

Designing and Implementing a System for Assigning Student Identifiers in New York

Barbara S. Clements, Ph.D.
Glynn D. Ligon, Ph.D.
Evaluation Software Publishing, Incorporated

This paper discusses and summarizes issues related to the creation and maintenance of a statewide student identifier system for the students in the State of New York. In September, 1999, Dr. Clements met with representatives of the New York State Education Department (NYSED) to explore options and assist in setting a context for designing a plan. In March, 2001, Dr. Clements and Dr. Ligon met with representatives of the NYSED, Regional Information Centers (RICs), school districts, and special programs to continue that discussion and to begin finalizing options for a plan.

April 2001

This document combines descriptions and background information from a variety of sources. Evaluation Software Publishing has consulted with several states directly and others indirectly on the design of their student identification systems. From those activities, crucial issues and key decisions have been identified. Those are brought together here with the state-specific issues identified for New York.

The Case for Student Identifiers

The Chief Information Officer for the New York State Education Department (NYSED) began the March 29, 2001 planning session with the statement that a statewide student identifier is already generally acknowledged as a desirable necessity. Therefore the discussion focused upon documenting the context within New York for an identifier.

This included:

1. What requirements exist?
2. What restrictions exist?
3. What preferences exist?
4. What additional legislation or regulations may be needed?

Attachment A provides the content recorded on charts during the March 29, 2001 discussion. This includes descriptions of current practices within Regional Information Centers (RICs) and districts.

Existing requirements and restrictions centered around New York's legislation related to restrictions on the use of Social Security Numbers, and the prior allocation of number ranges to districts and RICs in 1991.

- Section three hundred five of the education law, subdivision twenty-one (July 1988), requires that the New York Commissioner of Education establish the

procedures for a statewide system of assigning unique identification numbers for all students in public and nonpublic schools. The identification numbers are to be used for tracking students and for state reporting purposes. The numbers are to be assigned when a student enters school in New York and retained until the student reaches the age of 21. The commissioner is supposed to provide for appropriate procedures of insuring security and confidentiality of student information.

- In a Memo from Liz Taylor to Mike Radlick dated August 27, 1990, ranges of nine digit numbers were assigned to each Regional Center in New York State. Regional Information Centers were directed to be responsible for maintaining a file listing of all student numbers assigned throughout the region, and for assisting schools when assigning identifiers to new students. RICs were expected to provide ranges of numbers and work with school districts not on-line with the RIC (including the largest school districts).
- A document on NYS Unique Student IDs from the WNYRIC dated October 28, 1999, provides a historical perspective on what had been accomplished and offers recommendations for moving ahead on the assignment of student identifiers and establishment of the NY statewide database.
- New York Senate Bill SO6237B, dated June 13, 2000, restricts the use of Social Security Numbers for students. Basically the bill restricts any public display of the student's SSN on any student record for any public identification purpose.

These documents provided background information for the discussion, and validated many of the suggestions made.

Overview

The New York State Education Department finds the need for statewide unique student identifiers growing with the increasing demands for an accessible, centralized information source for decision making. In response to this, NYSED is planning to develop more automated information systems. These software applications envision information codified to save space and to allow for analysis of information across separate files. Although some extant information systems at the school and district levels may use only the student's name or a local identifier, those that anticipate linking related information in other files, longitudinal comparisons, or efficient searching for records across districts must have a common identifier system.

Today, the burden of managing identifiers has been heightened by the necessity of accommodating legal and personal demands for confidentiality. The identifier itself has become the object of confidentiality concerns, because the identifier becomes a virtual key to unlock access to personal information. In some cases, the identifier itself contains imbedded information about the individual. A discussion of student identifiers cannot be conducted in isolation from associated confidentiality and security issues.

Within this challenging context, local education agencies (LEAs) and state education agencies (SEAs) have developed or adopted an incredible array of student identifier systems. Most were developed with influence by political and legal forces. Many were

merely inherited from the software applications purchased to manage student information. Attachment A describes some of those used across the State of New York.

There is a very practical side to student identifiers. These identifiers become the index used by a computer program to find an individual's record(s). In a relational database, the identifier takes on a heightened sense of importance because information about an individual is stored in separate tables rather than in a single, long string of fields.

Some may expect that with an adequate student identifier there would no longer be a need to store a student's name in a database. The reality is that even the best identifier needs to be verified. Verification means having other pieces of information available to determine that the identifier is attached to the correct student's records. This verification does not have to occur within the main database (e.g., the files kept at the NYSED), but somewhere there must be a table that matches the identifier with other data elements such as name, birthdate, and gender to allow periodic verification.

This paper describes and discusses specific issues related to the creation and maintenance of a statewide student identifier system. Options for designing that system and recommendations for New York to consider are presented.

What is the Nature of a Functional Student Identifier for New York?

What is the essence of a student identifier according to the discussion on March 29, 2001? A functional student identifier for New York must:

1. Be a stand-in for a student's name.
 - a. There is a preference not to have the student's actual name in the state database at SDE.
 - b. Students' names change, the name components change order, they are abbreviated, and they are not reliably provided by a student or parent from one time to the next.
 - c. Names are of varying length, so a file has to allocate long fields that take up substantial space in a database.
 - d. At times we do not want a student's name showing on a document.

2. Be unique, unduplicated, and permanent.
 - a. New York students move, change names, go away and return, but databases must be able to link a student's records across time, across files, and across schools.

- b. The identifier must be unique to identify one and only one student or else records could be inappropriately combined for more than one student.
 - c. The identifier must not be an alias because the student must have one and only one number, so all of the student's records can be linked into one (virtual) set.
 - d. The identifier must be permanent because changes in a student's identifier impose extreme challenges to link multiple identifiers to the same student.
 - e. Permanent also means that the NYSED or some state-level process must support the selection of the identifier for a student. States that allow a parent to choose between using a Social Security Number or a state-assigned number must constantly manage changes in the parent's selection.
3. Be ubiquitous.
- a. The local and state records and systems in New York that use the same identifier must be linked to create a virtual set of records for an individual student.
 - b. Crosswalk tables can be used to translate the identifier used by one system to the identifier used by another. However, creating and maintaining crosswalk tables adds a step to usage of the data.
 - c. Records that do not use an identifier must be linked using less precise matching techniques, such as name and demographic characteristics.

The bottom line for a student identifier is that it must identify one and only one student “forever.” Forever is defined as being as long as records for a student are maintained within the database—even if the student leaves the state for a time and returns later.

In New York, forever is defined in statute as the age of 21. However, longitudinal studies and follow-ups of students into the workplace and postsecondary education depend upon these identifiers remaining unique as long as possible.

Options for Selecting a Student Identifier

The March 29, 2001 discussion explored the advantages and disadvantages of several options used in other states. Other SEAs have adopted a wide variety of identifiers. There is no generally accepted "best practice" to be copied. The approaches used most often are categorized below along with an analysis of the advantages and disadvantages discussed for each.

