Tracking Postsecondary Students Across State Lines
Results of a Pilot Multi-State Data Exchange Initiative

National Center for Higher Education Management Systems (NCHEMS)
Center for State Policy on Student Progression (C2SP)

March 2008

Background. Increased rates of inter-institutional transfer among postsecondary students pose growing challenges to policymakers and analysts seeking to understand and influence flow through the "educational pipeline." The most recent national studies indicate that about two thirds of those exiting high school and ultimately earning a baccalaureate degree have attended more than one institution and about a fifth have attended three or more (Adelman 2006). Further complicating the picture is the fact that some forty percent of those who change institutions cross state lines. This is important because state-level Student Unit Record (SUR) constitute the primary resource available to state policymakers for understanding the educational pipeline. SUR databases—now established in 42 states—can be used to track student transfer among public institutions within the boundaries of a given state and are increasingly used for this purpose (Ewell and Boeke 2007). But they have rarely been linked to examine patterns of interstate mobility.

Supported by the Lumina Foundation for Education, the National Center for Higher Education Management Systems (NCHEMS) began efforts to test the feasibility of linking state SUR databases in 2003 with a national inventory of state SUR resources. After analyzing the structure and data element contents of extant SUR databases in some detail, NCHEMS concluded that it was feasible to try to link data drawn from these different data environments to examine patterns of interstate transfer (Ewell, Schild and Paulson 2003). Accordingly, NCHEMS worked with two states, Ohio and Kentucky, in 2003-2004 to demonstrate the feasibility of the approach and the added value for policy research of pooling student records from multiple states. Six years of data involving more than 77,000 students in Ohio and over 26,000 students in Kentucky who began their studies in the fall terms of 1998 or 1999 were successfully exchanged and matched to determine the frequency of cross-state enrollments. Results indicated a relatively modest increase in cross-state enrollment for the majority of students (less than 1% for Ohio students enrolling in Kentucky institutions and 1-2% for Kentucky students enrolling in Ohio). But this phenomenon was more pronounced in border regions like Cincinnati, in where cross-enrollments of Kentucky students in Ohio exceeded 5% (see Appendix A).

These results suggested the continued value of cross-state SUR data exchanges, and led to the four state data exchange initiative described in this report.

Building the Partnership. Early in the Ohio-Kentucky data exchange initiative in December 2003, data administrators in several additional adjoining states were convened by conference call to determine their interest and capacity in participating in an extension of the concept. Participants included representatives from Ohio, Kentucky, West

Virginia, Tennessee, and Indiana. All participants agreed to take additional steps toward establishing a one-time five-state data exchange in the following year.

Early in the dialogue, and guided by experience gained in the Ohio-Kentucky data exchange, it was determined that only a few data elements would need to be exchanged. These included the Student Identifier, Institution, Year, Term, Date of Birth, Gender, Race/Ethnicity, State of Origin, County of Origin, Student Level, Full-time/Part-time Status, Current Field of Study, Level of Degree or Certificate Earned, and Degree Field of Study. The Student Identifier, Date of Birth, Gender, and Race/Ethnicity data elements were all included to test different approaches to matching records. The remaining data elements were included to exchange substantive information about enrollments at other institutions. It was agreed that six complete years of student enrollment data would be exchanged. All these data elements were defined consistently in the SURs maintained by the various agencies, and were governed by a common file layout (see Appendix B). The expectation was that once a match was returned to a participating state indicating that a given student was enrolled in another state, the participating state could re-merge the resulting record with its own SUR database using each record's unique identifiers to harness the hundreds of other data elements about that student to generate appropriate reports. As described in the subsequent section, this component of the exchange had to be modified because of privacy concerns.

As discussions about how to accomplish the exchange proceeded, moreover, it became clear that asking one of the participating states to serve as the host for accomplishing the required data match—the role that the Ohio Board of Regents had served in the prior Ohio-Kentucky exchange—was becoming awkward. This was primarily due to privacy and FERPA-related concerns associated with having a given state agency physically possessing and manipulating uniquely-identified student record data supplied by another state. As a result, the decision was made to contract with a third-party data administrator to receive the data individually from each participating state, physically accomplish the match, and return the enhanced data file individually to each participating state. Accordingly, NCHEMS sought a capable third-party to accomplish this task, and eventually contracted with a non-profit firm (HigherEducation.org) to play this role.

Finally, previous experience emphasized the need to formalize data exchange arrangements in writing in the form of an agreement or Memorandum of Understanding spelling out specific responsibilities and limits about data access and use. In the initial stages of the project, NCHEMS staff developed a basic template for such an agreement which was use as the model for a complete "Agreement for Services" document signed by NCHEMS and the participating states. This was originally modeled on agreement between NCHEMS, Ohio, and Kentucky used in the previous data exchange. Main features of the Agreement included contracting with a third-party data administrator to receive data from each state to construct database tables based on the designated data elements, write specifications and rules for matching student records, and write web-based student mobility reports. The Agreement also spelled out restrictions on access and precisely how data were to be transported from one agency to the third party data administrator.

As the process of developing the Agreement proceeded, it became clear that differences in laws and regulations among the participating states would render unworkable a generic multi-lateral data-sharing arrangement such as that used to govern the original Ohio-Kentucky exchange. Accordingly, the process was re-crafted in the form of separate bilateral Agreements between NCHEMS and each participating state. While these were substantively identical, they were crafted to include distinctive language specific to each state (see Appendix C) and were subject to legal review by counsel in each state. The need to establish individual bilateral agreements was an important lesson of the project at this point, and emphasizes the import role of an independent third-party broker, the role that NCHEMS played, in establishing any multilateral data exchange agreement among states.

Implementation Difficulties and Modifications to the Exchange Process. The period 2004-2007 within which this pilot unfolded were extraordinarily eventful with respect to the politics of data in higher education. The original target date for data exchange using the rules of engagement defined above was January 2006. But progress was slowed by concerns about privacy raised by the highly visible NCES proposal for a national unit record data system. This controversial proposal raised privacy concerns for state data administrators who had been "flying under the radar" with respect to record-matching operations, whose operations were now subject to greater scrutiny by institutions and national organizations. Privacy concerns were raised by constituents in several participating states in the wake of these events. This delayed implementation of the exchange because of the need for project administrators at NCHEMS to call and hold several face-to-face meetings in participating states to consult with institutions. Unfortunately, these developments ultimately resulted in Indiana's withdrawal from the project.

These national and state-specific developments also forced NCHEMS to re-think the approach to data exchange and develop more conservative data transmission and matching processes designed to make them "FERPA-proof." The decision to employ a third-party data administrator to accomplish the data match was one result of this process of re-envisioning, obviating the need for participating states to physically possess one another's records. In a related step, NCHEMS developed an encryption routine for the unique identifier used to identify each record involved in the match. This encryption routine was supplied to each participating state and applied to the records to be matched before they were given to the third-party data administrator. But the encryption routine was not supplied to the third-party data administrator. As a result, the third-party data administrator could use the identifier to match records, but could not identify the records further.

By far the most important FERPA-related decision, however, was for the third-party data administrator to de-identify the matched data file before returning it to each of the participating states. Doing this fulfills the letter of the FERPA prohibition against "re-disclosure" of identified data to another participating state—an issue that had been raised prominently in several states in previous months. De-identifying the returned data file, of course, made it impossible for participating states to link the enhanced enrollment record

back with their parent data records in their own SURs in order to conduct further analyses of student enrolling in other states. To partially address this difficulty, provisions were established to enable each participating state to append up to fifteen additional data elements of their choice to the records sent to the third-party data administrator, intended for further analysis once the enhanced file was returned. These data elements passed through the entire matching process and remained attached to their associated records after these records were de-identified and returned. States could then use these additional data elements to further disaggregate the returned records.

