Transcript: Integrated Data Systems: Connecting Education Data to Local Communities

Slide 1:

Baron Rodriguez: Hello and welcome to the U.S. Department of Education and PTAC's webinar on Integrated Data Systems: Connecting Education Data to Local Communities.

My name is Baron Rodriguez and I am with the Privacy Technical Assistance Center. With me, is Michael Hawes who is the Director of Student Privacy Policy at the Office of the Chief Privacy Officer with the U.S. Dept. of Education

Slide 2:

Baron: Before we start, a few logistics for you. For now, we have muted all phone lines. This helps to ensure that all participants can hear and information can be provided uninterrupted. We do have a Q&A and chat feature in the right hand lower corner of your webinar screen. At any time, please feel free to insert your questions. We plan to review some of the questions at the end, and if we don't get to them today, we plan to review them for further clarification of the guidance in the coming months.

As always, this recording, and associated slide deck will be posted in a few weeks on the PTAC website.

Slide 3:

Baron: Today's webinar is intended to provide you with a high-level overview of our most recent guidance on Integrated Data Systems & Student Privacy. To read the full guidance, please go to ptac.ed.gov. It is available on the front page in the "recently released documents" section.

Slide 4:

Baron: For today's agenda, we plan to define what we mean when we use the term, Integrated Data System. There are many variations of Integrated Data Systems and our intent is to show you an example, specifically Allegheny County Pennsylvania. Our goal is to show you how this works in a local community based setting how privacy laws intersect with community based IDS' and highlight some best practice considerations as it relates to student privacy when considering implementation of an IDS. Ultimately, you and your leadership will need to decide if an IDS achieves the objectives needed to measure your program policy goals and if those objectives outweigh the potential risks and costs of implementing an IDS.

Slide 5:

Baron: This first visual displays the variety of possible external inputs of external administrative data that may be considered to be part of an IDS. Note that the possibility of a plethora of federal, state, and local

laws may play into any linked personally identifiable data depending on the agencies and/or the data involved. For instance, with Head Start data, you will need to consider the recently released head start regulations on privacy. With Health & Vital statistics, you may need to consider HIPPA implications and with Employment & earnings, you'll need to consider U.S. Department of Labor laws. As always, we recommend you consult with your local legal counsel to not only consider the federal laws, but also any state and local laws that may have additional protections and/or restrictions on sharing.

Slide 6:

Baron: For purposes of this discussion, we want to clarify our focus for the day. The visual on the left represents a State Longitudinal Data System hosted IDS. These systems are usually hosted at a state level and have many different considerations than those of local community based IDS'. For the purposes of today's discussion we will be focusing on IDS' that are hosting local data with a local government or local university. The example we will provide later in the webinar will describe how that IDS was structured from high level.

Slide 7:

Baron: This slide provides some examples of both state longitudinal IDS' and local based IDS'. The states in blue are the examples of established state systems with full K12 participation. The orange dots represent the locally established IDS'. For instance, Nashville public schools has been partnering with local agencies to determine the impact of after school programs on student's reading. They look at real time data on the students such as school attendance, behavior, and academic coursework achievements to determine which programs are a good fit for students. The advantage to using an IDS is access to real time data can be more useful than waiting until the end of the year to look at the results programs have on students. An IDS can also help determine other programs that may be of use for students and/or their parents such as homeless vouchers or availability of free food programs within the local community.

Keep in mind that with the significant advantages that integrated data systems provide, there are significant investments in technology as well as data sharing agreements to ensure that data is secure and access is only provided to those authorized.

Slide 8:

Baron: Allegheny County Department of Human Services is the example we wanted to share with you today. As you can see, they already host a significant amount of internal sources such as child welfare, homeless, mental health and some ancillary head start and early intervention education programs. The external sources include K12, specifically Pittsburgh Public Schools and surrounding districts, birth records, the housing authority, and juvenile probation. As you can imagine, these external sources can be a wealth of information to inform and improve the educational outcomes for children and students.

One specific advantage to the Department of Human Services hosting the IDS in this scenario is that they already fully understand HIPAA constraints and privacy regulations. As the owners of certain health data, they can control access to the linked data to only those who are authorized by their agency.

