of agencies establish Data Sharing Agreements or Memorandums of Understanding when exchanging or sharing unit records with external parties. These agreements function as policies for the specific relationships, and the contents typically stipulate how the data will be used, disclosed, stored and destroyed/returned. In some cases, agencies will develop additional documents outlining the expectations of their data sharing relationships. The Montana University System, for instance, has a Memorandum of Understanding for its exchanges with the state department of labor. They also have a separate Data Exchange Security Plan with the department of labor, which specifies a Data Exchange Flowchart, and procedures around data transfer, data storage, data elements, and a timeline of data exchange.

Beyond such relationship-specific policies, however, sixteen of those interviewed do not currently have formal agency-level policies pertaining to the access, management, or disclosure of their unit record data. Two others are currently working to develop such policies. Among the more well developed policies are Kentucky’s Council on Postsecondary Education’s Institutional Data Policy (see [http://cpe.ky.gov/NR/rdonlyres/428A2CA2-D0B3-4801-B6D2-007B058EAA1B/0/1\\_datapolicy\\_attachmentcleanversion.pdf](http://cpe.ky.gov/NR/rdonlyres/428A2CA2-D0B3-4801-B6D2-007B058EAA1B/0/1_datapolicy_attachmentcleanversion.pdf)) and Academic Data Use Agreement Policy. Others are the Florida Department of Education’s

Integrated Education Data Systems Security and Access Policy & Procedure (see <http://fldoehub.org/research>) and the Pennsylvania Department of Education's Data Access Manual includes a section pertaining to the release of aggregate, deidentified, and identifiable unit records (see <http://www.edportal.ed.state.pa.us/portal/server.pt/community/pims-pasecureid/1839/documents/285865>).

Offices that are part of the state's higher education governance system office (i.e., the Board of Regents/Board of Governors, chancellor, or system president) often are subject to system-level policies or governing body documents. The Florida Board of Governors regulations, for instance, include a section specific to the State University System Management Information System, which is the formal name of their unit record system, as well as regulations pertaining to the Security of Data and Related Information Technology Resources. Others who noted system-level policies in place pertaining to postsecondary student unit record data are the University of Alaska Statewide System, the Minnesota State Colleges and Universities, Montana University System, Nevada System of Higher Education, City University of New York, Oregon Department of Community Colleges, South Dakota Board of Regents, Vermont State Colleges, and University of Wyoming.

Formal security and confidentiality policies have been created and distributed to users by most of the agencies. These policies outline administrative, technical and physical safeguards as well as individual privacy protections. In most cases, these include password distribution and control, incident handling, disaster recovery plans, record retention and disposal, tracking of user access, intrusion detection, and security software upgrades and/or patches. Several policies discuss the use of individual identifiers and/or Social Security numbers and measures to protect identities; for instance, the University of California System has a separate policy that outlines Guidelines for Release of Aggregated Student Data with Small Cell Sizes.

### **State Level Policies**

As with the case of agency level policies the states themselves vary quite a bit in terms of policies that apply to their postsecondary student unit record data. Several general types of state-level policies were identified in the interviews, including those that extend federal record privacy and freedom of information policies, as well as those that authorize the collection of postsecondary unit records.

A number of states have laws that regulate government practices involving the dissemination of data and/or public records generally and extend to student unit record data, including the Right to Know Act in Pennsylvania, the Open Records Acts in Colorado, New Jersey, Kentucky, and Kansas, the Public Records Law and Chapter 62 (Act Relating to the Protection of Personal Information) in Vermont, Chapter 119 (Public Records) in Florida, Oregon's Administrative Rules, New York's Public Officer's Law Personal Privacy Protection Law, Nevada Revised Statute NRS 239-B and NRS 603-A, Data Collections and Dissemination Act in Virginia, the Government Data Practices Act in Minnesota, the California Information Practices Act, and the Freedom of Information Act in Maryland. These laws typically incorporate FERPA provisions, speak to the protection, confidentiality,

and security of information collected by state agencies, and may also expand upon how agencies should handle freedom of information requests.

Several states – for example, Ohio – have policies that are specific to the data collected at the elementary and secondary education level, but do not extend specifically to the postsecondary realm. However, a number of states have established statutes or legislative mandates that explicitly permit the agencies to collect and use postsecondary data. These include Alabama, Kentucky, New Mexico, Oregon, and Washington. Those who were interviewed were very aware of the general interest in moving toward major synchronized systems across primary/secondary and postsecondary education, and in a few cases, states have – or are in the process of – mandating the creation of these integrated systems. For example, after a recent executive order from the Governor, New Mexico is currently working to establish an overarching, multi-agency policy that will create a P through 20 system.