With these rules in place to govern the exchange, Kentucky, Ohio, Tennessee, and West Virginia successfully produced data files in early 2007 to enable six years of enrollment history to be tracked for two starting cohorts (Fall 1998 and Fall 1999). Potential matching files for the six years were supplied for all students enrolled in all public institutions in each state. This resulted in a total of some fifteen million enrollment records to be matched. HigherEducation.org, the third-party data administrator, accomplished the required matches in the fall of 2007. HigherEducation.org then used the resulting enhanced records to produce a number of reports exploring the "value added" of the multi-state matching process in terms of out-of-state enrollment and degree completion (see Appendix D for the methodology used to generate these reports).

Results. Cohorts were constructed from these files using the methodology employed by the NCES Graduation Rate Survey (GRS). That is, the 1998 cohort consisted of first-time, full-time students, beginning in the summer or fall of 1998 and tracked though seven years of enrollment history. The 1999 cohort was constructed in parallel and tracked through six years of enrollment history. For each year, rates were calculated for:

- The percentage of the starting cohort that had earned a bachelors degree from a) the same institution at which the student began, b) another institution in the same state and, c) another institution in each of the other three states.

- The percentage of the starting cohort that continued enrollment at a) the same institution at which the student began, b) another institution in the same state and, c) another institution in each of the other three states.

Students were counted as completed or retained in a given year regardless of the term in which the activity took place. Once a given student completed a degree, the student was no longer tracked. Finally, because of substantial data challenges (see subsequent section), only four-year starters were tracked.

Results for the two cohorts for each of the four states are summarized in the tables below:

Summary Tables from Four-State Match--Annual
All Institutions

	KY	OH	TN	WV
1998 Cohort				

Bachelors Degree After 7 Years

	KY	OH	TN	WV
Number of Students	13784	37852	14665	10753
Same Institution	42.6%	55.2%	45.0%	39.5%
Same State	47.0%	57.3%	49.5%	43.6%
4-State Region	48.2%	57.8%	50.0%	44.1%

Retained 2nd Fall

	KY	OH	TN	WV
Same Institution	71.6%	77.6%	74.9%	72.1%
Same State	81.1%	91.8%	86.5%	81.0%
4-State Region	83.6%	92.3%	87.3%	82.6%

Retained 3rd Fall

	KY	OH	TN	WV
Same Institution	59.3%	65.9%	62.8%	57.9%
Same State	72.2%	82.2%	76.2%	66.7%
4-State Region	73.8%	82.9%	77.2%	68.5%

	KY	OH	TN	WV
1999 Cohort				

Bachelors Degree After 6 Years

	KY	OH	TN	WV
Number of Students	13788	37910	14929	10680
Same Institution	41.3%	53.6%	42.1%	37.4%
Same State	44.8%	58.4%	45.9%	40.6%
4-State Region	45.6%	58.7%	46.2%	41.0%

Retained 2nd Fall

	KY	OH	TN	WV
Same Institution	73.1%	76.6%	72.6%	71.1%
Same State	83.3%	91.2%	84.6%	79.9%
4-State Region	85.7%	91.7%	85.5%	81.5%

Retained 3rd Fall

	KY	OH	TN	WV
Same Institution	62.1%	66.3%	62.8%	58.4%
Same State	74.7%	83.0%	81.7%	67.5%
4-State Region	77.7%	83.7%	82.7%	69.3%

These results are quite consistent with those obtained in the original bilateral data exchange effort between Ohio and Kentucky. Graduation rates are boosted about three or four percentage points when the home state is taken as the unit of analysis, and about one percent or less when the tracking region is expanded to encompass the three bordering

states. But the pattern varies by state. Kentucky, for example, shows more than twice the "value added" in terms of boosted graduation rates of expanding the tracking region to its border states.

Though the increases remain modest, somewhat more interstate movement is apparent in retention rates. Here within-state retention is increased by an average of ten percentage points and retention within the four-state region between one half and three percentage points. But differences in these patterns among states are even more apparent. West Virginia, for example, shows a good deal less within-state inter-institutional mobility than the other three states, but has somewhat higher levels of cross-state mobility. Ohio shows a good deal of within-state transfer, but relatively little cross-state activity. And Kentucky has the highest—though still relatively limited—rates of cross-border transfer.[1]

Because it would be reasonable to assume that institutions located near state borders would have somewhat higher rates of interstate degree-granting and transfer, a subset of each cohort drawn from such institutions was tracked using the same procedures used for the full cohort.[2] Results are provided in the table below:

Summary Tables from Four-State Match--Annual Border Institutions

	KY	OH	TN	WV
1998 Cohort				
Bachelors Degree After 7 Years				
Number of Students	5115	8200	4300	2263
Same Institution	40.5%	54.9%	40.2%	38.2%
Same State	44.9%	60.9%	45.2%	41.7%
4-State Region	46.1%	61.8%	45.6%	42.6%
Retained 2nd Fall				
Same Institution	69.8%	77.4%	72.8%	72.2%
Same State	77.5%	90.4%	83.2%	82.3%
4-State Region	80.7%	91.5%	83.8%	85.2%
Retained 3rd Fall				
Same Institution	58.7%	67.3%	59.4%	59.4%
Same State	69.3%	81.4%	71.9%	67.8%
4-State Region	73.2%	83.0%	72.6%	70.5%

---

[1] Some of these difference are undoubtedly artifacts of geography as Kentucky is unique in that it shares a border with all three other states.

[2] The institutions were Austin Peay State University (TN), Middle Tennessee State University (TN), Tennessee Tech University (TN), Murray State University (KY), Western Kentucky University (KY), Northern Kentucky University (KY), University of Cincinnati (OH), Ohio University (OH), and Marshall University (WV).

	KY	OH	TN	WV
1999 Cohort				
Bachelors Degree After 6 Years				
Number of Students	5239	7978	4349	2061
Same Institution	39.3%	53.0%	38.1%	36.6%
Same State	42.2%	57.3%	41.4%	39.0%
4-State Region	43.1%	58.0%	41.6%	39.6%
Retained 2nd Fall				
Same Institution	72.7%	77.3%	69.9%	75.7%
Same State	80.8%	89.3%	71.8%	82.8%
4-State Region	84.0%	90.2%	72.6%	85.7%
Retained 3rd Fall				
Same Institution	62.6%	67.3%	60.2%	61.7%
Same State	75.2%	80.7%	76.1%	69.3%
4-State Region	79.5%	82.1%	77.0%	73.1%

As is apparent, Bachelors degree completion rates for border institutions are slightly, but not appreciably, enhanced by including other states. Somewhat larger increases are apparent for retention rates, with third-fall retention bumped by three or four additional percentage points in Kentucky and West Virginia. But these increases remain fairly modest in the light of initial expectations and are about one percentage point or less in Ohio and Tennessee.

One interesting point of comparison to understand interstate mobility in completions was generated by a parallel NCHEMS project using data from the National Student Clearinghouse (NSC). Because many institutions report enrollment and degree completion data to the NSC each year, the organization possesses a fairly robust national SUR database. Data reported to NSC were never intended to be used for cohort tracking, so have many limitations when used for this purpose. But to ascertain their potential, NCHEMS worked with NSC staff to construct GRS-type cohorts for five states—Kentucky, Minnesota, Nevada, New York, and Washington. Statistics were generated on retention and completion at the same institution, at an institution in the same state, and at an institution in any state. This meant that it was possible to obtain same-institution and same-state baccalaureate degree completion rates for 1998 cohort members in Kentucky to compare with NSC rates. Results are provided in the table below:

Multi-State Longitudinal Graduation Rates for Kentucky

4-Year Starters Six Years Later

	1998 Cohort		1999 Cohort	
	State Sharing	**NSC**	**State Sharing**	**NSC**
Number of Students	13784	12563	13788	10180
Graduated Same Institution	42.6%	39.2%	41.3%	40.9%
Graduated In State	46.9%	42.5%	44.9%	43.7%
Graduated Anywhere Tracked	48.1%	47.9%	45.7%	49.3%

	1998 Cohort		1999 Cohort	
	State Sharing	**NSC**	**State Sharing**	**NSC**
Number of Students	13784	12563	13788	10180
Graduated Same Institution	42.6%	43.2%	41.3%	44.3%
Graduated In State Gain	4.3%	3.3%	3.6%	2.8%
Graduated Anywhere Gain	1.2%	5.4%	0.8%	5.6%

As is apparent, the six-year completion rates for four-year institutions in Kentucky within the same institution and same state are within about three percentage points using the two methodologies. Considering the substantial limitations of the NSC data and cohort selection methodologies, this is quite good.