Later we will discuss how K12 is able to contribute to an IDS and be compliant with FERPA.

Slide 9:

Baron: This slide describes the primary purposes and scope of the Allegheny County Integrated Data System. You can see that they have specific, targeted goals such as addressing Chronic Absenteeism and improving school stability through improved placements of students. In addition, the integrated data allows Allegheny to look at homeless data to provide homeless prevention services at the school level.

Their model is to Analyze, critically reflect, and then take action based on data.

Slide 10:

Baron: In addition to the analytical data, when legally appropriate, the real time data from the IDS can allow caseworkers to have access to education records on whether foster children have been absent from school. Under the recently enacted Uninterrupted Scholars Act, Caseworkers are allowed to have access to foster children's education records. This allows caseworkers to make quick, informed decisions by seeing their children's education records integrated into their case management system.

Slide 11:

Baron: Depending on your familiarity with FERPA, it's important to remember the context of the elements that are relevant in the IDS context. Specifically FERPA protects those records that are related to a student and maintained by an educational institution. Generally, you need consent to share PII, however there are several exceptions such as the UISA that allow sharing under specific circumstances. We will talk through many of these in relation to IDS'. Also, keep in mind that FERPA consent/exception requirements do not apply to properly deidentified or aggregated data.

Michael Hawes: In order to discuss the creation and participation of a school district or state educational agency in an integrated data system, it's important to understand the legal requirements around the protection and use of education data in those contexts. We don't have time today to go through a full review of FERPA and its relevant provisions, but there are a few terms and definitions that it's important to understand in the IDS context.

At its core, FERPA requires the written consent of a parent or eligible student before disclosure of any personally identifiable information from the student's education record, unless an exception applies. There are a few terms in there that need to be parsed out a little bit.

The first is personally identifiable information. PII in the FERPA context has a very specific definition and includes the usual suspects: direct identifiers, like name social security number, student ID number, anything with a 1:1 relationship with the student. It also includes indirect identifiers, contextual information like a gender and race, anything that has a many:1 relationship with the student that could help someone find out who the student is. It also includes a third category in the legal definition, that is any information that alone or in combination with other information, would allow a reasonable person in the school community to re-identify the individual with any reasonable certainty. So if any linkable information could be used to identify a student, it counts as PII under the law.

Another term that needs definition in this context is education record. And education record is anything that's directly related to the student that's maintained either by or on behalf of an educational agency or institution. In the FERPA context, any PII from those educational records is protected from disclosure

without consent, unless an exception applies. Under the law, consent must be given by the parent or eligible student and must specify what information will be disclosed, to whom it can be disclosed, and for what purposes. There are exceptions to that consent requirement. The three most relevant exceptions in the context of integrated data systems are disclosure to other school officials, both within and outside the school, disclosures for audit and evaluation, and disclosures for certain types of studies. We will go through these exceptions in a little more detail later in the presentation.

Lastly, it's important to remember that there is a distinction under FERPA between disclosure of PII and disclosure of deidentified data. Truly deidentified data can be disclosed without a FERPA exception and without consent as long as you have gone through the proper steps to make sure that it is in fact deidentified.

Slide 12:

Michael: In order to participate in an integrated data system, a school district must ensure that there are appropriate agreements under FERPA for each of the aspects of participation in the IDS. The thing to remember here is that different exceptions may apply at different stages of the participation and use of the IDS.

It's helpful to think of this as a two-stage process. Different exceptions apply in the two different stages. This graphic highlights how you can approach this. In stage one you're talking about the establishment and decision to participate in the IDS, the initial provision of PII from education records to, in this case, the lead agency for the integrated data system. This is the provision of PII either by a school district or a state or local educational authority to the IDS lead for the purposes of establishing the linked data.

Then stage two, which is the use of the data that are linked and integrated within the IDS, includes approval or review of any particular uses that are being proposed for the integrated data, and ultimately the disclosure and or release of analyses, or the disclosure of identifiable data to a third-party for the purposes of conducting another evaluation or study. We'll talk more about these different steps in a moment.

Slide 13:

Michael: There are different ways that integrated data systems can be structured. You can have a fully federated integrated data system model, where the data are not residing in a single place, they are residing in their source agencies and are linked. That is not the model we are discussing today, although there are examples of fully federated integrated data systems out there.

