



## P20 WIN DATA-SHARING SYSTEM

By: Marybeth Sullivan, Legislative Analyst II

### **PRESCHOOL THROUGH 20 WORKFORCE INFORMATION NETWORK SNAPSHOT**

- Participating agencies are the State Board of Education, the Board of Regents for Higher Education, and the Department of Labor
- Data allows the agencies to examine individuals' paths through the K-12 and higher education system into the workforce
- Agencies use the system to share data that is "scrubbed" of personally identifiable information by data-matching software

### **QUESTION**

How does the Preschool through 20 Workforce Information Network (P20 WIN) work? Can individuals opt out of it?

### **SUMMARY**

P20 WIN is a data-sharing system that temporarily collects data, along with individuals' personally identifiable information (PII), from existing data systems within participating agencies. Then, it centrally processes this information using data matching software, strips and destroys the PII, and sends the results to the requesting agency.

The State Board of Education (SBE) (on behalf of the State Department of Education (SDE)), the Board of Regents for Higher Education (BOR), and the Department of Labor (DOL) are currently implementing a pilot phase of the P20 WIN system.

Any of these three agencies may use the system to request data that tracks individuals through elementary and high school into higher education and the workforce.

Exceptions to the federal Family Education Rights and Privacy Act (FERPA) permit participating agencies to share student data among themselves without obtaining individuals' permission or offering an opt-out option. However, the agencies have pledged in their P20 WIN Memoranda of Agreement (MOA) to safeguard the shared data in a variety of ways.

## **SYSTEM OVERVIEW**

### ***P20 WIN Origins***

According to BOR's P20 WIN [website](#), the system's design process began with funding from the National Center for Educational Statistics at the U.S. Department of Education in 2009. SDE asked the Connecticut Department of Higher Education (later BOR) to manage the effort to link agency data, and participating agencies formed the Interoperability System Council (ISC) to develop a model for how the data could be linked.

### ***Participating Agencies***

SBE, BOR, and DOL are the three participating P20 WIN agencies. According to BOR, the system is designed with the flexibility to expand to include connections to the Office of Early Childhood (OEC), UConn, and other organizations if approved by the Office of the Attorney General and the P20 WIN Executive Board.

UConn indicated in public hearing [testimony](#) during the 2014 legislative session that it had also agreed to participate in the P20 WIN system and hopes to by the end the calendar year. We could not find any formal participation agreements signed by the university. (Previously, the agencies were working with the UConn Health Center to implement P20 WIN, but according to BOR, their agreement ended on March 31, 2014.)

OEC does not currently participate in P20 WIN. [PA 14-39](#) transferred preschool and school readiness program oversight from SDE to OEC, but it is unclear whether OEC will also maintain preschool data.

### ***Functions***

P20 WIN is designed to gather data from SBE, BOR, and DOL to identify, among other things:

1. high school indicators that predict students' success in college or the workplace,
2. teacher preparation programs that produce graduates whose students have the strongest academic growth, and
3. postsecondary certificates and degrees that lead to significant earning differences.

To do this, P20 WIN links individuals' data records across the agencies, which include:

1. education records of students in public preschool, elementary school, and high school, which SBE provides;
2. education records of students enrolled in Connecticut State Universities and community colleges, which BOR provides; and
3. workforce data, which DOL provides.

Two Memoranda of Agreement (MOA) govern the three participating agencies' P20 WIN collaboration: one between [SBE and DOL](#) and the other between [BOR and DOL](#).

## **SYSTEM GOVERNANCE**

Under the *P20 WIN Data Governance Policy*, three bodies share P20 WIN system leadership responsibilities: the Executive Board, the Data Governing Board, and the Data Steward Committee. Collectively, they provide policy creation and system implementation, maintenance, and improvement.

### ***Executive Board***

The chief executive officer of each participating agency sits on the Executive Board, which sets a prioritized research agenda and has the ultimate decision-making authority for P20 WIN. All decisions of the Executive Board must be made by a unanimous vote of its members.

### ***Data Governing Board***

Each Executive Board member appoints one member from his or her agency to serve on the Data Governing board. The members recommend policy and may commit resources to support their agencies' P20 WIN participation. Members also review and approve requests for data and establish and enforce policies related to cross-agency data management, including data confidentiality and alignment with the vision for P20 WIN and applicable law. All decisions of the Data Governing Board must be made by a unanimous vote of its members.

### ***Data Steward Committee***

The Data Steward Committee consists of two subject matter experts from each agency: a primary data user, such as an institutional researcher, and a primary information technology specialist. The committee develops recommendations and provides oversight for the technical implementation of the P20 WIN system and is responsible for the availability, security, and quality of data shared. It also (1) recommends to the Data Governing Board policies or practices that need to be developed or improved and (2) implements the necessary processes to carry out approved policies.