Implications and Lessons Learned. Overall, the data exchange project demonstrated that it is feasible to exchange records drawn from multiple state SUR systems to create more accurate estimates of cohort retention and completion. But the value added of doing this within some regions may be modest. Comparison with similar data drawn from the NSC suggests that nationwide tracking would be required for this value-added to be appreciable. Because only one region was tested, however, it remains possible that some multi-state groupings—for example, states in New England—would show a higher value added.

From a procedural and technical standpoint, moreover, the project has a number of implications for future work of this kind:

- Third-party organizations are essential for such a project. While it is in principal possible for one of the state partners to match records from multiple states, this allows the state performing this role direct access to records from other states. This is avoided by using a third party. In addition, using a third-party organization with substantial experience in record matching and data analysis is more efficient, though it may involve some direct costs. This also allows a standard methodology to be developed by the organization that can be applied to multiple states. Similarly, we found that generic agreements to share data across multiple states do not work well. Because each state constitutes its own unique legal environment, exchange agreements are best crafted for each state individually with an administrative third-party organization. NCHEMS played this role in this demonstration. Not only did this allow the preparation of bilateral agreements between NCHEMS and each participating state, but it also allowed NCHEMS to be the keeper of the encryption routines for all student identifiers, independent of both the participating states and the third-party data administrator.

- Data quality is likely to be an issue. Data were of uneven quality across states. For example, while all state file submissions had some duplicate records, one state file had large numbers of duplicate records all with slightly different data values for the same enrollment period. These were likely generated when a record was updated, but not overwritten. These situations required individual decisions about which record should be used. In addition, several students were reported as first-time, full-time students by more than one institution. These required similar individual decisions about which record to use or whether to use the case at all. It is likely that similar data issues will continue to arise with any exchange and a set of standard decision rules about how to handle them should be evolved.

- Political issues trump data issues and will slow the effort down. The extended duration of this project was the result of a number of factors that may be idiosyncratic. The biggest of these was the emergence of the NCES national unit-record data system proposal, which heightened the visibility of the privacy issue and polarized the higher education community. But even if this had not occurred, it would probably have been necessary to meet with institutional representatives in some participating states to gain their trust and cooperation. In the case of the four-state exchange, the need for this varied substantially across states, and this too should be anticipated. But the exchange can only proceed at the pace of the slowest partner. This means that substantial lead time needs to be built into any such project.

- State ability and interest in using the resulting data may be limited. Because of budget shortfalls, many states are cutting back analytical staff. As a result they may be in a constrained position to actually do much with the resulting data, or manage an exchange in the first place. Indeed, the data files have now been in the hands of the participating states for six months as of this writing and they have

not been accessed by local analysts. All reports off the data files have been produced by the third party data administrator.

Although the value-added of multi-state data exchange was not in this case large, the ultimate success of this effort suggests that it is yet another tool that policymakers can use to explore the dynamics of an increasingly volatile higher education enrollment picture. To maximize the utility of this tool, further efforts should be undertaken to develop standard documentation and procedures to govern them, through such organizations as the State Higher Education Executive Officers (SHEEO) and the Postsecondary Education Standards Council (PESC).

References

Adelman, Clifford (2006). *The Toolbox Revisited: Paths to Degree Completion from High School Through College*. Washington, DC: Office of Vocational and Adult Education, U.S. Department of Education.

Adelman, Clifford (2007). The Post-Matriculation Geographic Mobility of Traditional-Age Undergraduates*: Where Do They Come From? Where Do They Go?* Washington, DC: Institute for Higher Education Policy (IHEP).

Ewell, P. T. and Boeke, M (2007). *Critical Connections: Liking States' Unit Record Systems to Track Student Progress*. Indianapolis, IN: Lumina Foundation for Education.

Ewell, P. T., Paulson, K. P., and Schild, P. R. (2003). *Following the Mobile Student: Can We Develop the Capacity for a Comprehensive Database to Assess Student Progression?* Indianapolis, IN: Lumina Foundation for Education.

Appendix A

Selected Results of Ohio/Kentucky Data Match

Students beginning collegiate study in Ohio were examined to determine if they enrolled or earned a degree at a Kentucky institution. This analysis involved a total of 77,709 students entering college from the fall of 1998 through the fall of 2002, tracked through the fall of 2004. Of these Ohio starters, 91 (0.1%) had earned degrees from Kentucky institutions within six years and 241 (0.3%) had enrolled in Kentucky institutions during the peak enrollment year. As expected, these proportions were somewhat higher for students beginning their study in Ohio in the Cincinnati metropolitan region, with 0.7% earning degrees and 1.8% enrolling in Kentucky during the peak enrollment year.

The parallel analysis involved 26,418 students beginning collegiate study in at public institutions in Kentucky during the same time period. Of these Kentucky starters, 136 (0.5%) had earned degrees from Ohio institutions within six years and 294 (1.1%) had enrolled in Ohio institutions during the peak enrollment year. Mirroring the Ohio results for the Cincinnati metropolitan region, 1.9% of students beginning their study at Northern Kentucky University (the only public institution in Kentucky in this region operating continuously during this period) completing degrees in Ohio within six years, and 5.4% enrolling in Ohio during the peak enrollment year.

These are admittedly relatively small proportions of total completion. Among all Ohio starters, for example, 30.2% completed a baccalaureate or associate degree in six years from the institution at which they began and a further 4.0% earned a degree from another Ohio institution. And among all Kentucky starters, 28.6% completed a baccalaureate or associate degree in six years from the institution at which they began and a further 7.1% earned a degree from another Kentucky institution. But in border regions, those who go to another institution and complete degrees, though small in absolute numbers, are quite likely to cross state lines. For instance, more than 20% of Ohio starters in the Cincinnati metropolitan region who earned degrees at institutions other than where they started did so in Kentucky. For those beginning at Northern Kentucky University, more than 35% of those who earned degrees at institutions other than where they started did so in Ohio. These analyses also reveal that cross-border flows in border regions can be asymmetric: more students appear to begin in Kentucky and enroll in Ohio than the reverse.

Furthermore, the direct matching procedure appeared to capture enrollments and completions that are not able to be located in any other way. For example, of 24,601 students starting in Ohio in 1998 who had neither re-enrolled nor graduated from an Ohio institution by fall 2004, 164 (0.67%) were identified as enrolled in Kentucky by either the matching procedure or the National Student Clearinghouse but almost 30% of these were added uniquely by the matching procedure and not identified by the Clearinghouse.

These results suggest that the matching procedure adds some value in increasing the accuracy of graduation and persistence rate reporting. And the boost in added value for border regions in particular metropolitan areas suggests that overall state results may be affected more and more as all states bordering a given state join a consortium.