## **DATA SECURITY**

The *Data Security* section of this report describes “best practices” occurring across the nation with respect to unit record data access management and security. The topic of data security is certainly a sensitive topic, and understandably some agencies were hesitant to publicly describe the specifics of their secured environment. This section therefore synthesizes the overall practices in place around student unit record data security and highlights exceptional/leading activities, without identifying agencies in a manner that might compromise their security.

Standard physical and access security measures are in place at all of the agencies. Database servers are physically secured in limited access computer rooms and virtually secured on a safe network behind firewalls. Full server/data backups with secure off-site storage occur on a regular schedule, and several agencies have documented disaster recovery plans for their data. Server login is limited to local domain administrator accounts, and login access to database software is limited to authorized agency staff. Password policies require regular changing of passwords as well as specifying minimum length and combinations of characters, numbers, and symbols. In a number of cases, the agency data is resident on the states’ higher education system mainframe, so they also piggyback on the broader system security mechanisms and policies. Also, several reported undergoing recent IT security audits by state auditors or by external consultants that reinforced and/or upgraded their internal security practices. As noted in *Organization Level Policies* (p. 18), most agencies have documented security policies in place. These include end user practices – for instance, never exchanging unit records through email was frequently mentioned in the interviews as a standard end user practice – as well as confidentiality expectations.

In terms of personally identifiable data, while almost every agency collects social security number as an identifier, a number have come up with extra measures to encrypt and/or remove this variable from their systems after matching occurs. A few agencies match their data in “black box” environments that are not online. Social security numbers are removed

before the data is uploaded into the live database, and randomized IDs are assigned to the records. A number of agencies do not collect student directory information (i.e., name or street address) and/or only have students' birth years, which is another layer of identity protection.

Several systems have row-level security access in place, which permits highly customized user access. While older systems do not have the capacity to limit access within a given file, row-level security permits the specification of viewing or updating access to specific fields and/or records within a file. Also, many newer systems have role-based access, which defines permissions to data according to the user's business responsibilities and needs. One role created in these types of systems is often that of "Researcher," which allows the user access to the reporting instance and sometimes de-identified unit level data. Sophisticated systems have tracking features in place that enable the main agency office to monitor which role(s) and user(s) access what data at any given point in time.

### **PROTECTION AND ACCESSIBILITY: CHALLENGES AS IDENTIFIED BY AGENCIES AND BEST PRACTICES IN RESPONDING TO CHALLENGES**

Given the structures, policies, and practices described above that are in place at the various agencies, respondents were asked to identify the primary challenges that they face in responding to requests from data. In addition, to inform best practices that might assist in responding to challenges pertaining to the access and security of student unit record data systems, two experts in the area of unit record data systems were contacted based upon their knowledge and experience in dealing with the access and oversight of large state student unit record systems. These experts are Jay Pfeiffer and James Parsons. Jay Pfeiffer is currently a Senior Associate in the Florida office of MPR Associates Inc., which provides educational services to a range of clients with a focus on data driven solutions, including research and evaluation, policy analysis, Web design and technology tools, training and technical assistance, and materials development and dissemination. Before joining MPR, Jay was the Assistant Deputy Commissioner for Accountability, Research, and Measurement at the Florida Department of Education. In the Department, he had been responsible for the development and management of the Florida K-20 data warehouse. James Parsons is the Assistant Director of the Texas Schools Project (TSP), which supports academic research and evaluation to improve academic achievement, teacher effectiveness, increase transitions to and success in postsecondary education, and improve labor market outcomes of students in Texas and the nation. As one of Texas's Education Research Centers, TSP maintains a data warehouse of PK-12 data from the Texas Education Agency and college data (2 and 4 year) from the Texas Higher Education Coordinating Board.

Each expert interview lasted approximately 60 minutes and focused on how the access and security protocols of student unit record data systems might be improved. Also, the experts were asked to reflect on the challenges that agencies identified in their interviews with regard to managing external requests for data from their unit record system. These challenges fall into five general categories: "Managing Relationships with External Requestors of Data,"

“Interfacing and Data Sharing Practices,” “FERPA Compliance,” “Maximizing Available Resources,” and “Implementing or Upgrading Public Reporting Capabilities.” This final section of *Protection and Accessibility* describes each challenge, followed by possible solutions, opportunities for improvement, and “best practices” offered by the responses of the experts.