1. **Local-Assigned Identifier:** This option allows each district (or school) to assign identifiers in whatever manner it is accustomed or would adopt. Some New York LEAs have extant identifiers from the 1991 assigned range of numbers or from a local process, and some are assigned at the school level by student information management software applications. The NYSED could use these pre-existing identifiers.

Option 1.a. Authorize LEAs to continue to assign identifiers as is current practice.

Option 1.b. Collect pre-existing local identifiers in year one only, and then assign all new numbers from the NYSED system.

Option 1.c. Add a district identifier to the front of locally assigned identifiers to make them unique across the state.

Advantages: Local control is maintained. LEAs are not required to make changes in their identifier system. Historical local files continue to be compatible with the identifiers. Dependent upon local processes in place, an identifier can be assigned at the time of registration for new students.

Disadvantages: Uniqueness across the state is not likely. Mobile students would receive a new identifier in each district/school. The formats and characteristics of local identifiers would have to be considered in the establishing of parameters for acceptable identifiers. NYSED's central database design and the checking for aliases and duplications would be more complicated. 1.c might work with a district identifier added to the front of local identifiers to make them unique statewide. However, where the local identifiers are only unique within a school, both a district and a school identifier would have to be added. This has implications for the length of the identifier.

This strategy breaks down when students move from one district to another. Each district would have to accommodate the characteristics of other districts' identifiers or there would be aliases created in the assignment of numbers to individual students. The NYSED would have to abandon uniqueness statewide to fully accommodate option 1.a.

Option 1.b would require the setting of criteria for a local identifier to be acceptable, e.g., no longer than the NYSED identifier, same characteristics in regard to alpha, numeric, and special characters, etc. This might eliminate too many local identifiers from use.

Option 1.c. works only if receiving schools verify a prior identifier from a prior district rather than assigning a new local identifier. The addition of a

three-character district identifier to the front may make the identifier longer than practical for some local student information systems.

2. **Algorithm to Assign Identifier:** An algorithm could be devised that encrypts the student's name, birthdate, gender, place of birth, and possibly other data elements to create a unique combination. The algorithm could be secured to protect the contents of the identifier.

Advantages: The algorithm could be distributed as a software application to registrars for use at the time of registration. Parents and students would not have to remember the identifier when they move, nor would registrars have to contact prior schools.

Disadvantages: The required length of the identifier to ensure uniqueness might be excessive. Students who have identical names and demographic characteristics would get the same identifier. Twins at times have the same names and identifying characteristics. Students who happen to get the same identifier upon their first registration will always get the same one generated by the algorithm, and will need a solution each time.

A hacker (an unscrupulous computer user) might obtain or break the encryption routine.

Changes in names or mistakes in the entry of data elements used to run the algorithm would result in incorrect identifiers. Requiring the use of names, birthdates, etc. directly from an official birth certificate would be advisable.

Alternative: The algorithm could be used to "estimate" a mobile student's identifier. Some states use a sound encryption to generate possible matches of students in their databases. Then a manual process is used to make the final match. New York uses this technique presently with STAC identifiers.

3. **State-Assigned Identifiers:** The NYSED could create a pool of valid and available identifiers, and students new to the state would be assigned a permanent identifier from the pool.

Option 3.a. Assign identifiers from a common statewide pool.

Option 3.b. Assign LEAs or RICs a block of identifiers from the state pool.

Advantages: The NYSED fully controls the characteristics of the identifiers. The validity of identifiers can be verified by the NYSED. Uniqueness is assured within the pool of potential identifiers. Option 3.b. facilitates assignment of identifiers to new students at the time of registration. The identifier cannot be deciphered and directly linked to confidential data sources.

In 1991, a range of nine-digit numbers was assigned to each RIC or large city school district. Some have been using these numbers, including New York City. Maintaining the assignment of numbers corresponding to those already in use would limit the burden and transition effort for instituting a statewide NYSED identifier.

Disadvantages: LEAs must rely upon the NYSED for their identifiers. Access to their identifiers and the ability to assign them at registration will require sound management.

The 1991 assignment of numbers has not been followed precisely by those using them. However, a computerized system would have no problem reserving currently assigned numbers before reallocating ranges to districts.

4. **Social Security Number (SSN):** Use of the SSN as an identifier is legal and in practice in a number of states. With the requirement by the IRS for dependents to have a SSN, almost all students enter school with a SSN assigned. The use of SSN for student identifiers has been a common practice by postsecondary institutions for decades.

Option 4.a. Require SSN.

Option 4b. Encourage use of SSN, but provide for an alternate identifier at the parent's request.

Option 4.c. Collect the SSN as a data element, but use another identifier.

Advantages: SSN is unique on a nationwide basis. SSN is almost universal in its assignment in the U.S. The assignment and maintenance of the SSN system is handled by the Federal government; thus, the burden of determining an identifier and assigning it is avoided by NYSED. With SSN, the parent and student can assist in providing the identifier when transferring from one school to another. When students move across states, those states using SSN can use it to verify records. Postsecondary institutions are assisted in their applications processes when secondary schools can provide SSN on transcripts.

SSN can be used to share information or conduct studies across agencies that use it. Some states exchange information about families across agencies to determine eligibility for services. For example, several states use SSN and other family information to link across Aid to Families with Dependent Children and other public assistance files to establish a student's eligibility for the school meals program, to count the number of economically disadvantaged students to qualify a campus for Title I funding, and to establish a student's eligibility for vocational and job training programs.

More controversial is the ability to exchange student information with law enforcement agencies such as the local police or the Immigration and Naturalization Service. Confirmation of a student's identity and enrollment status can assist their investigations. At times, LEAs and SEAs are legally required to cooperate and provide information. Within that context, ensuring that the information provided actually belongs to the correct individual is critical.

SSN is useful when conducting former student studies. Employers, the armed forces, and postsecondary institutions can use it to correctly match former students with their current employees or enrollees.

Disadvantages: Some parents are uncomfortable providing SSN, others strongly oppose its use. A very small number of students will not have one. Parents have occasionally provided their own SSN for their children.

To use SSN, the NYSED would be obligated to provide an alternative identifier to parents or students who refuse to provide it. The requirement to design and implement an alternative identifier is equivalent to having two identifier systems in place. Students may move back and forth between the use of their SSN and their request for an alternative identifier. Tracking these changes in a longitudinal database is difficult.

In states using SSN, an attorney general's opinion, legislative authority, or state board of education authority is typically secured first. Consideration of SSN adds time and effort to the planning, review, and public comment process.

In contrast to a nominal identifier, the printing or display of SSN on education documents demands a higher degree of diligence from everyone handling those documents. In fact, New York law prevents the display of the SSN on student records.

Throughout this document a "Direction Indicated" will be described for each issue discussed. The group did not vote on options, and ESP does not represent these as

recommendations. These options are stated as the directions that appear to follow the general sense of the discussion.