Appendix B

Specifications and File Layout for Exchanged Data Elements

NCHEMS/Lumina Multi-State Study

Scope of Work for 3rd Party Data Match

Input Files

Each of the five participating states will provide the third party vendor with data for all enrollments and completions for each academic term from Summer 1998 to Summer 2005 – totaling approximately 20 million records. The variables in the submission are included in the proposed file layout (attached). A unique study ID number will be assigned to each record prior to their submission to the vendor. [*NCHEMS will provide the study number to each of the states.*]

Data Matching

Using the uniquely-aligned study ID numbers from each state's file, the vendor will identify the matched records for every academic term in the combination of the other 4 states' files. For **each** instance of a match, the vendor will append the following records to the appropriate record in the state specific files:

- Campus
- Year
- Term
- Student level
- Credit hours enrolled
- Field of study
- Degree of study
- Level of degree/certificate completed
- State of campus

Return Data Files

Once the matched data have been appended to each of the individual state files, the vendor will strip the encrypted student identifier and return the files to the states.

File Layout

Input File	Field Definition	Characters	Data Type	Matched Records	Return File
Study ID Number – Used for Matches	#########	9	Alpha Numeric		Study ID Deleted
Campus	IPEDS Unit ID	6	Alpha Numeric	X	X
Year	####	4	Alpha Numeric	X	X
Term	1 = Fall, 2 = Spring, 3 = Summer	1	Alpha Numeric	X	X
Year of Birth	State Definition (optional)	4	Alpha Numeric		X
Gender	State Definition (optional)	1	Alpha Numeric		X
Race/Ethnicity	State Definition (optional)	1	Alpha Numeric		X
State of Origin	FIPS Code	2	Alpha Numeric		X
County of Origin	FIPS Code	5	Alpha Numeric		X
Student Level (Admission Area)	1 = First-Time, 2 = Freshmen, 3 = Sophomore, 4 = Junior, 5 = Senior	1	Alpha Numeric	X	X
Credit Hours (FT threshold = 12)	1 = Full-Time, 2 = Part-Time	1	Alpha Numeric	X	X
Current field of study	6 Digit CIP Code	6	Alpha Numeric	X	X
Degree Field of Study	6 Digit CIP Code	6	Alpha Numeric	X	X
Level of Degree/ Certificate Completed	IPEDS Award Level Code (1=Less than 1-Year Certificate/Diploma, 2=Less than 2-Year Certificate/ Diploma, 3=Associates Degree, 4=Certificate/Diploma Greater than 2-Year and Less than 4-Year, 5=Bachelor's Degree	1	Alpha Numeric	X	X
State of School	FIPS Code	2	Alpha Numeric	X	X
State Specific Data	State Defined (Optional)	Up to 15	Alpha Numeric		X

Appendix C

Example of Data Exchange Agreement

AUTHORIZATION TO ACT AS AGENT IN CONDUCTING A RESEARCH STUDY USING STUDENT UNIT RECORD DATA

Background and Purpose
[The Kentucky Council on Postsecondary Education (CPE)] serves as the coordinating body for higher education in the state of [Kentucky] and desires to conduct a study of the interstate mobility of students enrolled in the state's public higher education institutions. The National Center for Higher Education Management Systems (NCHEMS) has substantial experience in higher education research, and is authorized to do business in the state of [Kentucky]. [CPE] desires to authorize NCHEMS to act as its agent in assembling data on postsecondary enrollments in public institutions of higher education in the states of Indiana, Ohio, Tennessee, and West Virginia in accordance with the terms set forth hereinafter.

Data drawn from electronic records in these states will be used to produce aggregate statistics for [CPE's] use in evaluating the effectiveness of its education and training programs. The data will not be used to make decisions that affect the rights, benefits or privileges of any individuals.

Justification and Authority
In accordance with the Federal Family Educational Rights and Privacy Act (FERPA), and in particular 34 CFR 99.31 (a)(3)(iv) and 99.35, [CPE] is a state educational authority that for the limited purpose of this Agreement, designates NCHEMS as its authorized representative for the purpose of assembling data for a specific evaluation of publicly-funded education and training programs. Procedures used in this agreement will be governed by FERPA, the Gramm-Leach Bliley Act (GLB), all applicable state laws, and the [CPE's] established Data Policy.

Implementation
Operational procedures to carry out the study are as follows:
1. NCHEMS will provide specifications for a data file to be prepared by [CPE] containing a limited number of data elements for students enrolled in [Kentucky's] public higher education institutions. These specifications will include a file layout indicating data element types and lengths, and a definition of each data element (see Attachment A).
2. NCHEMS will supply [CPE] with a method for assigning a unique Study Identification Number to each qualifying record. NCHEMS will supply the same method for assigning a Study Identification Number to each of the other four participating states.
3. [CPE] will produce a data file containing the specified data elements for students meeting the criteria for inclusion in the study. The Study Identification Number will be substituted for the student's Social Security Number (SSN) and will be appended to the record in place of the SSN.

4. [CPE] will deliver this data file to a Third-Party Vendor identified by NCHEMS in a secure fashion using [specify method here]. The other four participating states will deliver similar data files to the Third-Party Vendor in a similar fashion.
5. The Third-Party Vendor will use Study Identification Numbers in the file supplied by [CPE] to search for matching enrollment information in the data files supplied by the other four participating states. Each time a match is found, selected data elements that describe the enrollment will be appended to the record.
6. The Third-Party Vendor will strip all unique identifiers from the resulting augmented data file and return the de-identified file to [CPE] in a secure fashion using [specify method here].
7. [CPE] will conduct analyses on this de-identified file to calculate selected statistics on out of state enrollments using a template reporting format supplied by NCHEMS. Following IPEDS conventions (see NCES Statistical Standard 4-2-10), these analyses will ensure that each cell contains at least three cases.
8. [CPE] will maintain the data file in a secure environment until the conclusion of the study and will destroy the data file within six months of the termination of the study. Destruction will be by shredding, burning, or electronic erasure.

Duration
This Agreement will remain in effect for one year from the date of signing. At the end of that period, this Agreement may be renewed for a period of up to two additional years if both parties approve that renewal and certify that 1) the program will be conducted without change and 2) the program has been conducted in compliance with this Agreement.

Fees
No charges or fees will be charged by either party. This study is being supported through a grant to NCHEMS by the Lumina Foundation for Education.

Records usage, duplication and re-disclosure restrictions
1. NCHEMS agrees not to re-disclose data received from [CPE] to any outside party or agency except as permitted under the terms of this Agreement.
2. NCHEMS agrees that the data will not be used to make a decision about the rights, benefits, or privileges of any individual identified in the course of the study, but will be used strictly for the purposes of analysis and evaluation and that any end product will be in the form of aggregate statistical data without any personal identifiers.
3. Information exchanged will not be duplicated unless essential to the conduct of the study. All stipulations of this Agreement will apply to any duplication of records or files.
4. [CPE] agrees not to re-disclose the data to any organization that is not a party to this Agreement including institutions and other state agencies. Any reports prepared by [CPE] using these data will present aggregate data only.

Program Records; Privacy Rights; and Public Access
  1. NCHEMS staff or its designees will not personally view the contents of the data files supplied by [CPE].
  2. If the purpose for which files are sent does require NCHEMS or its designees, or [CPE] personnel to print, display, or otherwise personally view the contents of the file, for example to avoid or correct a malfunction of a data matching process, the personnel shall do so in a manner that prevents the disclosure of the contents of the file by persons not involved in the process.
  3. The parties agree to provide written instructions regarding sub-paragraphs one and two above to affected employees.
  4. Each electronic file provided by [CPE] and in the possession of NCHEMS or its designees that contains individual student record data, and each hard copy of such data (if applicable), shall be stored in a secure location, such as a locked desk or file cabinet, except when in use for the purposes for which it was provided. Automated records shall be stored in secured computer facilities with strict ADP controls.
  5. Under no circumstances shall either party disclose any personally identifiable information about a student supplied or generated under this Agreement to any outside party.