### **CHALLENGE: MANAGING RELATIONSHIPS WITH EXTERNAL REQUESTORS OF DATA**

Two areas of concern emerged from the interviews in terms of how agencies manage relationships with external researchers and other third parties requesting unit record data. First, agencies must address growing demand from these groups; and second, agencies must attend to a lack of requestor knowledge about the data and how it might meet their research needs.

#### **Increasing demand and prioritization of requests**

Several respondents noted that as more state residents become aware that the SUR data exists and may be available to them, the number of requests (mostly reporting/aggregate type requests) has increased. In terms of how agencies might best manage this demand for data from external groups, the typical practice is to prioritize requests from policymakers, while extracts for third parties are usually given lower priority. The turnaround for third party requests, therefore, may take as long as six months to a year. In most cases, because the purpose of the postsecondary data is to improve educational outcomes in the state, what the governor, state legislature, state Department of Education and/or higher education Board of Governors wants should (and will) take priority over providing data to external researchers, both now and in the future. Therefore, agencies must be cognizant about the most effective and efficient ways to allow for external researcher access to these data.

#### **Best practices suggested by the experts**

In Dr. Pfeiffer’s opinion, the current practice of agency staff creating extracts and providing these to approved researchers on an individual basis is an “antique way of doing business these days.” Instead, he suggests that agencies must move toward ways that will allow external researchers to take more responsibility in the data retrieval process. Models similar to those used by the National Center for Education Statistics (NCES) or the U.S. Census Bureau for providing restricted data access offer benchmarking opportunities for agencies managing unit record databases. These national-level data systems operate in a way that permit individual researchers to take responsibility for extracting and utilizing the data themselves. Also, these systems have sophisticated security controls that restrict what kind of data that external users can view and/or download. Leveraging the opportunity for controlled external access – both to de-identified unit record data and aggregate level data – is an important piece of research and development that should be occurring in parallel with the development of the unit record databases themselves.

#### **Requesters’ lack of comprehensive knowledge of the data**

Another commonly mentioned challenge was the time-consuming negotiation involved with external parties who request unit record data extracts, and the experts agreed that this is an ongoing issue. Often these individuals are not able to clearly articulate what they

are requesting and how that translates into the data that is collected in the warehouse. Or the requestors will not clearly articulate what questions they are trying to answer and what data may be needed to do so. The lack of clarity many times requires multiple follow-ups with the agency in order to add data elements. Ensuring that there is a mutual understanding of the variables that are part of the system, as well as clarifying the requestor's data needs to answer their research questions, often requires a great deal of communication between agency staff and data requesters.

A related challenge is end user comprehension of the data when actually conducting their analyses. In several cases, agencies reported situations where researchers misinterpreted or misunderstood the contents of the data with which they were working. If requestors misinterpret the data, produce a study, and are ultimately called into court to testify about the schools, this could conceivably result in misinformed or erroneous policies due to the dissemination of inaccurate findings. In part to address this issue, some agencies such as the California Community Colleges and the University of California Office of the President include a stipulation in their researcher contracts that require any reports to be provided back to the agency for the opportunity to review the findings before publication (see example text in *Data Request Process* section, p. 15).

### **Best practices suggested by the experts**

Given the challenges related to communications and lack of understanding of the data, the experts offered several suggestions for improving and optimally managing the data request process. First, access rules and the process by which requests are considered and approved or denied need to be transparent to internal staff as well as third parties interested in requesting data. These rules should specify who the contact person is at the agency for each type of request and what the researchers need to do to get permission to receive the data. Often external researchers perceive a great deal of uncertainty about the process of obtaining data, but by clearly outlining what a proposal must include and the rules for vetting these proposals, agencies can reduce this uncertainty effectively. For example, the Texas Schools Project access process is described and the required forms are available at <http://www.utdallas.edu/research/tsp-erc/access.html>, and the Florida Department of Education's request information is available at <http://www.fldoehub.org/research>.