### **DATA TYPES AND COLLECTION PROCESS**

P20 WIN does not collect and store educational and workforce data. Instead, it temporarily collects such data, along with individuals' PII, from existing data systems within participating agencies. Then, it centrally processes agency data using DOL's data matching software, strips and destroys the PII, and sends the results to the requesting agency.

### ***Agency Data***

BOR lists the data elements that are part of education and workforce records shared between participating agencies in its online [P20 WIN Data Dictionary](#). Table 1 provides a selection of data elements by agency.

**Table 1: Selected Data Elements Shared Among P20 WIN Participating Agencies**

<b>Agency and Data Sources</b>	<b>Data Category</b>	<b>Data Elements</b>
<b>SBE</b> (through <a href="#">SDE</a> )  (Cohort of students who graduated in the 2009-10 academic year)	Assessment	Name of standardized test taken (e.g., CAPT, CMT, MAS, SAT, PSAT); school year and semester when test taken; student's grade level when test was administered; version/generation of test; student's performance level on test; conversion of student's raw score to a common scale to allow for numerical comparison between students
	Attendance	Days student attended; days student could have attended
	College	Type of college attended (e.g. two- or four-year, public or private); state where college was located; start and end date of college attendance
	Graduation	Name of high school and school district from which student graduated; Date of college graduation; major and type of degree earned
	Student	Birth year; gender; race and ethnicity; presence of a special education individualized education plan (IEP) at any time during high school; English Language Learner (ELL) status at any time during high school; eligibility for federal free/reduced lunch program
<b>BOR</b> (For regional community-technical colleges: academic years 1999-00 through present)  (For state universities: academic years 2009-10 through present)	Academic	Full- or part-time enrollment status; attempted and earned credit hours; enrollment in remedial English or mathematics courses; remedial and non-remedial credit hours attempted per term; GPA per term
	Financial aid	Amount of award; parent's and student's contribution toward the Expected Family Contribution; type of financial aid award (e.g., loan, school, grant, or work)
	Graduation	Anticipated graduation date; graduation status; age at graduation
	Placement test	Description of test taken (e.g., Accuplacer, SAT, ACT, Arnett Nursing); date on which test was taken; test score
	Program	Identification and description of major, program, and degree
	Student	Highest education level; term of first admission; current city, state, and zip code; race and ethnicity; gender; high school name and graduation year; higher education institution; semester of matriculation; State Assigned Student Identification (SASID) number; system-assigned unique identifier (PIDM); birth year; enrollment type (e.g., spring only; readmitted; non-credit; secondary, non-tech prep; continuing; transfer; new; visiting)
	Transfer	Start and end date of enrollment at transfer institution, as reported by the National Student Clearinghouse
<b>DOL</b> (4 <sup>th</sup> quarter 1999 through present)	Employment	Employee's first initial, last name, and social security number; quarter and year in which wages were received; total wages paid by all employers in quarter; industry code of employer paying highest wage

\*Source: P20 WIN Data Dictionary

**PII**

According to BOR, the agency data must be personally identifiable in order for DOL’s matching software to function. The matching software uses PII to match data about the same individual among multiple agencies. Appendix A in both MOAs lists the PII that must accompany agency data:

1. name (first, middle, last, suffix);
2. date of birth;
3. gender;
4. graduating high school code and name;
5. town of residence;
6. Social Security number;
7. state assigned student identifier (SASID); and
8. address.

**Data Collection and Sharing**

Agencies share PII only when P20 WIN receives a request for data from one of the three participating agencies that has been approved by the P20 WIN Data Governing Board. Table 2 summarizes the data collection and sharing procedure outlined in the MOAs using a hypothetical data request from BOR.

**Table 2: Data Collection and Sharing Procedure**

<b>Steps</b>	<b>Details</b>
1. BOR requests data from P20 WIN	<ul style="list-style-type: none"><li>• BOR submits a request to the Data Governing Board for the number of students who completed a high school bilingual program that went on to earn bachelor’s degrees from a Connecticut state university, along with their post-graduate wages earned.</li></ul>
2. Data Governing Board reviews the request	<ul style="list-style-type: none"><li>• Board members whose agency’s data is being requested (SDE, BOR, and DOL) determine whether they wish to share their data with BOR (sharing is not mandatory). The board approves or denies BOR’s data request.</li></ul>