Termination

This Agreement may be terminated without cause by [CPE]. Upon termination, NCHEMS agrees cease work, to destroy all data generated to date, and to provide assurances to [CPE] that it has done so.

Contact Information

Name, Title, Address, Phone, Fax, and email for [CPE].
Name, Title, Address, Phone, Fax, and email for NCHEMS.

Notices

All notices under this Agreement must be provided to the authorized officials identified above.

Approval

The undersigned authorized officials of [CPE] and NCHEMS commit their respective agencies to the terms of the MOU.

_____   _____
[CPE] sign   Date

_____   _____
NCHEMS sign   Date

Appendix D

Documented Methodology for Producing Reports

NCHEMS Graduation & Retention Methodology
4-year public schools by state

The Graduation & Retention report provides a more detailed look at the cohort's progress than seen in the Cohort Tracking report by reporting term-by-term data and by breaking out the degree and retention categories by control of the institution. This report also focuses on retention rather than enrollment.
The 1998 & 1999 cohorts (defined as first-time, full-time, beginners in either the summer or fall term of the cohort year) will be followed through the next several academic years reporting on graduation and retention as described in the "**Report Columns and Definitions**" section.

Like the Cohort Tracking report, the Graduation & Retention report uses the same data set, the same cohort definitions, and the same views for term data. The details are reiterated in the "**Same Foundation as Cohort Tracking Report**" of this document for ease of reference.

The SQL queries used to generate the data supplied in this report is included in the section "**SQL Queries to Populate the Report.**"

The more detailed look provided by the Graduation & Retention report uncovered some additional anomalies in the data that may influence the understanding of the numbers provided in the Cohort Tracking report.

- **Duplicate enrollment records** – many students have identical, or nearly identical records, submitted by the same state for the same campus within the same year and term. See "**Appendix: Duplicate enrollment records.**"
- **Multiple enrollment records** – many students legitimately are enrolled at more than one institution in a given year/term
- **Multiple cohort records** – several students were reported as first-time, full-time students by more than one school. See "**Appendix: Multiple Cohort Records.**"

These anomalies prompted the data on the Graduation & Retention report retention numbers separately for those who had only one enrollment record in the specified term from those who had multiple enrollment records in the specified term. These also explain why the sum of the break-outs by institution control can be greater than the number of unique students represented.

Report Columns and Definitions

Condition	
SUMMARY	
TOTAL Completed prior or not enrolled	The number of unique students in the cohort having no enrollment record this term.
Retained (multiple/duplicate enrollment)	The number of unique students in the cohort having more than one enrollment record this term (duplicate records submitted for the same institution or records from more than one institution) who have an enrollment record that does not have a matching (same institution) Bachelor's completion record. Students who were enrolled this term and completed with a Bachelor's degree from the same institution during this term are not included as retained.
Retained (single enrollment)	The number of unique students in the cohort having only one enrollment record this term whose enrollment record does not have a matching (same institution) Bachelor's completion record. Students who were enrolled this term and completed with a Bachelor's degree this term are not included as retained. Students who were enrolled this term and completed with a Bachelor's degree from the same institution during this term are not included as retained.
TOTAL RETENTION (Unique students enrolled no Bachelor's)	The total number of unique students in the cohort retained (enrolled but did not complete with a Bachelor's degree from the same institution) this term.
TOTAL GRADUATES (Unique students with Bachelor's)	The number of unique students in the cohort completing this term with a Bachelor's degree.
Cumulative Graduates	*The cumulative number of unique students in the cohort who have completed with a Bachelor's degree as of this term.*
Correction Number (students graduated & retained)	The number of unique students in the cohort who were counted both as a graduate and as being retained. Some students were enrolled in more than one institutionduring a term and graduated from one institution while being retained by another institution.
TOTAL COHORT (Unique students)	Number of unique students who were first-time, full-time, students at a 4-year public institution in the summer or fall term of the cohort year and state indicated.

ENROLLMENT DETAIL	
Single enrollment record	
Single enrollment: TOTAL RETENTION (unique students)	Number of unique students in the cohort having only one enrollment record this term who were retained (did not graduate with a Bachelor's degree this term).
Single enrollment: Retained at enrolling institution	Number of unique students in the cohort having only one enrollment record this term who were retained (did not graduate with a Bachelor's degree this term) in their enrolling/cohort institution.
Single enrollment: Retained in same state: 2 year/less than 4	Number of unique students in the cohort having only one enrollment record this term who were retained (did not graduate with a Bachelor's degree this term) in the same state as their enrolling/cohort institution in a school whose IPEDS control designation is "At least 2 but less than 4 years".
Single enrollment: Retained in same state: 4 or more years	Number of unique students in the cohort having only one enrollment record this term who were retained (did not graduate with a Bachelor's degree this term) in the same state as their enrolling/cohort institution in a school whose IPEDS control designation is "Four or more years".
Single enrollment: Retained in same state: unknown	Number of unique students in the cohort having only one enrollment record this term who were retained (did not graduate with a Bachelor's degree this term) in the same state as their enrolling/cohort institution in a school whose IPEDS control designation could not be determined most likely due to an incorrect UNITID in the state-supplied data.
Single enrollment: Retained in [Kentucky/Tennessee/Ohio/West Virginia]: 2 year/less than 4	Number of unique students in the cohort having only one enrollment record this term who were retained (did not graduate with a Bachelor's degree this term) in the state indicated [Kentucky/Tennessee/Ohio/West Virginia] in a school whose IPEDS control designation is "At least 2 but less than 4 years".
Single enrollment: Retained in [Kentucky/Tennessee/Ohio/West Virginia]: 4 or more years	Number of unique students in the cohort having only one enrollment record this term who were retained (did not graduate with a Bachelor's degree this term) in the state indicated [Kentucky/Tennessee/Ohio/West Virginia] in a school whose IPEDS control designation is "Four or more years".
Single enrollment: Retained in [Kentucky/Tennessee/Ohio/West Virginia]: unknown	Number of unique students in the cohort having only one enrollment record this term who were retained (did not graduate with a Bachelor's degree this term) n the state indicated [Kentucky/Tennessee/Ohio/West Virginia] in a school whose IPEDS control designation could not be determined most likely due to an incorrect UNITID in the state-supplied data.
TOTAL RETENTION (Calculated – may not match unique students)	*The calculated total of students retained based on the sum of the retention numbers above. This number may be higher than the "Total Retention (unique students)" above because of the cohort containing dual enrollment records for several students. For example, individual students may show up in counts for being retained at the enrolling institution and for being retained at another institution in the same state.*