Direction Indicated: Adopt a state-assigned identifier from a pool of available numbers. Allow schools and districts to continue the use of locally assigned identifiers within their information systems. Require the reporting of the NYSED identifier on all state reports. Also collect the locally assigned identifier as an additional verification field.

Analysis of Issues Related to Options

These issues are not mutually independent. In the analysis that follows, the options available will interact with the options described for another.

1. Uniqueness--At what level must the identifiers be unique (e.g., school, district, RIC, state, nation)?

An identifier must not be duplicated within a population or be an alias for a single individual within the population. The population defined here encompasses all elementary and secondary students in New York. Therefore, uniqueness must be maintained at the state level for a statewide identifier to ensure accuracy and completeness of data. The current identifiers assigned by schools and districts to their students are not unique across other districts. In fact, some commercial student information management systems adopted by RICs, districts, or schools may provide uniqueness only within a school building for a single year.

Statewide uniqueness can be achieved if districts assign identifiers that are unique within the district and begin with a district number. However, this complicates the process of verifying and using those numbers when students move across districts. The variance in length across districts complicates this option.

There are certain benefits to using an identifier that is unique nationwide. At this time, the only such identifier is the SSN. SSN provides functionality for tracking former students into postsecondary education, verifying the identity of students across states (with agencies using SSN), and exchanging useful data with other state and Federal agencies that provide services to families and individuals (e.g., verifying eligibility for services). These benefits can also be achieved by collecting SSN as an additional data element irrespective of its use as the student identifier.

Direction Indicated: Ensure student identifiers are unique at the state level.

2. Burden--What level of burden should be imposed upon LEAs?

Burden is defined as the time, effort, and resources required to implement the student identifier system. This includes the necessity of creating the system, assigning the identifiers, verifying an individual's identifier, and entering the identifiers wherever they are required. Burden also includes the effort to make the transition from an existing identifier system to a new one; or to add the new identifier into an existing system or perform a crosswalk from the local identifier to the state identifier each time a report is exchanged.

Clearly the level of burden must be limited to achieve compliance (both voluntary and practical) with the identifier process. Too high a level of burden will introduce unwanted errors as a consequence of the attention to detail required. Burden must be balanced by benefit. In the case of identifiers for New York students, the benefits have already been determined to be high because they are critical to the functionality of the entire NYSED student information system. The options that impose the least burden, school and district identifiers, fail to provide the functionality required as described by other issues. Burden is typically an issue to recognize and to manage rather than being one of the deciding issues when a student identification system is selected.

A moderate level of burden can be achieved by allowing the continuing use of local identifiers within local information systems at the discretion of schools and districts. Crosswalking to the state identifiers at the time of state reporting is commonplace across states. An alternative is for the state identifier to be recorded in the local information systems as a separate field to be included with data extracts at the time reports to the NYSED are created. In cases where the local student information system software does not allow for second identifiers, the crosswalk option would be necessitated.

Direction Indicated: Accept a moderate level of burden in exchange for the clear benefits from a functional student identifier within the information system.

3. Assignment--At what level will the identifiers be assigned to individual students (e.g., school, district, state, national)?

The NYSED must determine the pool of identifiers available to be assigned. However, assignment of the identifiers at the level of registration (either at the school or at the district) provides the quickest and least burdensome alternative. This issue is also related to uniqueness. The schools (or district office where central registration occurs) must follow a procedure that ensures unduplicated identifiers. Therefore, identifiers should be assigned at the lowest level possible without losing their uniqueness.

A reality is that parents and students cannot be relied upon to carry their student identifiers from one school to the next. Mobile families too often cannot identify

their last school/district, do not have records with them, and cannot remember student identifiers. Thus, a system must be in place to avoid assigning a new identifier when an alias exists.

If a local identifier is to be assigned at registration and used for local purposes, then the assignment of the state identifier can be delayed until some time before the next reporting to the state.

If the state identifier is to be used in the local student information system as the primary identifier, then a process for the identifier to be assigned promptly must be in place. An on-line student locator system that provides the identifier to be assigned or a list of available identifiers would be required at registration.

Direction Indicated: Determine the pool of identifiers available at the NYSED and assign identifiers to individual students at the level where registration occurs (e.g., school or district).

NOTE: SEAs typically require that their official student identifier appear on all reports and data submissions from the LEA. There is usually not a requirement that the state's identifier be used on all local files and records. A district or school could opt to use its own identifier system for local applications such as scheduling and grade reporting. A crosswalk table could be used to translate local identifiers to the state identifiers whenever reporting to the state is required.

4. Timing--When will the identifier be assigned to a student (e.g., at registration, as soon as possible after registration, at the time of the first report to NYSED)?

If the state identifier is to be used for local records, schools need an identifier immediately upon enrollment of a new student. Certain forms are completed at that time and begin to go their separate ways. Ensuring that the student's identifier is on each form immediately saves changes and mismatches later. The difference between assigning identifiers immediately and within a few hours is arguable. However, any system that takes days to assign an identifier presents a very different level of burden--and potential for errors that must be cleaned up later. An on-line student locator system that provides the identifier to be assigned or a list of available identifiers would be required at registration.

Direction Indicated:

For districts using the state identifier as their local identifier, it should be assigned at registration and be available to schools before enrollment forms begin to be distributed to their respective offices. If registration occurs at the school, then the school should be enabled to assign the identifier.

For districts NOT using the state identifier as their local identifier, the timing of the assignment can be any time before the next state reporting.

5. Verification Level--Where will the identifier be verified (school, district, state)?

When a student moves from one school or district to another, the student's identifier must be verified upon registration in the new district. Verification is the process made available to ensure that the identifier assigned to a student is valid and correct. Valid means that the identifier is one actually included in the pool of identifiers to be assigned. Correct means that the identifier is accurately matched to the student.

The identifier can be verified immediately upon registration or later as part of a validation process at the state level. The earlier the verification occurs, the fewer changes will be required later if an identifier is changed/corrected. The closer the verification occurs to the parent and student, the higher the probability of accuracy. Verification at the time of registration, when the parent and student are most likely to be present, is best. This requires that the person registering the student be authorized to assign the identifier according to a set of precise rules, or that the person have direct access to the assignment process.

Verification conducted at the state level using available demographics in the database after submission of individual records is the least efficient. Some correct identifiers can be incorrectly challenged based upon duplications in the data elements used for verification, e.g., students with the same name, birthdate, and gender. In these instances, verification is then delegated back to the school. This state-level verification is a required component of the system, but the frequency of potentially incorrect identifiers can be greatly reduced by adequate controls at registration.

An on-line student locator system would provide the verification necessary.

Direction Indicated: Verify the identifier at the time of registration, if possible, when parents and students are available to answer questions and provide documents.

6. Assignment/Verification Process—How will the identifier be assigned or verified?

The assignment/verification process includes several steps:

First, the registrar determines if the student already has an identifier assigned.

Second, the registrar secures the existing identifier or causes a new one to be assigned.