Today we are talking about a data warehouse model: agencies provide data on a periodic basis to the IDS lead, where it is integrated and linked and then stored securely in that repository. It is important to remember that our focus today is on FERPA, but there may be additional federal, state and local laws that apply in your context. Always check with your own legal counsel whenever you are sharing data with third parties to make sure that you are abiding not just by FERPA, but also by any other laws that may also be relevant to you.

Slide 14:

Michael: What does it mean to participate or become a partner in an integrated data system? One step is that it requires a commitment to providing data to that integrated data system on a regular basis. An integrated data system is an ongoing project, not just a single linkage of data across agencies. It is the commitment to engage in ongoing evaluation or research that links data over time. It's not a single project, it's an ongoing relationship.

Another thing that participation in an IDS means is that there is going to be some entity that is managing that system who has ongoing access to either identifiable or semi-identifiable education records, and the fully identifiable or semi-identifiable data from those other agencies as well.

Participation in the IDS means a general openness to participate in multi-agency or cross-agency research and evaluation. Each individual project must be approved and be permissible under all of those relevant laws, but the purpose of an integrated data system is to engage in a cross-agency evaluation, where you're looking at the relative impact of multiple factors from multiple data sources.

The two FERPA exceptions to the consent requirement that are most relevant when establishing an integrated data system are the school official exception and the audit and evaluation exception.

The school official exception can only be used by school districts. This allows school districts to bring in third parties to perform services within the district. The school official exception can only be used by districts.

The other alternative is the audit and evaluation exception. This allows disclosure without consent for the purposes of auditing or evaluating a federal- or state-supported education program. This exception can be used for disclosure without consent by school districts and state and local educational authorities. Both will work for the establishment of the IDS.

Slide 15:

Michael: Let's talk about the audit and evaluation exception first, since this has the broadest relevance. The audit and evaluation exception allows a school district or a local or state educational authority to disclose PII from students' education records without consent to their authorized representatives. The regulation specifies the definition of authorize representatives, but it's essentially any third-party who is acting on behalf of the school district or SEA or LEA, who has been properly designated and who will be performing an audit or evaluation of a federal- or state-supported education program.

In terms of establishing the IDS, you have to think about the construction, integration and maintenance of the data as part and parcel of the audit and evaluation exception and the evaluation of an education program. This is necessary because integration of these data will then facilitate future audit and evaluation of these federal- or state-supported education programs by establishing and maintaining those linked data. This is permissible under the audit and evaluation exception, and in your written agreements between the school district, LEA, or SEA and the IDS lead, you would then specify how it fits in the audit and evaluation exception, that this purpose is to facilitate these future evaluations by providing and establishing linked data.

Slide 16:

Michael: The school official exception, which school districts and LEAS can use, allows districts to disclose PII from education records without parental consent, to a third-party, provided that

- third-party is performing some function or service that the district would otherwise use its own employees to perform
- The third-party is under the direct control of the district with regard to how they are using and maintaining the PII from education records
- the party can only use data in a manner consistent with how the school or district has a defined its school official's legitimate interest in its annual notice to parents of their rights under FERPA.
- The third-party cannot re-disclose the PII from education records exception under the direction of the school and as permitted by FERPA.

In this context, if this district wants to use the school official exception to establish an IDS, they are essentially saying that the IDS lead is serving as a school official for the purposes of creating an integrated data infrastructure, including the linkage of education data with other IDS data sources, while maintaining security of those data and controlling access. This can be seen as functions that the school district would otherwise perform themselves if they had the resources and the expertise.

In a nutshell, there are two possible exceptions that can be used to establish the IDS. School districts can use the school official exception, and schools, districts, local educational agencies, state educational agencies can use the audit and evaluation exception.

Slide 17:

Michael: Once you have established an integrated data system, you then want the data to be available to perform various cross-agency evaluations and research. The ideas for that research, and the proposals for how those linked data can be used, come from a variety of stakeholders, including government agencies, research institutions, the general public, service providers within the community. The IDS lead plays a central role in the access and data analysis process. The key thing to keep in mind with the evaluating and responding to these information request is the primary distinction between release of identifiable data and release of deidentified data.