ENROLLMENT DETAIL	
Multi/dupe enrollment records	
Multi enrollment: TOTAL RETENTION (unique students)	The number of unique students having more than one enrollment record this term (duplicate records submitted for the same institution or records from more than one institution) who were retained (did not graduate with a Bachelor's degree this term.)
Multi enrollment: Retained at enrolling institution	Number of unique students in the cohort having more than one enrollment record this term (duplicate records submitted for the same institution or records from more than one institution) who were retained (did not graduate with a Bachelor's degree this term) in their enrolling/cohort institution.
Multi enrollment: Retained in same state: 2 year/less than 4	Number of unique students in the cohort having more than one enrollment record this term (duplicate records submitted for the same institution or records from more than one institution) who were retained (did not graduate with a Bachelor's degree this term) in the same state as their enrolling/cohort institution in a school whose IPEDS control designation is "At least 2 but less than 4 years".
Multi enrollment: Retained in same state: 4 or more years	Number of unique students in the cohort having more than one enrollment record this term (duplicate records submitted for the same institution or records from more than one institution) who were retained (did not graduate with a Bachelor's degree this term) in the same state as their enrolling/cohort institution in a school whose IPEDS control designation is "Four or more years".
Multi enrollment: Retained in same state: unknown	Number of unique students in the cohort having more than one enrollment record this term (duplicate records submitted for the same institution or records from more than one institution) who were retained (did not graduate with a Bachelor's degree this term) in the same state as their enrolling/cohort institution in a school whose IPEDS control designation could not be determined most likely due to an incorrect UNITID in the state-supplied data.
Multi enrollment: Retained in [Kentucky/Tennessee/Ohio/West Virginia]: 2 year/less than 4	Number of unique students in the cohort having more than one enrollment record this term (duplicate records submitted for the same institution or records from more than one institution) who were retained (did not graduate with a Bachelor's degree this term) in the state indicated [Kentucky/Tennessee/Ohio/West Virginia] in a school whose IPEDS control designation is "At least 2 but less than 4 years".
Multi enrollment: Retained in [Kentucky/Tennessee/Ohio/West Virginia]: 4 or more years	Number of unique students in the cohort having more than one enrollment record this term (duplicate records submitted for the same institution or records from more than one institution) who were retained (did not graduate with a Bachelor's degree this term) in the state indicated [Kentucky/Tennessee/Ohio/West Virginia] in a school whose IPEDS control designation is "Four or more years".
Multi enrollment: Retained in [Kentucky/Tennessee/Ohio/West Virginia]: unknown	Number of unique students in the cohort having more than one enrollment record this term (duplicate records submitted for the same institution or records from more than one institution) who were retained (did not graduate with a Bachelor's degree this term) in the state indicated [Kentucky/Tennessee/Ohio/West Virginia] in a school whose IPEDS control designation could not be determined most likely due to an incorrect UNITID in the state-supplied data.

TOTAL RETENTION (Calculated - will not match unique students)	*The students with multiple enrollment records may have been enrolled in more than one campus this term or they may have had more than one record submitted for the same campus; each student/campus combination was only counted once so the sum of the multi enrollment retention will not necessarily be more than double the multi enrollment total retention number. However, the total could be more than double; in at least one case, a student was enrolled in three schools during a term making the total more than double the unique students.*

GRADUATION DETAIL	
Associate Degree (Unique students)	
Graduation: Associate's at enrolling institution	Number of unique students in the cohort who earned an Associate's degree in this term from their enrolling/cohort institution.
Graduation: Associate's in same state: 2 year/less than 4	Number of unique students in the cohort who earned an Associate's degree in this term in the same state as their enrolling/cohort institution in a school whose IPEDS control designation is "At least 2 but less than 4 years".
Graduation: Associate's in same state: 4 or more years	Number of unique students in the cohort who earned an Associate's degree in this term in the same state as their enrolling/cohort institution in a school whose IPEDS control designation is "Four or more years".
Graduation: Associate's in same state: unknown	Number of unique students in the cohort who earned an Associate's degree in this term in the same state as their enrolling/cohort institution in a school whose IPEDS control designation could not be determined most likely due to an incorrect UNITID in the state-supplied data.
Graduation: Associate's in [Kentucky/Tennessee/Ohio/West Virginia]: 2 year/less than 4	Number of unique students in the cohort who earned an Associate's degree in this term in the state indicated [Kentucky/Tennessee/Ohio/West Virginia] in a school whose IPEDS control designation is "At least 2 but less than 4 years".
Graduation: Associate's in [Kentucky/Tennessee/Ohio/West Virginia]: 4 or more years	Number of unique students in the cohort who earned an Associate's degree in this term in the state indicated [Kentucky/Tennessee/Ohio/West Virginia] in a school whose IPEDS control designation is "Four or more years".
Graduation: Associate's in [Kentucky/Tennessee/Ohio/West Virginia]: unknown	Number of unique students in the cohort who earned an Associate's degree in this term in the state indicated [Kentucky/Tennessee/Ohio/West Virginia] in a school whose IPEDS control designation could not be determined most likely due to an incorrect UNITID in the state-supplied data.
Graduation: Total Associate's degrees	The total number of Associate's degrees earned by students in the cohort. This number may be greater than the number of unique students earning Associate's degrees in this term because students may have earned an Associate's degree at more than one institution in this term. The difference also may be attributable to several students having dual enrollment records in the cohort.
Graduation: Unique Students earning Associate's degrees	The number of unique students in the cohort earning Associate's degrees this term.

GRADUATION DETAIL	
Bachelor's Degree (Unique students)	
Graduation: Bachelor's at enrolling institution	Number of unique students in the cohort who earned a Bachelor's degree in this term from their enrolling/cohort institution.
Graduation: Bachelor's in same state: 2 year/less than 4	Number of unique students in the cohort who earned a Bachelor's degree in this term in the same state as their enrolling/cohort institution in a school whose IPEDS control designation is "At least 2 but less than 4 years".
Graduation: Bachelor's in same state: 4 or more years	Number of unique students in the cohort who earned a Bachelor's degree in this term in the same state as their enrolling/cohort institution in a school whose IPEDS control designation is "Four or more years".
Graduation: Bachelor's in same state: unknown	Number of unique students in the cohort who earned a Bachelor's degree in this term in the same state as their enrolling/cohort institution in a school whose IPEDS control designation could not be determined most likely due to an incorrect UNITID in the state-supplied data.
Graduation: Bachelor's in [Kentucky/Tennessee/Ohio/West Virginia]: 2 year/less than 4	Number of unique students in the cohort who earned a Bachelor's degree in this term in the state indicated [Kentucky/Tennessee/Ohio/West Virginia] in a school whose IPEDS control designation is "At least 2 but less than 4 years".
Graduation: Bachelor's in [Kentucky/Tennessee/Ohio/West Virginia]: 4 or more years	Number of unique students in the cohort who earned a Bachelor's degree in this term in the state indicated [Kentucky/Tennessee/Ohio/West Virginia] in a school whose IPEDS control designation is "Four or more years".
Graduation: Bachelor's in [Kentucky/Tennessee/Ohio/West Virginia]: unknown	Number of unique students in the cohort who earned a Bachelor's degree in this term in the state indicated [Kentucky/Tennessee/Ohio/West Virginia] in a school whose IPEDS control designation could not be determined most likely due to an incorrect UNITID in the state-supplied data.
Graduation: Total Bachelor's degrees	The total number of Bachelor's degrees earned by students in the cohort. This number may be greater than the number of unique students earning Bachelor's degrees in this term because students may have earned as Bachelor's degree at more than one institution in this term. The difference also may be attributable to several students having dual enrollment records in the cohort.
Graduation: Unique Students earning Bachelor's degrees	The number of unique students in the cohort earning Bachelor's degrees this term.

Prior to this analysis, all data submitted by the four states was loaded into one database table rawdataapril30_sorted which has the following fields:
- id
- field1
- field2
- field3
- field4
- field5
- field6
- field7
- field8
- field9
- field10
- field11
- field12
- field13
- field14
- field15
- field16
- field17
- field18
- TermSort

To complete this analysis, several views were created against the table rawdataapril30_sorted:

- Views for all records submitted by all states for each year/term
  - v_NCHEMS_1998_aSpring
  - v_NCHEMS_1998_bSummer
  - v_NCHEMS_1998_cFall
  - v_NCHEMS_1999_aSpring
  - v_NCHEMS_1999_bSummer
  - v_NCHEMS_1999_cFall
  - v_NCHEMS_2000_aSpring
  - v_NCHEMS_2000_bSummer
  - v_NCHEMS_2000_cFall
  - v_NCHEMS_2001_aSpring
  - v_NCHEMS_2001_bSummer
  - v_NCHEMS_2001_cFall
  - v_NCHEMS_2002_aSpring
  - v_NCHEMS_2002_bSummer
  - v_NCHEMS_2002_cFall