Another opportunity for agencies to improve the knowledge of external requestors regarding the use of the data is to focus on properly documenting the data elements so as to assist external researchers or other third-party researchers and reducing user confusion. The need for proper documentation is pronounced especially when these external requestors plan to use the data for the purpose of longitudinal analyses. While many agencies have some type of formal documentation (such as a data dictionary) in place, these dictionaries often do not inform how to use longitudinal records of the data. Although data dictionaries are often updated annually, most agencies do not provide information that tie those annual dictionaries together over time to give a clear picture of how the total longitudinal data is collected, measured, and modified from year to year. Failure to update changes in longitudinal data information may cause problems for external researchers, especially when the definition of elements changes over time but the

variable name does not. If agencies are more attentive to documenting these changes and construct longitudinal records of their data, researcher comprehension will be improved.

Given the concerns described in the preceding paragraph, the Texas Schools Project is organizing a series of analyses intended to provide external researchers with a better understanding of their data. The types of topics that these studies will consider include: how “long” is longitudinal? How many students in any given cohort can be tracked over time? How are students who can be tracked over time different from those students who cannot be tracked? How long can someone follow specific student subgroups such as economically disadvantaged students versus others? How far back can someone trace the graduating class of 2009? The Texas staff plans to develop a series of short one to two page reports that will give a better indication of what the data contain so that this information can be used to inform the effective use of their student unit record data.

### **CHALLENGE: INTERFACING & DATA SHARING PRACTICES**

Another theme that emerged in the interviews is the challenge of understanding how to link separate data sources, such as labor and/or incarceration records with educational records. An additional concern is a lack of understanding about who is responsible (“owns”) the data and who controls its use/release to third parties. This is obviously especially important given the emphasis on building P through 20 systems.

### **Best practices suggested by the experts**

According to Dr. Parsons, a critical aspect of creating these interfaces is technical, in terms of how to link all of the data elements together. Although Social Security number has been the commonly used unique identifier, some agencies no longer collect it due to privacy concerns and rely instead on their own ID number or a state ID number. In Texas, for instance, the K through 12 systems would like to move to a non-Social Security unique identifier. The burden will be placed on the postsecondary institutions to collect this new identifier, but it will be a technical challenge to match that data with workforce data (which must rely upon Social Security number because of IRS and Social Security Administration reliance upon them). While data can be linked using secondary identifiers (e.g., name, birth date, gender), this is an expensive undertaking and not as effective as using a common unique identifier. This is an important issue for agencies to consider, although due to privacy concerns, many in the public are pushing to have agencies move away from the use of the Social Security number.

### **CHALLENGE: FERPA COMPLIANCE**

The interpretation and the application of FERPA to each agency’s specific circumstances was frequently cited as a challenge in the interviews with the different SHEEO agency representatives. For example, one respondent expressed ambivalence about the process of determining exactly what her agency was able to disclose to the various requesting entities. Also, respondents noted some apparent contradictions in FERPA in terms of balancing data transparency/release with data protection. One respondent said that his biggest hurdle is that other agencies pick and choose what data they will provide to his agency, and FERPA is a “card of convenience” that they will pull out when they do not want to provide data or data elements. This legal confusion will pose a big challenge in the

future, especially as the federal government is calling for broader data-driven accountability that will require more extensively networked databases.

### **Best practices suggested by the experts**

Due to the ambiguous and often contradictory FERPA interpretations and lack of clear administrative rules, different agencies follow different processes and come to different conclusions about managing and releasing their data. There is clearly not a one-size-fits-all solution to applying FERPA, but the case of Texas provides an example for those struggling with how to ensure their processes are FERPA-compliant. Texas worked closely with the Family Compliance Office in constructing their unit record data release procedures. The process by which this occurred – and the final design of the approved process – is outlined in an Urban Institute Brief written by Dan O’Brien, the Director of the Texas Schools Project at the University of Texas-Dallas (see [http://www.urban.org/UploadedPDF/1001267\\_texasferpastory.pdf](http://www.urban.org/UploadedPDF/1001267_texasferpastory.pdf)). The document provides a point of reference, at least, for agencies attempting to define a FERPA-compliant process pertaining to the release of data for research purposes.

Also, after data are provided to external researchers, following up to ensure that the data are used only for the expressed original purpose is an important aspect of FERPA compliance but an area where many agencies fall short. Although, as the law specifies, data may be used for the reasons that were approved only, in Florida at least, there have been situations in the past where external researchers have used extracts to build their own data warehouse for purposes beyond what they originally defined. Therefore, as agencies begin to regularly provide data to external researchers, implementing some form of monitoring apparatus to ensure that data are properly destroyed is important.