Third, the registrar records the identifier in local records for use in state reporting.

The registrar can accomplish the first step by asking the parent or student, or examining paper or electronic records from a prior school. Parents and students too often do not have the records, and at times cannot precisely identify contact numbers or addresses for the prior school (e.g., districts with county or descriptive names rather than city names). In the case of migrant worker families, enrollment in a prior school may have been too brief to generate an official record.

An alternative is to establish a statewide reference file (locator system), such as a web page, that can be queried to determine the existence of a previously assigned identifier. The registrar could access the reference during the registration process. The second step could be accomplished using the statewide reference to learn the identifier or to request assignment of a new one. In the absence of such a reference, the registrar must contact the prior school. This is a critical point. Students who have existing identifiers can be assigned an alias identifier simply because that is easier than contacting a prior school, or because the prior school cannot be contacted or does not respond promptly. This might result in duplicate identifiers for a single student, rather than a single unique identifier. If such a number is meant to serve as a temporary placeholder for the permanent identifier, procedures would be needed to replace the temporary identifier with the permanent identifier as soon as possible.

Direction Indicated: Registrars should assign and verify identifiers at the time of registration using a statewide reference for finding existing identifiers and a state-approved process for assigning new numbers.

7. Verification Data Elements—What data elements are required for verification of a student’s identifier?

When a student’s identifier is in doubt, other unique combinations of information about the student must be used for verification. This is typically “directory information” as defined by the Family Educational Rights and Privacy Act (FERPA) along with a few other pieces of information included to increase the probability of describing a single student. Directory information typically includes:

- Name
- Birthdate
- Gender
- Prior schools/districts of enrollment, enrollment dates, and grade level

Additional detail information could include:

- Race/Ethnicity
- Place of Birth
- Parents' Names
- Date of First Immunization

Security and confidentiality issues must be considered. However, the more information available for query, the more likely existing identifiers will be found and used.

One state provides for a two-phase identification system. If the use of directory type data elements results in multiple matches, then other data elements, such as parents' names, race/ethnicity, and place of birth, are made available to the registrar for making a correct identification.

Direction Indicated: Include as many of these data elements as practical in the verification resource.

8. Confidentiality--Who may know the identifier?

FERPA and local policies will define the answer. If the identifier is a nominal code without intrinsic meaning, then it may be viewed as directory information. This assumes that links using this identifier are not generally available to the public.

Prudent practice would call for the identifier to be treated as confidential, because knowledge of the number would place the holder one step closer to access to confidential information.

A general consensus was expressed in the March meeting that the NYSED should not collect a student's name or identifier. This would protect the NYSED from the risk of violating FERPA and other mandates for confidentiality. The NYSED must have a method for uniquely distinguishing each student's record and for matching student information across files and years. This can be accomplished along with the confidentiality preference by encrypting the actual NYSED identifier when it is passed from the local level to the NYSED.

Direction Indicated: Allow only education employees with a need to know to access the student identifier at the local level. Encrypt the identifier when it is passed to the NYSED and stored in statewide files.

9. Imbedded Information--What meaning will be built into each number?

SSN has no useful intrinsic meaning imbedded in the numbers. (Generally, the number may imply a region, year of assignment, or sequence, but any algorithm producing the number is obscure.)

Imbedded information typically adds to the length of a number. For example, county/district codes may add six characters to a number. Birthdate may add eight.

A truly unduplicated, random number carrying no meaning has the advantage of requiring less restrictive security and confidentiality precautions.

If the district number is imbedded, this could identify the student's first district of enrollment; however, that information can be carried in other fields within the database as well. In fact, any useful information that might be imbedded in the identifier can also accompany the identifier as a separate field.

Direction Indicated: Do not imbed information in the student identifier.

10. Length--How many characters can be in each number?

Shorter numbers can be entered, transcribed, and maintained with fewer errors. A common length provided for an identification number on generic scanner documents is 10. SSN is currently 9, but moving to 10 numbers has been discussed. To accommodate 3-4 million active students and to retire numbers for former students until their age reaches 30, requires 8 numbers. This provides 100 million minus one unique numbers.

Direction Indicated: Limit the student identifier to 9 or 10 digits.

11. Characters--What should be the nature of the characters in the identifier?

Any number, letter, or symbol could be used. Symbols and letters present problems with recognition and accuracy in entering—especially when mixed with numbers. Certain letters (e.g., o, l, i, z, E, b/d, q/p) are sometimes confused with numbers or each other.

Using both numbers and letters provides for many more combinations for unique identifiers, and thus the ability to have shorter identifiers. Problematic numbers and letters could even be eliminated from use (e.g., neither 0 nor o ever assigned).

Some state systems use letters and numbers in combination (e.g., a state assigned number beginning with a letter to distinguish it from a SSN.)

Letters require 26 bubbles for each character on a scanner form compared to 10 for numbers. Combinations of letters and numbers require 36 bubbles. Letters come in capital and lower case forms that may or may not have meaning, but often cause confusion as to their use. Numbers are easier to distinguish from each other, they can carry intrinsic and extrinsic meaning, and they are more “universal” across languages and cultures. Numbers can be assigned without risking the creation of meaningful and undesirable combinations as with letters.

Even in the absence of considerations that require restricting the length of the identifier, use of only numerals is preferable. They are easy to distinguish. They can be entered with efficient keystrokes using a number pad. They require less space and are associated with less bubbling error on scanner forms.

Direction Indicated: Use only numerals for the student identifier.

12. Rubric--What conditions will be imposed on the numbers?

If an algorithm or imbedded meaning is to be used, then those considerations will answer this question. However, if a random number is used, then several rules can be followed to reduce data entry and clerical errors.

Leading and final zeros are sometimes accidentally, or by software design, dropped when numbers are entered or moved across databases. When the remaining numbers are justified left or right, then a reader or a computer application can misinterpret the identifier.

Consecutive identical numerals at times are incorrectly entered too few or too many times.

Eliminating all the cases described above would reduce the available pool of numbers by about 15%.

A final check digit (a number calculated by formula from the other digits) is sometimes used to provide a quick way to locate invalid numbers. With this methodology, if the verification formula checking the validity of a number does not generate the final digit as in the number reported, then there is an error. This check digit could be the 10th character. If a check digit is used, and it is not allowed to be zero, then some of the numbers eliminated because they end in zero would be recovered.

The check digit does not have to be a part of the identifier. The digit can be held in a separate field accompanying the identifier. With this option, the check digit is not always available.

Direction Indicated: Use unduplicated random numbers with the exception of any with an initial or final zero, or any sequence of three or more identical numerals. Calculate a final check digit that can be used as a final digit or as a separate field.

Policy Implications

Several of the issues discussed raise policy questions. In general, any use of SSN is considered as having policy implications. This designation is not a legal determination, but a practical recognition of the controversy that is associated with the use of SSN. As stated earlier, SEAs using SSN have typically acted under or sought their attorneys general's opinion, legislative mandate, or an official departmental regulation.