Deidentified data can either be deidentified aggregate data or summary statistics, or it can be deidentified individual level data. The process you would use to deidentify it would differ in those two contexts.

The IDS lead must create this divide between what they are releasing generally to the public, as deidentified data, and what they're releasing under specific restrictions of the law in identifiable form.

Slide 18:

Michael: Sharing deidentified data is not considered a disclosure under FERPA. Deidentification as a process is trickier than a lot of people initially think. A first step to deidentify the data is removing all direct and indirect identifiers. This is where you strip off anything that can uniquely identify the student, be it name, student ID number, Social Security number, address, all of those direct and indirect identifiers that obviously re-identify an individual. That is not enough to fully deidentify a data file. A lot of other contextual information can be pieced together to re-identify an individual if you have enough pieces of information that you can link.

Deidentifying education data, and particularly of data that are linked between education and the different agencies participating in the IDS, will often require the application of one or more statistical methods to reduce the risk of reidentification from the resulting file. Those methods could include one or more applications of aggregation, suppression, blurring, or perturbation. We don't have time to go into the details of each, but we do have some resources on our website that we can point you to.

In a nutshell in order to release the deidentified data, you need to strip these direct identifiers, perform one or more statistical methods to reduce the reidentification risk, and you need to perform a disclosure avoidance analysis on the resulting file to determine if the risk of re-identification has been properly mitigated.

Slide 19:

Michael: If the IDS lead is looking to release identifiable data, then you have FERPA restrictions and the restrictions of other laws that come into play. The IDS lead may only re-disclose PII from a students' education record to a third-party in very limited situations in cases where the project falls under either FERPA's audit and evaluation exception or under FERPA's studies exception. In either case, all the required criteria that are specified in the regulations need to be met, and the recipient needs to comply with them.

In both cases a written agreement is required. That might be the written agreement that established the IDS, if it was properly laid out in the governments documents, or in many cases it may require the development of a separate agreement between the education agency and the third-party.

In any case, particularly under audit and evaluation, the purposes of the evaluation or study need to be properly spelled out in the terms of the agreement, along with provisions related to data security, data access, and the prevention of further re-disclosures. We have resources available online that can point out what the specific written agreement requirements are for these two exceptions.

Slide 20:

Michael: Regardless of which method you use to share data, whether it is identifiable data or deidentified data, the IDS lead can only make use of the data that are integrated into the IDS for the very specific purposes that are permitted within the legal framework that the education partner established with the IDS lead when becoming an IDS partner.

There is a spectrum along which those governance requirements can be established in those documents. On the end relating to most control, you need to have specific written agreements for each data use.

On the other end may be a number of pre-approved research questions that are already drafted into the governance document.

The level of control depends largely on how fleshed out the purposes of the IDS have been in advance of the creation of the IDS, and the level of trust that exist between the IDS lead and the education partner. There is no right or correct position, a lot of it will depend on the context of your specific community.

Slide 21:

Michael: It should be noted our discussion today has been on the use of integrated data systems for program evaluation and research. In most cases the exceptions we talked about will not permit the operation of an IDS to provide direct services to students. There are some specific situations in which direct service workers may be able to access PII from students' education records through the IDS for service provisions. We are not going to cover those today, but there are a few including for example the Uninterrupted Scholars Act.

Slide 22:

Michael: FERPA can be very complicated when you are looking to integrate education records with other data sources. The way to conceive it and do it within the confines of the law would be to conceptualize the creation and use of an IDS as a two-stage framework.

Stage one is the initial creation and governance structure of the IDS, where you are deciding to participate and provisioning the education records into the system.

Stage two is the use of the data by the IDS lead, working and approving research requests, and creating any outputs from those research or evaluation projects either to the public or to third-party entities.

Different FERPA exceptions apply in each case.

Slide 23:

Baron: Now that we've covered the legal requirements, let's discuss some best practices as it relates to data systems.

First, it's important to be very deliberate in communicating what data you are collecting and how it will benefit the community and/or parents. In absence of information, the community's natural tendency will be to assume the worst. Secondly, ensure that when you do communicate, you consider your wording to be something that is understandable by the general public. Consider running the language by a focus group or group of parents to get their initial reaction. It's important that when you do communicate, you consider multiple media, such as websites, flyers, and discussions at community meetings.