### **CHALLENGE: MAXIMIZING AVAILABLE RESOURCES**

The issue of limited resources is certainly a common theme that emerged from this project. In a word, “resources” was the most frequently cited challenge identified during the interviews. For example, budget cuts and lack of new financial inflows limits the ability of small staffs to attend to a growing workload.

### **Best practices suggested by the experts**

One option for addressing financial constraints is to implement charges for access to the data as well as offer professional research services like program evaluations conducted by agency staff, as the Texas Schools Project has done. Although the agency’s initial funding came from an infrastructure grant provided by a foundation, its continuing line of fee-for-research projects covers a great deal of overhead and personnel costs. When implementing fee schedules, making these public in terms of what types of services are offered and what the charges are (i.e., a “menu”) is a useful means of preempting inquiries from potential requestors about such costs.

In addition to financial resources, agencies should maximize available non-financial resources by sharing data with others external to the agency. A critical element of sharing data with others external to the agency is what Dr. Pfeiffer labeled “quid pro quo,” or ensuring that all parties involved benefit from the relationship. Where external researchers

are concerned, their work with the data may offer a valuable opportunity for agencies to maintain their relevancy in the broader educational and policymaking context. The research that external groups can – and will – produce using SUR data should be leveraged by the SHEEO agencies and the states themselves. States need to tap this researcher expertise to find answers from a policy and an evaluative standpoint. Once researchers work through the data, they can serve as “expert witnesses” to help make the case to legislators and governors for particular policies or program direction. To document these opportunities for possible leveraging in the future, a few agencies ask researchers to formally explain how their work will benefit current and future policy practices and reforms when they request the data. For instance, the Florida Department of Education (DOE) requires external requests of their data to be accompanied by a Statement of Benefit. This must explain how the research will provide information about educational needs in Florida, or provide information about the effect of alternative educational practices in Florida, or measure the effect of an implemented state policy or program. Also, researchers are asked to demonstrate how the DOE might use the publication/research in its final form and which office will benefit from the research conducted.

### **CHALLENGE: IMPLEMENTING OR UPGRADING PUBLIC REPORTING CAPABILITIES**

A final topic that emerged from the interviews pertained to enhancing reporting capacities. A few respondents expressed a desire to improve upon different aspects of their system (e.g., add a reporting interface for public and/or internal users or upgrade internal reporting tools to more user-friendly programs) but funding limitations preclude doing so. In some cases, respondents noted that when their data warehouses were originally constructed the systems met their needs, but over time evolving demands from the state as well as external requestors have rendered the warehouses less able to meet current needs. For instance, the North Carolina Community College data warehouse is set up for reporting purposes on a semester or annual basis. The agency has found that while it meets their reporting needs, it falls short in terms of use as an analytical tool. When the agency attempts to conduct longitudinal data analyses, merging or building the data is a complex process.

A related challenge cited by several respondents is how, from a political standpoint, to best manage the reporting of data in both unit record and aggregate form. External researchers sometimes propose topics that the agencies are concerned will produce controversial or unpopular results. Requests for reports about issues that the agency or its partner colleges and universities do want to be in the public domain also occur. Therefore, trying to understand whether a report request or research proposal will generate something controversial is a challenge. This is of particular concern to the agencies that collect data from private universities, since these institutions might opt out of contributing data should it be used to produce something that reflects unfavorably upon them.

### **Best practices suggested by the experts**

Regarding how agencies can create tools or upgrade existing ones that would meet the needs of public users and ultimately reduce the reporting burden on staff, Dr. Pfeiffer observed that the expansions in coverage of many of the data warehouses have produced the potential to find answers to almost every question that an individual might generate.

Yet the sheer scope and magnitude of data quickly becomes overwhelming to most in the general public. Therefore, a huge sophisticated reporting system or interface is likely not an optimal strategy. Instead, to determine an optimal strategy, agencies must first 1) determine what information needs to be made available as well as 2) how the average person might quickly and easily work with the information.

Where the first point is concerned, agencies must begin by identifying the main questions that many people are asking, then tease out logical follow-up questions from the main ones. For instance, if a common question posed by public inquiries is how many students graduate from the state's community colleges every year, one logical follow-up might be, are the graduation rates the same for students who enroll in remedial classes?