The level of confidentiality ascribed to the identifier is a direct outgrowth of applicable FERPA, New York, and local provisions. If the identifier contains meaningful elements, even if they are encrypted, a policy decision would be advised to determine if the identifier as a consequence was confidential. Even in the absence of any meaningful content, e.g., a truly random identifier, the identifier might be designated as confidential except for educators and other governmental employees with a need to know. This designation in part could be based upon the potential for accessing other information using the identifier if it were to be known.

The level of confidentiality would impact the openness of any verification reference. For example, a web page accessible by districts to look up identifiers or to assign an identifier for a new student would need to be compliant with adopted policy.

Associated with the definition of the identifier and its use within the overall student information system, is the issue of defining the basis for "a need to know." This phrase communicates a very general meaning that a person must have a legitimate educational reason to access confidential information. A clear statement of who has a need to know under what circumstances and which data elements they can know is a basic component of a functional student information system. The role of the student identifier should be defined within this statement as well.

In general, an identifier other than SSN would not need a high level of confidentiality as long as prudent firewalls and access procedures are in place to manage access to the databases within which confidential information exists.

Summary

These issues and conclusions support the use of a unique, numeric identifier determined by NYSED. A major factor influencing this is the assumption that New York schools and districts will have functional and universal Internet access to NYSED. This makes the use of a network-based (e.g., Internet) locator reference practical. Without the locator, verifying identifiers and ensuring the assignment of unique identifiers to new students is more challenging.

SSN should be considered for collection and inclusion in the locator and central database. The added value from having the SSN in a student's record is easily documented.

At the conclusion of the discussion on March 29, 2001, the CIO summarized the issues as follows.

Issue Summary for New York Student Identifier Process

Issues on Which General Consensus Exists:

1. The New York State Education Department will maintain the student identifier system and the official directory of IDs that have been assigned.
2. LEAs will be able to look up individual students to verify previously assigned IDs, to input data on individual new students for assignment of an ID, and/or use a batch input process.
3. New York will begin with an assigned set or range of ID numbers for each entity (e.g., RIC or district) authorized to assign IDs. Existing IDs will be maintained and will be retired from the available pool of numbers.
4. LEAs will create a field in their local student information systems for the state ID. This will allow for reporting of both state and local IDs and matching across state and local files.
5. LEAs will be able to access historical data from state files for analysis and tracking of students.
6. LEAs will be able to access data about their students for a local decision support system.
7. The benefits of collecting SSNs (e.g., follow-up to postsecondary, social services, and vocational services) will be detailed to encourage their collection in LEA information systems.
8. The burden on LEAs for reporting to the state will be reduced through coordination of collections and use of existing data.

9. The aim is to include all students in New York (e.g., public, private, special programs, etc.).

Unresolved issues:

1. Funding
2. What are the data elements to accompany the ID for positive identification of individuals?
3. What are the data elements the NYSED will collect in the individual student record system?
4. What is the set of ID numbers that have been assigned or are available?

What design matches the recommendations?

Attachment B is a diagram showing the process of assigning, verifying, and maintaining student identifiers with the options discussed. This diagram also illustrates the relationships among the student identifier, a web-based locator, and the central NYSED student database.

The NYSED Central Student Database: This central student database is shown as only very generally representative of the actual database to be created. The central database would not be required to contain the student's name and identifier. In Attachment C, a confidential encryption process is shown. This process would translate the NYSED identifier reported by the districts to an internal key. The internal key would become a field in the central database as a proxy for name and identifier.

An important point here is that the internal key would be crosswalked and encrypted. An unauthorized user would not be able to use the internal key to decrypt the student's official identifier.

The Encryption Process: The decrypted and modified file is then uploaded to the central student database. The confidential crosswalk table must be protected extremely well from unauthorized access. The internal key carries no intrinsic meaning and provides a user with no linking to a student's name, identifier, or other files without going through the confidential crosswalk table.

The Identifier Database and Locator System: A web page is envisioned to provide timely access to a reference file of student identifiers

Two options exist for NYSED's use of the locator.

- a. The locator can serve simply as a reference for student information and the assignment and verification of student identifiers. The locator database would be updated each time a submission is received by NYSED.
- b. The locator can provide the added value of tracking the enrollments and withdrawals of students statewide. NYSED would mandate that every time a student enrolls in a new school, even if the move was within the same district, that the registrar would connect to the locator and enter the entry date. This would provide an up-to-date reference for schools searching for information as contrasted with "a," which provides snapshots of student locations after each submission date.

Value-Added Uses for the Identifier Database and the Locator System

In the consideration of the design and use of a locator, there are several valuable additional uses to which the information in the identifier database could be applied.

- 1. Electronic Records Request:** The locator can offer a feature to send an electronic request for a student's record to the student's prior school. Transaction set 146, Request for a Student Record, ANSI X12 Standards, SPEEDE/ExPRESS, provides a format for such a request.
- 2. Migrant Programs:** Students eligible for Title I Migrant program services could be identified. This process would be compliant with the Federal requirement for timely records exchanges for migratory families.
- 3. Dropouts:** Tracking and verification of mobile students can reduce the reported dropout rates by providing a way to document transfers who otherwise would be considered dropouts.
- 4. Placement:** The locator has the potential to provide a new school with valuable placement information about the student. Placement in proper courses, support services, and programs can save a new school the time and resources required for assessments. The student can be provided more continuous services and avoid changes that might be required upon completion of a reassessment or arrival of records from a prior school. In addition, students with special needs, e.g., vision or hearing modifications, emergency procedures, or free meals, can be accommodated. The inclusion of data elements useful for placement decisions changes the nature of the locator and raises extended confidentiality and access issues.

The drawback of including these value-added features in the initial locator is that each one raises additional issues and brings into the design process additional players and considerations. A prudent approach is to design the locator to accommodate these

features and uses, but implement the primary objective of the locator—to verify and assign student identifiers.

Suggested Confidentiality for NYSED Decisions Related to the Student Identifier

Public entities are not obligated to publish or provide to the public details related to systems and designs intended to protect the security and confidentiality of information systems. This protection is intended to prevent unauthorized access beyond that allowed by law or policy. Therefore, the details associated with Attachment B, and certainly any decisions made related to design and implementation of the student information system and the student identifier should be designated as internal documents, not for public distribution.

Steps Remaining

After carefully considering the issues and options available, what are the next major steps?

1. Document the current use across the state of the 1991 number ranges. Reserve the ranges that are in current use, and create an available pool of numbers to be reallocated to districts. Create the reallocation.
2. Select from among the value-added functions possible those to be targeted either in the short run or long-term. Incorporate the requirements and target dates for each into the planning that follows.
3. Determine the features and functions desired for the statewide student locator system, and decide upon the process to be followed to design and build it.
4. Determine the directory information to be used in the locator system as well as any additional descriptive data elements desired.
5. Determine the features and functions for the encryption process to convert actual NYSED identifiers into an encrypted identifier for reports to NYSED, and decide upon the process to be followed to design and build it.
6. Determine political and legal issues that may require legislative or administrative action. Develop a plan for that action.
7. Set a date for districts to comply with the implementation of the statewide identifier. Provide necessary and helpful information to prepare them for assigning new numbers and securing existing ones.