Table 2: -Continued-

<b>Steps</b>	<b>Details</b>
3. SDE, BOR, and DOL assemble their data	<ul style="list-style-type: none"> <li>• Upon board approval, the agencies whose data is part of the request (SDE, BOR, and DOL) each assemble their own “input file.” Each sending agency’s input file contains separate, unique records for each individual. The sending agencies assign a unique generic ID to each unique record in the file.</li> <li>• Each sending agency’s input file is split into two separate files:               <ul style="list-style-type: none"> <li>- the “matching file,” which contains the unique generic ID and the PII fields required for matching to other agencies’ data fields, and</li> <li>- the “data file,” which contains the unique generic ID and all the data fields needed to satisfy the request, but not the PII fields used for matching.</li> </ul> </li> <li>• The sending agencies send their “matching file” to DOL and their “data file” to the requesting agency (BOR).</li> </ul>
4. DOL matches data between agencies and creates a matrix	<ul style="list-style-type: none"> <li>• DOL:               <ul style="list-style-type: none"> <li>- uses data-matching software to produce an “output file;”</li> <li>- strips and destroys PII fields from the matching file and the output file; and</li> <li>- uses the remaining data fields to create an ID matrix, which contains a “group” or “cluster” number and the unique generic ID’s of each record in each group.</li> </ul> </li> <li>• DOL sends the ID matrix to the requesting agency (BOR).</li> </ul>
5. BOR receives results	<ul style="list-style-type: none"> <li>• BOR receives the ID matrix from DOL and the data files from the sending agencies.</li> <li>• BOR uses the ID matrix as a key to link the data files for analysis.</li> </ul>

## **DATA ACCESS**

Under the P20 WIN system, the agency requesting data never receives individuals’ PII. Only the agency sending data and DOL have access to the PII, which is destroyed once data is matched between agencies and sent to the requesting agency. Also, only the three participating agencies may request data through P20 WIN.

### ***Opting-Out and FERPA***

Individuals cannot opt-out of P20 WIN, since FERPA and its implementing regulations allow the participating agencies to share students’ PII without their consent under certain circumstances.

FERPA generally gives a student the right to halt the release of PII and prohibits educational institutions from disclosing PII without the student’s written consent or, if the student is a minor, the student’s parents (20 USC § 1232g(a)-(b)). However, FERPA regulations permit education records to be disclosed to authorized representatives of state and local education agencies, among others, in connection

with an audit or evaluation of federal- or state-supported education programs, or for the enforcement of or compliance with federal legal requirements that relate to those programs (34 CFR § 99.31(a)(3)(iv)). Such disclosures do not require a student's consent.

## **AGENCIES' PRIVACY PLEDGES**

The P20 WIN MOAs and their appendices contain various pledges from the three participating agencies to protect individuals' data. They also contain a separate set of pledges from DOL.

### ***Agencies' Privacy Pledges***

All participating agencies agree that:

1. no data will be included in P20 WIN unless the sending agency approves (a) its inclusion and (b) the individual(s) or entities who have authority to access the resulting data set;
2. they reserve the right to audit policies, procedures, and systems involved in processing and safeguarding any PII and accompanying data;
3. they will comply with all applicable federal and state statutes and regulations, including FERPA, the Gramm-Leach-Bliley Act, the Health Insurance Portability and Accountability Act, and related state agency contracting policies in protecting all PII and other protected confidential information; and
4. they will each be responsible for losses arising out of (a) their own acts or omissions that result in a breach of PII or failure to comply with laws that protect confidential information and (b) their own negligence or misconduct.

### ***DOL's Privacy Pledges***

In addition to the above pledges, DOL agrees that it will, among other things:

1. transmit and store all education records using information security best practices, including commercially available firewalls, virus scanning, security access control software, logical data encryption, and secure tunnels;
2. limit physical access to confidential information and PII;
3. not transfer or disclose PII to any person or entity whose access to PII is not specifically authorized in the MOAs;

4. ensure that all data will be transmitted via secure connection, such as a Virtual Private Network (VPN) or Secure FTP File Transfer Protocol (SSL) connectivity;
5. not store data on a local drive;
6. not allow support personnel to store or transmit data through unencrypted email or on a portable storage device, such as a USB flash drive, cell phone, portable laptop, or external hard drive, except for encrypted system backup tapes;
7. require and maintain an appropriate confidentiality and nondisclosure statement from each employee or subcontractor with access to the data;
8. have data security policies and procedures that protect against FERPA violations when the data is at rest and during transmission;
9. allow the other participating agencies to perform on-site system audits and inspections to ensure confidentiality compliance;
10. maintain tracking logs to document data senders and recipients, transferred data, and destroyed data, among other things; and
11. permanently delete and shred education records used for matching within five business days of completing the matching process.

MS:ts