Where the second point is concerned, the average person's comfort with technology and basic computer skills must drive the design of any reporting interfaces. The Census website, for instance, has very ingenious reporting tools, but they require substantial time for most end users to master and produce the desired output, so few people actually use them regularly. Dr. Pfeiffer noted that several years ago in Florida, the Department of Education was interested in identifying what data the state could provide to teachers. They held focus groups all around the state with teachers to determine what the Department might develop to support the teachers' efforts. The type of information that teachers requested was quite basic, essentially how to work with a spreadsheet. Because even teachers who are educated lack in basic computer skills, it is natural that the general public like simple and direct tools on websites for reporting purposes. In sum, there is no clear, one-size-fits-all solution as to how to best report data back to the public and make it transparent. However, identifying the needs of the target audience and working to ensure that these needs are met in the most straightforward and comprehensible manner is critical.

## **CONCLUSION**

When considering how to manage the challenges identified during the interviews, the experts observed that it is important not to lose sight of the fundamental purpose of why each unit record data system was established. As Dr. Pfeiffer noted, the purpose of many student unit record data systems is to administer the states' education programs, i.e., they are administrative databases. Policymakers, administrators, and constituents often do not fully grasp this primary "management and operation of schools" function of their state's data warehouses, nor do they understand what it means to administer a state education program from these data. Instead, these groups may assume that the systems' primary function is to meet federal reporting requirements or to conduct research; in some instances, such as the Texas Schools Project, this is true, but these reasons are not necessarily why many other systems were built. Reporting and research are certainly important, but in most cases neither is the primary purpose for which the systems were designed. Fundamentally, state unit record data have an overarching purpose of improving the states' ability to understand and administer their P through 20 programs.

## **Appendix A. Interview Protocol.**

### ACCESS TO DATA

*The first set of questions focus on who is granted access to your agency's data and under what circumstances this occurs. Note that we are interested in who is able to access – both update and report on – the “master” dataset that your agency maintains. We are not interested in access to each individual campus's dataset, only the combined master dataset that is your responsibility.*

1. What are the different groups that request access to your agency's SUR data? Clarification: we are interested in unit record requests (either access to the database or unit record-level extracts that your office generates for the requestor) – not aggregate-type reports.
2. Do individual postsecondary campuses in your state have updating and/or reporting access to your agency's SUR data?  
*If response is Yes*
  - What are the postsecondary campuses that have this updating and/or reporting access to your unit record data? Is it all of them or a subset?
  - What form(s) of updating and/or reporting access do the authorized people on each postsecondary campus have to your agency's unit record data?
3. What different levels of access are available to state or federal employees/representatives when they request your agency's SUR data?
4. What different levels of access are available to third party researchers (in other words, researchers who are not employed by your state or by institutions located within your state) when they request your agency's SUR data?

### DATA REQUEST PROCESS

*The second set of questions considers the typical process that your agency follows to make your data available to requestors.*

5. Is your agency's data request process the same for all requestors or does it differ depending on who the requestor is?
6. Can you explain how the data request process typically works?
  - What office/group of people/or person in your agency is responsible for receiving data requests?
  - What office/group of people/or person in your agency is responsible for approving data requests?
  - What office/group of people/or person in your agency administers SUR data requests after they are approved?
  - Is there any information required of an applicant requesting SUR data? If so, what are the requirements?
7. Is a delivery date given to a requester when a request is made?
8. Can you estimate the typical turn-around time for a data request?
9. Could you identify the possible formats by which data can delivered to a requester?
10. Please describe at least one typical challenge that your agency encounters when responding to requests for data and how your agency responded to this challenge.

### TECHNICAL ORGANIZATION

*The next two questions deal with technical aspects of access to your agency's SUR data.*

11. Are there any data retrieval systems in place to facilitate access to the SUR data?
12. Are there any technical interfaces in place to enhance data accessibility for stakeholders?