8. Design a process for the initial assignment, collection into NYSED, and cleaning up of duplicates. Require schools and districts to verify and correct all identifiers. Monitor this process to assign and verify identifiers to a high level of accuracy as a prerequisite to collection of student records at the state level.

This paper was prepared for the New York State Education Department by Evaluation Software Publishing, Incorporated. The analyses and recommendations presented here are those of ESP. They may or may not reflect the thinking or any final decisions made by NYSDE.

GLOSSARY OF TERMS USED IN THIS PAPER

The definition of terms as they are used in this paper is important to facilitate a precise communication of the issues and recommendations described.

Access: Viewing or gaining possession of information or data elements from a database or other product of an information system

Algorithm: A set of steps for determining the identifier to be assigned to an individual, specifically when the individual's name or other personal characteristics are used in the determination of the identifier

Alias: A prior identifier that is not used when a new identifier is assigned to the same individual

Assignment: Linking an identifier to an individual

Burden: The amount of effort (or resources) required (to comply with the use of a statewide student identifier system)

Confidential: Describes information and data elements that must not be accessed by persons without the authority to know them

Correct: An identifier that is the actual and legitimate identifier for an individual

Directory Information: Data elements not designated as confidential by law, regulation, policy, or practice (e.g., name, birthdate, school, grade level, gender, place of birth, parents' names, address, phone number, etc.)

Duplicated: When an identifier is assigned to more than one individual

Encrypted: Encoded (student identifiers and other personally identifying data elements) information that cannot be linked to an individual

Enrollment: The listing of an individual on the membership roll of a school for the purposes of attendance

Family Educational Rights and Privacy Act (FERPA): The Federal law protecting the confidentiality of student educational records

Freedom of Information Law (FOIL): New York's law defining public accessibility to information

Identifier: A set of characters used as a proxy for name to identify uniquely an individual (Identification number is not used because that term denotes a numeric identifier. ID is not used because it at times refers to the physical medium, e.g., card, used to document an identifier.)

Locator System: Process and automated application that allows for looking up a student to determine whether or not a pre-existing identifier has been assigned

Masking: Deleting from publication any values that are based upon a small number of individuals such that the identity or value of a single individual might be determinable

Nominal: When an identifier contains no intrinsic or embedded meaning other than to designate an individual (An identifier created using an algorithm is not nominal because meaning is imbedded in its characters even though that meaning may be disguised.)

NYSED Identifier: The unique and permanent number assigned to a student as determined by the statewide process within New York

NYSED Encrypted Identifier: The encrypted form of a student's NYSED identifier that is used within the files of the NYSED

Permanent: When an identifier is used for an individual as long as any record of that individual is maintained in the database

Regional Information Centers (RICs): New York's regional support centers, which manage the student information systems for many school districts

Registrar: Any school representative who registers a student; the person who collects registration information from a parent or student, including the securing of an identifier or the assignment of a new one (In this use, registrar is a role or activity, not a job title.)

Registration: The provision by an individual of personal information to establish eligibility for enrollment (Registration may occur at a central site or at the school building. A student may be registered but not yet enrolled in a school.)

Secure: Describes information and data elements protected from unauthorized access by processes that manage access

Student: In New York, for the purposes of a statewide identifier system, an individual up to the age of 21 who is in contact with the New York education system for the purposes of being served

Student Information System (SIS): A computer application that maintains educational records for individual students

Student Number Ranges: Prescribed sets of potential identifiers allocated to a district or RIC for assignment to individual students (An actual range may not be contiguous, nor must all numbers within a range be assignable.)

Ubiquitous: When the same identifier or identifier system is used whenever and wherever data about an individual exists throughout an entity's information system

Unique: When an identifier appears only once in the pool of valid identifiers and is assigned to only one individual

Valid: An identifier that is from the authorized pool of identifiers and meets all criteria for being an identifier

Verification: Confirming that an identifier is correct for an individual

Verification Reference: A resource for looking up a student's identifier

ATTACHMENT A

Chart Notes

March 28, 2001

New York State Education Department

Discussion of Student Identifiers

Facilitated by Barbara S. Clements and Glynn D. Ligon, Evaluation Software Publishing

1988 Mandate –to assign a number to all New York students

- Public schools
- Non-Public schools
- Kept until but deleted at age 21
- Confidentiality required
- Assigned upon first entry into New York education system
- Provided for tracking and reporting
- Does not include home schools

Student groups included:

- Out of State
- Residential
- Atypical Schools
- Migrants

Early Intervention

- What should earliest age be?
- What is the practicality for early ages?

Need to create a list of all student groups getting IDs including:

- Migrant
- Spec Ed
- STAC

Assessment ID (within school)

- Progress by grade from K to 8
- Courses/credits for high school
- Exit exams –graduation and dropout

Links to Higher Education

- To vocational rehabilitation
- To GED
- Electronic transcript to college
- Research – consider both access, confidentiality

Characteristics of Current Systems

Rochester

- SIS
- IDs assigned to students “birth” to grade 12.
- 8/9 digit ID unique within Rochester
- No SSN
- Includes charter schools

NYC

- 9 digit
- “Any student” including summer school
- No SSN
- Unique within NYC

Yonkers

- 9 digit (6 digits with a zero fill)
- No SSN (also Orion #s – SASI)
- Not from NY range
- Central registration
- Unique IDs within NY

Buffalo

- 9 digit unique in Buffalo
- Some older student numbers kept to crosswalk
- No SSN

Nassau

- 7 different SIS systems
- Without unique IDs
- IDs use NY range

East Suffolk

- 4+ SIS systems
- Variety of ID number systems

MidHudson/Ulster

- Many systems not using the NY #s
- Variety of ID number systems

OCM

- “Birth to 12”
- Centralized SIS
- 9 digit shared/unique
- Unique across districts
- Using NY range

Wayne F-L

- Unique within district
- Not using NY range

Albany

- Unique within district
- Not using NY range

Lower Hudson/S. Wes---

- School or districts #s
- Not using NY range

SCT

- “All” SASI
- “Generally” unique across RIC
- Includes non-public
- Monroe
- Unique within districts in 2000
- 6 digit #s

Erie 1

- ½ districts use same SIS
- Within SIS unique #s
- ½ unique within district

LEAP ID

- 9 digit
- Unique to LEA
- LEA ID & student ID
- Grades 4, 8 & 5
- Q – is student ID permanent?
- Matching info not sent to NYSED
- Can’t track changes within the year