- o v_NCHEMS_2003_aSpring
- o v_NCHEMS_2003_bSummer
- o v_NCHEMS_2003_cFall
- o v_NCHEMS_2004_aSpring
- o v_NCHEMS_2004_bSummer
- o v_NCHEMS_2004_cFall
- o v_NCHEMS_2005_aSpring
- o v_NCHEMS_2005_bSummer
- o v_NCHEMS_2005_cFall
- o v_NCHEMS_2006_aSpring
- o v_NCHEMS_2006_bSummer
- o v_NCHEMS_2006_cFall
- View to represent each cohort analyzed
  - o v_cohort_KY4yrPublic_NCHEMS_1998_zSummerFall
  - o v_cohort_KY4yrPublic_NCHEMS_1999_zSummerFall
  - o v_cohort_OH4yrPublic_NCHEMS_1998_zSummerFall
  - o v_cohort_OH4yrPublic_NCHEMS_1999_zSummerFall
  - o v_cohort_TN4yrPublic_NCHEMS_1998_zSummerFall
  - o v_cohort_TN4yrPublic_NCHEMS_1999_zSummerFall
  - o v_cohort_WV4yrPublic_NCHEMS_1998_zSummerFall
  - o v_cohort_WV4yrPublic_NCHEMS_1999_zSummerFall

A sample of the year/term views is listed below. Each view differs only by name and by values in the where clause for year and term (field4 and field5):

```
CREATE VIEW dbo.v_NCHEMS_1998_aSpring
AS
SELECT    id,
     field1 AS RecordType,
     field2 AS StudyID,
     CASE (field3) WHEN '' THEN 'nocamp' ELSE field3 END AS Campus,
     field4 AS [Year],
     field5 AS Term,
--       CASEfield6 AS BirthYear,

     case
     when len(field6) = 0 then null
     else field6
     end 'BirthYear',

     field7 AS Gender,
--     field8 AS RaceEthnicity,
     case
     when len(field8) = 0 then null
     else field8
     end 'RaceEthnicity',

     field9 AS StateOfOrigin,
     field10 AS CountyOfOrigin,
--     field11 AS StudentLevel,
     case
     when len(field11) = 0 then null
     else field11
     end 'StudentLevel',

--       field12 AS CreditHours,
     case
```

```
            when len(field12) = 0 then null
            else field12
            end 'CreditHours',

            field13 AS CurrentFieldOfStudy,
            field14 AS DegreeFieldOfStudy,
--          field15 AS LevelOfDegree,
            case
            when len(field15) = 0 then null
            else field15
            end 'LevelOfDegree',
--          field16 AS SchoolState,
            case field16
              when 36 then 39
              else field16
            end 'SchoolState',
              field17 AS StateSpecific,
            field18 AS EndOfFileMarker,
            TermSort
FROM        dbo.rawdataapril30_sorted
WHERE       (field4 = 1998) AND (field5 = 2)
```

The examples in the remainder of the document follow the 1998 KY 4yr public cohort and the queries used to populate this report. The year and state values were modified accordingly to obtain the data for other cohorts and states.

The cohort view is similarly constructed (has the same fields); however the filters used differ and an additional table is joined in order to limit the cohort to those enrolled in 4-year public schools.

```
CREATE VIEW dbo.v_cohort_KY4yrPublic_NCHEMS_1998_zSummerFall
AS
SELECT    id,
          field1 AS RecordType,
          field2 AS StudyID,
          CASE (field3) WHEN '' THEN 'nocamp' ELSE field3 END AS Campus,
          field4 AS [Year],
          field5 AS Term,
--        CASEfield6 AS BirthYear,

          case
          when len(field6) = 0 then null
          else field6
          end 'BirthYear',

          field7 AS Gender,
--        field8 AS RaceEthnicity,
          case
          when len(field8) = 0 then null
          else field8
          end 'RaceEthnicity',

          field9 AS StateOfOrigin,
          field10 AS CountyOfOrigin,
--        field11 AS StudentLevel,
          case
          when len(field11) = 0 then null
          else field11
          end 'StudentLevel',

--          field12 AS CreditHours,
          case
          when len(field12) = 0 then null
          else field12
          end 'CreditHours',
```

30

```sql
        field13 AS CurrentFieldOfStudy,
        field14 AS DegreeFieldOfStudy,
--      field15 AS LevelOfDegree,
        case
        when len(field15) = 0 then null
        else field15
        end 'LevelOfDegree',
--      field16 AS SchoolState,
        case field16
          when 36 then 39
          else field16
        end 'SchoolState',
          field17 AS StateSpecific,
        field18 AS EndOfFileMarker,
        TermSort
FROM      dbo.rawdataapril30_sorted a
join ipeds_ic b on a.field3 = b.unitid and control like 'four%' and level like 'public%' --4year public
WHERE    (field4 = 1998) -- year
AND (field5 in (1,3))  -- summer/fall
and field11 = 1 -- first time
and field12 = 1 -- full time
and field1 = 'E' --enrollment
and field16 = '21' --KY
```

Members of the cohort met the following criteria:
- Record Type = Enrollment record
- Student Level = First-time
- Credit Hours = Full-time
- Term = summer or fall
- State of School = KY
- Year = 1998
- IPEDS Institution control = 4 year public

Note: The ipeds_ic table was created from DCT data on the control of the institutions for which data was submitted to NCHEMS.