#### POLICES PERTAINING TO THE PROPER USE AND DISCLOSURE OF SUR DATA

*The next group of questions asks about state policies and your agency's policies that specifically deal with the use and disclosure of your agency's SUR data.*

13. Do you have any internal policies within your agency that specifically address the groups to whom your agency's SUR data will be released?
14. Are there any other broader policies that specifically address the groups to whom your agency's SUR data will be released?
15. Are there state law(s) that govern the proper disclosure and/or subsequent use of your agency's SUR data?
  - Are there sanctions to parties who disclose SUR data in violation of these policies?
16. Are there any internal department or agency policies in place in addition to state laws that govern the proper disclosure and/or subsequent use of your agency's data?
  - Are there sanctions to parties who disclose SUR data in violation of these policies?
  - Does your agency ever conduct audits of groups after the requested data has been distributed to them?
17. Are there policies in place to assure that the data are destroyed when no longer needed for the purposes under which the request was granted?
18. Has your state appealed to the U.S. Department of Education for an opinion on your SUR data access protocol?
19. [If not answered by prior question], are there existing state legal opinions or legislation addressing FERPA requirements?
20. One goal of our study is to identify best practices across the country related to SUR data issues. Given that, have you looked to any other state(s) as a model upon which to base your SUR data access protocol?

#### DATA SECURITY AND CONFIDENTIALITY

*The final group of questions pertains to policies and practices pertaining to the security and confidentiality of your agency's SUR data.*

21. What data elements in the SUR system does your agency consider to be "personally identifiable information"?
22. Is there a formal policy (either that identified in the previous section or a different one) that specifies how personally identifiable data is protected?
23. Does this policy (or a different policy) determine when and from whom consent for dissemination of personally identifiable data is required?
24. Are there any separate policies or procedures for disclosure in cases where populations include only a few individuals?
25. What hardware, software, and security applications and/or procedures are in place to ensure that SUR data are not lost, stolen, vandalized, illegally accessed, or otherwise rendered useless?
26. Are there any additional comments you would like to provide, or are there any other sources you think I should check to find out more information about the topics covered here?

**Appendix B.  
Recruitment Email Text.**

Dear [name],

I hope this email finds you well! I am contacting you about a project sponsored by the State Higher Education Executive Officers Association (SHEEO), a portion of which researchers from the University of Michigan are conducting. You likely received notification about the project from Hans L'Orange and Tanya Garcia at SHEEO within the last few weeks. Essentially, we are working to update the 2006 Lumina study of Student Unit Record systems, as well as extend our understanding to include data access and security information. The Michigan researchers are responsible for this latter portion of the study and, because there hasn't been much information collected on the topic, have decided that the most efficient way to do so would be through interviews with state representatives. This is in addition to the online survey administered by SHEEO.

If possible, I'd like to schedule a time over the next few weeks to conduct our survey through a telephone conversation. You can expect it to take approximately 30 minutes. If you could respond to this email with two or three 30-minute slots of days/times that would work between now and [date], I will get back to you right away with a confirmation.

Also, SHEEO provided us with your name as a contact for [agency]. To give you a better sense of the types of questions that will be asked, I've attached to this email a draft of the interview protocol. If you believe that another individual might be better positioned to answer data access-type questions, please let me know that person's name and I will contact him or her instead.

Please let me know if you have any other questions that I did not address in this email. Thank you so much for your help with this project, and I look forward to hearing from you!

Best regards,

[Name]

## **Appendix C. Interviews Conducted.**

Alabama Commission on Higher Education~  
University of Alaska Statewide System  
Arkansas Department of Higher Education  
California Community Colleges  
The California State University and Colleges  
University of California System  
Colorado Department of Higher Education  
District of Columbia, Office of the State Superintendent  
Board of Governors, State University System of Florida  
Florida Department of Education  
Kansas Board of Regents  
Kentucky Council on Postsecondary Education  
University of Maine System~  
Maryland Higher Education Commission  
Minnesota Office of Higher Education  
Minnesota State Colleges and Universities  
Missouri Department of Higher Education  
Montana University System  
Nevada System of Higher Education  
New Jersey Commission on Higher Education  
New Mexico Higher Education Department  
City University of New York  
State University of New York  
New York State Education Department Office of Higher Education  
North Carolina Community College System  
Ohio Board of Regents  
Oregon Department of Community Colleges and Workforce Development  
Oregon University System  
Pennsylvania Department of Education Office of Postsecondary and Higher Education  
Pennsylvania State System of Higher Education  
South Carolina Commission on Higher Education  
South Dakota Board of Regents  
Vermont State Colleges  
State Council of Higher Education for Virginia  
Washington Higher Education Coordinating Board~  
West Virginia Higher Education Policy Commission  
University of Wisconsin System  
University of Wyoming  
Wyoming Community College Commission  
Utah State Board of Regents, Utah System of Higher Education

~ information provided through email rather than telephone interview