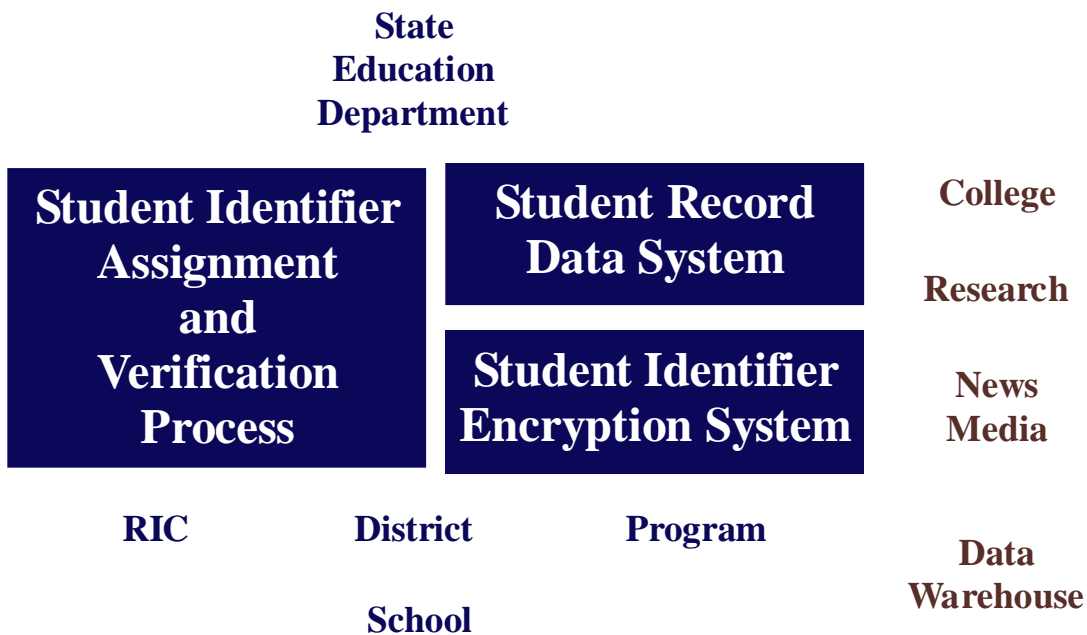
STAC

- Unique across programs
- 6 digit
- 2 ways to assign
 - Paper form...try to determine using name
 - Also check against Soundex
 - Use SIS ID or NYC ID
- Close matches checked by clerk
- Used as long as student within system
- 650,000 students

Possible IDs

- LEAP ID
- STAC ID
- SSN – encrypted or changed
- Random # - could be encrypted
- ID with information built in
- Random number from within an assigned range

Components of a statewide student information system



Locator System Characteristics

- Assigned at registration at school or program
 - +
- Enter into the district system
 - +
- Report to RIC
 - +
- Report to state

Locator Issues

- Children who don't want to be found

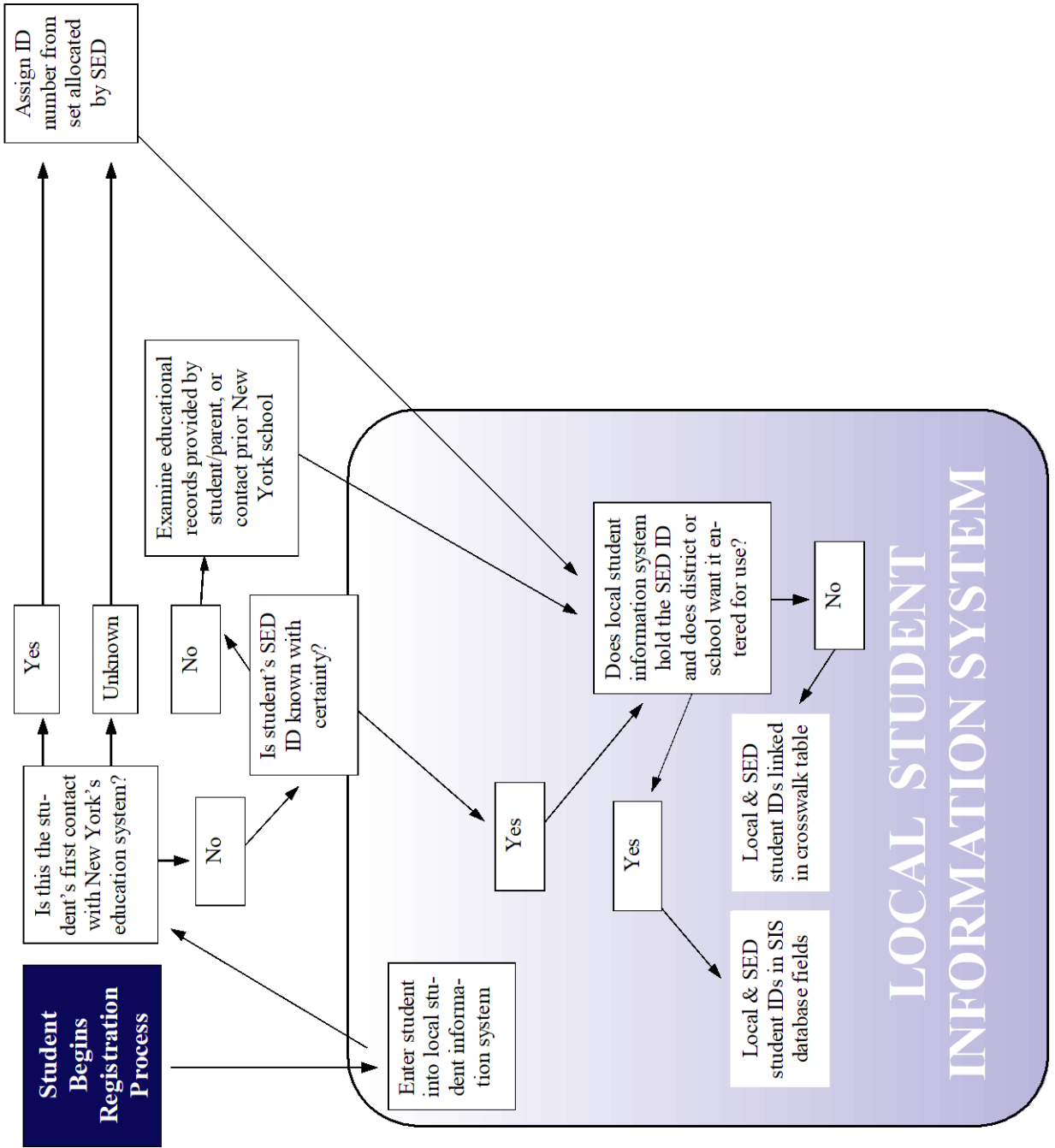
- Use for finding “lost” children
- Get model confidentiality policies; district, states
- Witness protection program required erasing all records
- Who has the right to “correct” data?
- Process to decide which data to accept/override
- Guidelines for Freedom of Information requests
- What constitutes directory information?
- Who will have access to the data?