Finally, it's important to be transparent about your policies, procedures and technology to protect the data. Explain how access is limited to only those who need to know. Provide an example of why that person would need access to the integrated data (for instance the example we gave of a caseworker to intervene on behalf of a student who is consistently missing school). This also is an excellent example of showing the value of the combined data. Consider publishing what problems this system will help alleviate in your community. Make sure you also consider a feedback mechanism (such as a phone number or email address/website) for community members to voice their concerns.

Slide 24:

Baron: Let's get into more detail about some specific best practices.

- Decision Making Authority: This is arguably going to be your most contentious, yet most important structure you establish. In order to establish this, trust and formal agreements must be achieved. The objective of this recommendation is to assign appropriate levels of authority to agencies involved as it relates to individual data as well as the linked data sets. This also will define the roles of the agencies involved and formalize decision making for the IDS.

- Standard Policies and Procedures: Adopting and enforcing clear policies and procedures within an IDS data governance structure is necessary to ensure that everyone in the organization understands the importance of data quality and security—and that IDS staff are motivated and empowered to enforce these policies. Some example policies include, but are not limited to: Data release procedures, data breach policies and procedures, and data security standards and policies.

- Data Inventorying: Having an inventory of all data within the IDS that require protection is a critical step. Maintaining an up-to-date inventory of all sensitive records and data systems, including those used to store and process data, enables the IDS to target its data security and management efforts. Classifying data by sensitivity helps the data management team recognize where to focus security efforts.

- Data Records Management: Specifying appropriate managerial and user activities related to handling IDS data is necessary to provide data stewards and users with appropriate tools for complying with the IDS' security policies.

- Data Quality: Ensuring that data are accurate, relevant, timely, and complete for the purposes they are intended to be used is a high priority issue for any organization. The key to maintaining high quality data is a proactive approach to data governance that requires establishing and regularly updating strategies for preventing, detecting, and correcting errors and misuses of data.

- Data Access: Probably the second most important practice is defining and assigning differentiated levels of data access to individuals based on their roles and responsibilities within the IDS organization is critical to preventing unauthorized access and minimizing the risk of data breaches. We suggest data minimization, that is limiting access to the fewest number of individuals as possible with access to linked PII.

- Recordation: Keep in mind that FERPA's recordation requirements apply, so an IDS needs to have a mechanism to record any disclosures of education records containing PII.

- Data Security & Risk Management: Ensuring the security of sensitive and personally identifiable data and mitigating the risks of unauthorized disclosure of these data should be a top priority for an IDS. For instance has a comprehensive security framework been developed? Has a risk assessment of vulnerabilities related to inadvertent disclosures and malicious individuals? Do you have a plan for a data breach? Do you regularly monitor or audit your data security and/or access to the system? Do you regularly review your data sharing agreements for compliance with ever changing state, federal, and local laws?

Finally, as Michael discussed earlier, have appropriate procedures such as rounding and cell suppression being implemented properly to ensure there is no inadvertent disclosure of PII on any public reports out of the IDS?

Slide 25:

Baron: Ultimately, an IDS can be a tremendous asset for a local community. The ability to have access to real-time linked data can answer policy questions or provide case workers with information needed to ensure equitable and appropriate interventions and educational opportunities for children. These systems can provide the local community with empowering information regarding what programs are working well and assist the average citizen and policymaker with information not usually available.

It's very important to remember the potential true costs associated with linkage, technology, data sharing agreements and being appropriately transparent about an IDS. To implement the linked data is only a partial success. It's critical to ensure your system has the necessary controls and communication strategies to truly make the system a community asset.

Slide 26:

Baron: We've provided a list of resources as part of this webinar. These slides will be made available on our website at ptac.ed.gov in the next couple of weeks. We appreciate the opportunity to share this exciting new guidance and truly hope this webinar has been helpful to you.

Slide 27:

Baron: At this time, please type your questions in the chat box or Q & A box in the lower corner of your screen. We will pause for a couple of minutes to see if there are any questions and start the Q&A portion in 1 minute.

Should questions come up later, please send an email to PrivacyTA@ed.gov.

Thank you for participating in today's webinar.