SQL Queries to Populate the Report

```sql
set nocount on
select 'Graduation & Retention per term'
select 'Cohort: Kentucky 1998 4 year public institutions'
select 'Spring 1999'


/*  Longitudinal file for cohort to enrollment & completions in one term */
create table #base (
    [studyid] [char] (9),
    [campus] [char] (6) NULL,
    [schoolstate] [int] NULL ,
    [studyidE] [char] (9) NULL ,
    [campusE] [char] (6) NULL ,
    [schoolstateE] [int] NULL ,
    [controlE] [nvarchar] (255) NULL ,
    [multiE] char (3) NULL,
    [studyidC] [char] (9) NULL ,
    [campusC] [char] (6) NULL ,
    [schoolstateC] [int] NULL ,
    [controlC] [nvarchar] (255) NULL ,
    [levelofdegreeC] [char] (1) NULL
    )
    insert into #base (studyid, campus, schoolstate, studyidE, campusE, schoolstateE,
controlE, studyidC, campusC, schoolstateC, controlC, levelofdegreeC)
    select studyid, campus, schoolstate, studyidE, campusE, schoolstateE, controlE,
studyidC, campusC, schoolstateC, controlC, levelofdegreeC from
    (
    select studyid, campus, schoolstate, studyidE, campusE, schoolstateE, controlE,
studyidC, campusC, schoolstateC, controlC, levelofdegreeC
    from v_cohort_KY4yrPublic_NCHEMS_1998_zSummerFall cohort
    left join
            (
            select distinct studyid 'studyidE', campus 'campusE', schoolstate
'schoolstateE', control 'controlE'
            from v_NCHEMS_1999_aSpring
            left join IPEDS_IC i on campus = unitid
            where recordtype = 'E'
            )
    enrollment on cohort.studyid = enrollment.studyidE
    left join
            (
            select distinct studyid 'studyidC', campus 'campusC', schoolstate
'schoolstateC', control 'controlC', levelofdegree 'levelofdegreeC'
            from v_NCHEMS_1999_aSpring
            left join IPEDS_IC i on campus = unitid
            where recordtype = 'C' and (levelofdegree = '3' or levelofdegree = '5')
            )
    completions on enrollment.studyidE = completions.studyidC and enrollment.campusE =
completions.campusC
    ) base

update #base
set multiE = 'yes' where studyid in
(
            /* any student in the cohort who has more than one enrollment record in
the period*/
            select distinct studyid from v_NCHEMS_1999_aSpring where recordtype = 'e'
            and studyid in (select studyid from
v_cohort_KY4yrPublic_NCHEMS_1998_zSummerFall)
            group by studyid
            having count(*) > 1
)


update #base
```

```sql
set multiE = 'no' where studyid in
(
            /* any student in the cohort who has more than one enrollment record in
the period*/
            select distinct studyid from v_NCHEMS_1999_aSpring where recordtype = 'e'
            and studyid in (select studyid from
v_cohort_KY4yrPublic_NCHEMS_1998_zSummerFall)
            group by studyid
            having count(*) = 1
)


set nocount on
/*
Completed prior or not enrolled
*/
select count (distinct studyid) 'CompletedPrior_Spring1999' from #base
where studyidE is null


/*
Several students are enrolled in two institutions during a term.
If they graduate from one and not from the other, they will be counted in both the
retention & the graduate counts.
Including them in both populations leaves a calculated cohort count different from the
actual cohort count.
This number evens things out.
*/
select count(distinct studyid) 'correction_studentsretained&graduated' from
(
select studyid from (
select studyid, 'retained' 'sortfield' from #base where studyidE is not null and
(studyidC is null or levelofdegreeC = 3)
union
select studyid, 'graduated' from #base where studyidC is not null and levelofdegreeC =
5
) test
group by studyid
having count(*) >1
--order by studyid
)test


/*
Retained
*/
select multiE, count (distinct studyid) 'Retained_MultiE_Spring1999'  from #base
/* enrolled and either not completed or completed with Associate's degree */
where studyidE is not null and (studyidC is null or levelofdegreeC = 3)
group by multiE
order by multiE


select count (distinct studyid) 'Retained_Total_Spring1999'  from #base
/* enrolled and either not completed or completed with Associate's degree */
where studyidE is not null and (studyidC is null or levelofdegreeC = 3)


/*
Retained in same institution
*/
select multiE, count (distinct studyid) 'Retained_SameInstitution_Spring1999' from
#base
/* enrolled and studyid & campus matches and either not completed or completed with
Associate's degree */
where studyidE is not null and (studyid = studyidE and campus = campusE) and (studyidC
is null or levelofdegreeC = 3)
group by multiE
order by multiE
```

```sql
select count (distinct studyid) 'Retained_Total_SameInstitution_Spring1999' from #base
/* enrolled and studyid & campus matches and either not completed or completed with
Associate's degree */
where studyidE is not null and (studyid = studyidE and campus = campusE) and (studyidC
is null or levelofdegreeC = 3)


/*
Retained in same state (different institution)
*/
select multiE, controlE, count (distinct studyid) 'Retained_SameState_Spring1999' from
#base
/* enrolled and studyid & campus matches and either not completed or completed with
Associate's degree */
where studyidE is not null and (studyid = studyidE and campus <> campusE and
schoolstate = schoolstateE) and (studyidC is null or levelofdegreeC = 3)
group by multiE, controlE
order by multiE, controlE

/*
Retained in different state
*/
select multiE, case schoolstateE when 39 then 'Ohio' when 47 then 'Tennessee' when 21
then 'Kentucky' when 54 then 'West Virginia' else 'unknown' end as schoolstateE ,
controlE, count (distinct studyid) 'Retained_OtherState_Spring1999' from #base
/* enrolled and studyid & campus matches and either not completed or completed with
Associate's degree */
where studyidE is not null and (studyid = studyidE and schoolstate <> schoolstateE)
and (studyidC is null or levelofdegreeC = 3)
group by multiE, schoolstateE, controlE
order by multiE, schoolstateE, controlE

/* Total students with each type of degree */
select case levelofdegreeC when 3 then 'Assoicate' when 5 then 'Bachelor''s' end
'levelofdegreeC', count(distinct studyid) 'TotalStudentsWithDegree' from #base where
studyidC is not null
group by levelofdegreeC

/* Completed at same institution */
select case levelofdegreeC when 3 then 'Assoicate' when 5 then 'Bachelor''s' end
'levelofdegreeC', count (distinct studyid) 'Completed_SameInstitution_Spring1999' from
#base
where studyidC is not null and campus = campusC
group by levelofdegreeC
order by levelofdegreeC

/* Completed in same state but different institution */
select controlC, case levelofdegreeC when 3 then 'Assoicate' when 5 then 'Bachelor''s'
end 'levelofdegreeC', count (distinct studyid) 'Completed_SameState_Spring1999' from
#base
where studyidC is not null and schoolstate = schoolstateC and campus <> campusC
group by controlC, levelofdegreeC
order by controlC, levelofdegreeC

/* Completed in a different state */
select case schoolstateC when 39 then 'Ohio' when 47 then 'Tennessee' when 21 then
'Kentucky' when 54 then 'West Virginia' else 'unknown' end as schoolstateE , controlC,
case levelofdegreeC when 3 then 'Assoicate' when 5 then 'Bachelor''s' end
'levelofdegreeC', count (distinct studyid) 'Completed_OtherState_Spring1999' from
#base
where studyidC is not null and schoolstate <> schoolstateC
group by schoolstateC, controlC, levelofdegreeC
order by schoolstateC, controlC, levelofdegreeC

drop table #base
```

Appendix: Duplicate enrollment records

Three of the four states have duplicate enrollment records submitted for the same state, year, term, student, and campus.

KY – 1 duplicate: records appear identical
TN – 16 duplicates: most records appear identical; one or two differ by current field of study
OH – 463,848 duplicates: most records differ by credit hours, student level, or field of study

Interestingly, all but one school had duplicates only in the SPRING term and the different fields were credit hours (PT/FT), student level, and current field of study.

Appendix: Multiple Cohort Records

During a call with NCHEMS, questions arose about the cohort groups and whether they were affected by duplicate enrollment records. The cohort groups used for the 4-year public institution analysis were reviewed to determine the affect that duplicate enrollment records may have on the report. The cohort was determined by selecting the first-time, full-time, enrollment records, for either the summer or fall term of the year.

Three states (Kentucky, Ohio, and Tennessee) with students having two cohort records can be explained by dual enrollment (more than one institution reporting the student as a first-time full-time student. One state (West Virginia) had two cohort records per student based on entries in two different terms by the same institution in addition to having two cohort records attributable to dual enrollment.

KENTUCKY 4 year public schools: 1998 cohort

13,787 records
13,784 unique students
----
3 students with two cohort records

All three students had records from two different schools in the same term

KENTUCKY 4 year public schools: 1999 cohort

13,792 records
13,788 unique students
----
4 students with two cohort records

All four students had records from two different schools in the same term.

OHIO 4 year public schools: 1998 cohort

37,862 records
37,852 unique students
----
10 students with two cohort records

All 10 students had records from two different schools with two different fields of study.
For four students, the records also differed by term (summer/fall).
For one of those four students, the records also differed by county of origin.
For one of the ten students, the records differed by state & county of origin.

OHIO 4 year public schools: 1999 cohort

37,919 records
37,910 unique students
----
9 students with two cohort records

All nine students had records from two different schools, most of which also differed by current field of study.
For one of those students, the records also differed by term (summer/fall).

TENNESSEE 4 year public schools: 1998 cohort

14,668 records
14,665 unique students
----
3 students with two cohort records

All three students had records from two different schools in the same term.
Two of those students also had different fields of study.

TENNESSEE 4 year public schools: 1999 cohort

14,931 records
14,929 unique students
----
2 students with two cohort records

Both students had records from two different schools in the same term.

WEST VIRGINIA 4 year public schools: 1998 cohort

10,778 records
10,753 unique students
----
25 students with two cohort records

For 3 students, the records were for different schools and different fields of study in the same term.
For 22 students, the records were for a different term (summer/fall) in the same school. A few differed by field of study.

WEST VIRGINIA 4 year public schools: 1999 cohort

10,687 records
10,680 unique students
----
7 students with two cohort records

For 5 students, the records were for different schools in the same term.
For 2 students, the records were for a different term (summer/fall) in the same school.