On-Line Resources

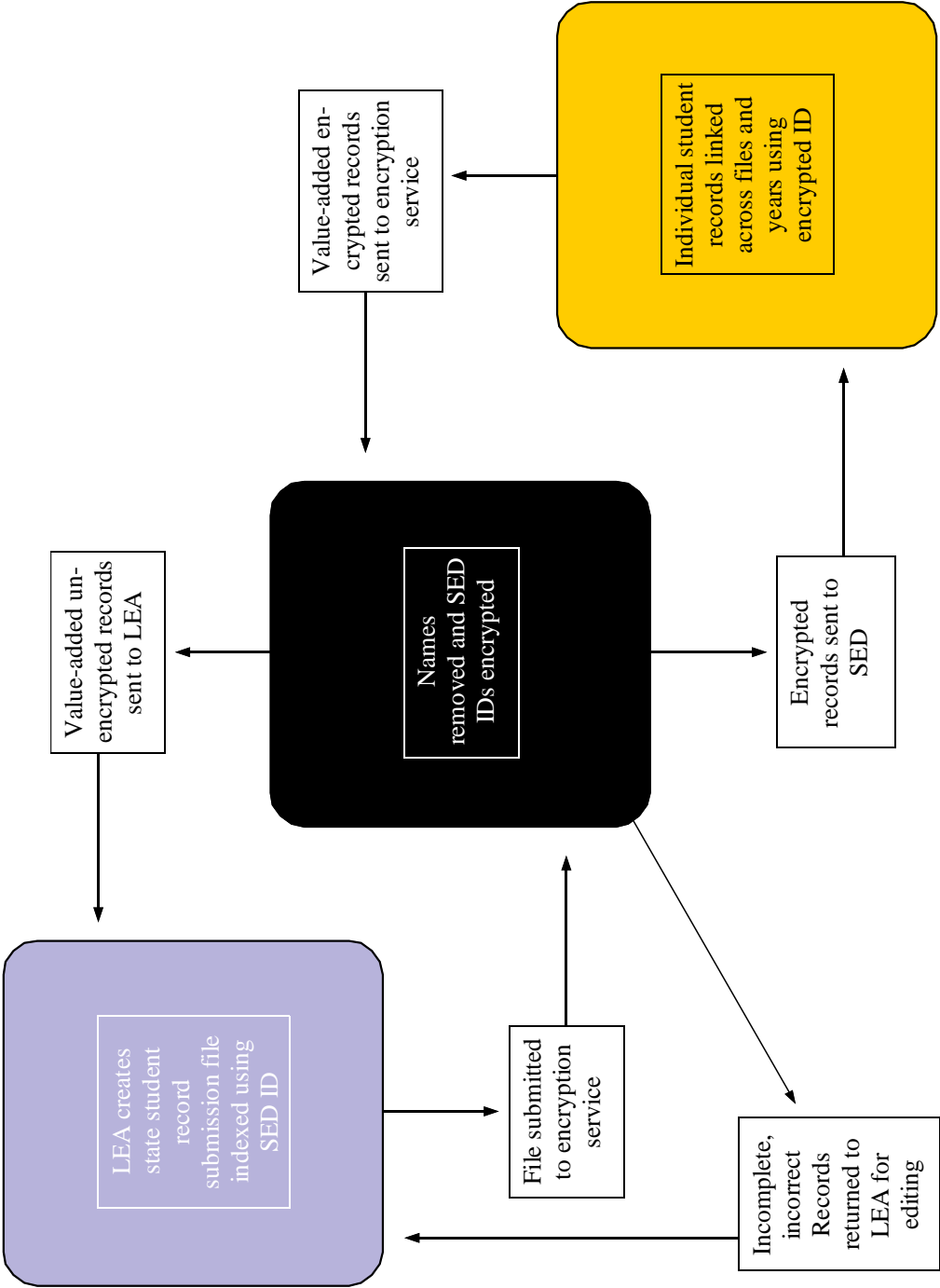
- www.educationadvisor.com
- www.nces.ed.gov
 - “Protecting the Privacy of Student Records”
 - “Safeguarding Your Security”
 - “Technology @ Your Fingertips”
 - “Automating Your Student Record System”

ATTACHMENT C

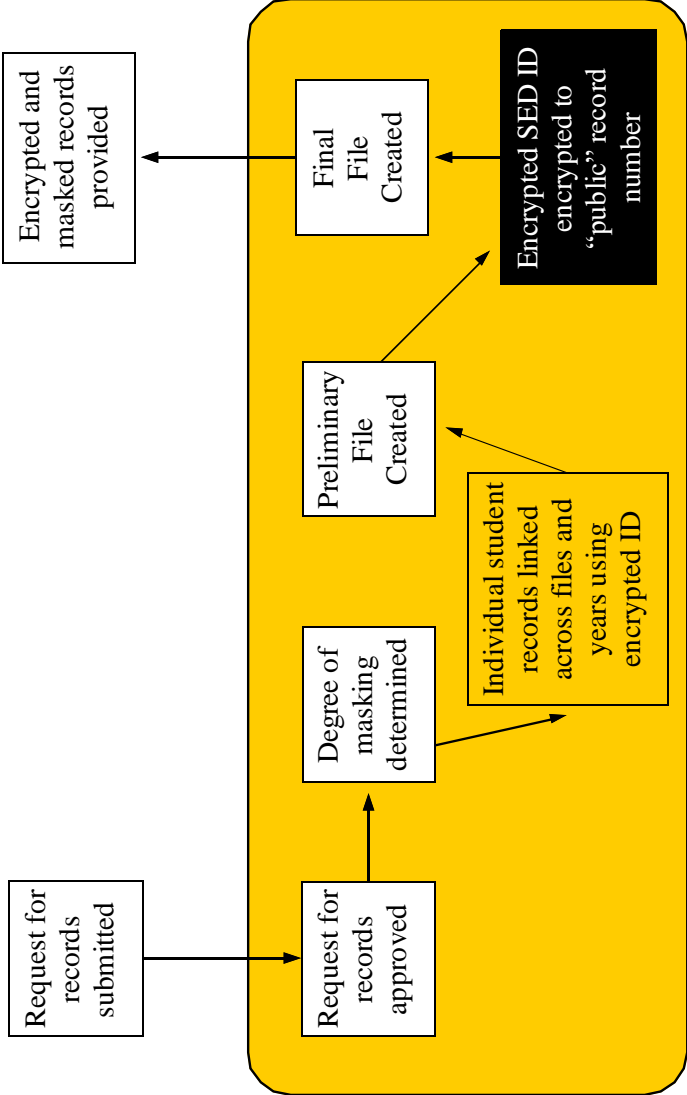
Basic Process for Assigning and Verifying Student IDs



Encryption Process for Protecting the Confidentiality of Student Data in SED Files



Process for Responding to Requests for Student Data in SED Files



Process for Exchanging Student Transcripts

