



## Student Privacy Website Review: LEA Sample Requirements

### LEA Sample Requirements

To support a review of the current state of privacy compliance and transparency relating to data practices in school districts across the country, a sampling plan is needed to determine which Local Education Agencies (LEAs) will be reviewed across four years of data collection (2019, 2020, 2021, and 2022). The goal of this study is derived from Performance Measure 3.2c of the U.S. Department of Education's Strategic Goal 3.2. We aim to calculate the "percentage of LEA websites from statistically representative sample reviewed for inclusion of transparency best practices and compliance with legal requirements relating to third party contracting." The goal is to take an annual snapshot of LEA websites for each year of the study, as well as conduct a final trend analysis, of how observations of LEA best practices and compliance in providing information on federal student privacy law changes over the four years of the study based on what is posted on their websites. Each year of the report will be nationally representative, and as will the final trend analysis. Each district included in the reviews will receive a letter report describing the results of their website review, in order to enable them to understand the specific ways in which the privacy protection information on their websites are consistent with best practices and the specific areas in which they could improve.

### District Sampling Frame

The foundation for the 2018-19 district frame will be the preliminary 2017-18 Common Core of Data (CCD) non-fiscal file. CCD is the Department of Education's primary database on public elementary and secondary education in the United States. CCD is a comprehensive, annual, national database of all public elementary and secondary schools and school districts.

The sampling frame for this study would differ from the CCD in the following ways:

- **LEA Types 1, 2, 7 and 9.** The goal is to include all districts that have students and where it would be reasonable to expect that they would have websites describing privacy practices. For New York City, we will remove all 33 Type 2 school districts and include it as one district. Types included are: Regular public school district that is NOT a component of a supervisory union (Type 1), Regular public school district that is a component of a supervisory union (Type 2), Independent Charter Districts (Type 7) and Specialized public school district (Type 9)<sup>1</sup>. This

<sup>1</sup> **Type 1:** Regular public school district that is NOT a component of a supervisory union: A regular public school district is a local government administrative authority that governs the education system at a specified local level on behalf of the public and the state that is not a component of a supervisory union.

**Type 2:** Regular public school district that is a component of a supervisory union: A regular public school district is a local government administrative authority that governs the education system at a specified local level on behalf of the public and the state that is component of a supervisory union.

**Type 7:** Independent Charter District: An independent charter district is an education unit created under the state charter legislation that is not under the administrative control of another local education agency and that operates one or more charter schools – and only charter schools.

**Type 9:** Specialized public school district: A specialized public school district is a school district that operates one or more schools that are designed for a specific educational need or purpose.

Source: FILE C029 – Directory File Specifications – V13.3 (SY 2016-17), U.S. Department of Education, Washington, DC: ED Facts. Retrieved December 27, 2018 from <http://www.ed.gov/edfacts>.



will exclude the following LEA types: Supervisory union, Service agency, State Agency, Federal Agency, and Other education agency.

- **LEAs currently actively open.** This would include LEAs with the following updated status based on the CCD: 1-Open, 3-New, 4-Added, 5-Changed Boundary/Agency, and 8-Reopened. This would exclude LEAs with the following updated status based on the CCD: 2-Closed, 6-Inactive, and 7-Future.
- **Only LEAs with students,** which would be determined by merging the LEA-level CCD data file with the school-level CCD data file. The school-level CCD data file would be restricted to schools with an enrollment of 1 student or greater. Only LEAs from the LEA-level CCD data file that match with the school-level CCD data file will be included in the frame. Those excluded districts appear to be administrative units rather than schools.
- **Exclude BIE schools.**

Note that the LEA sampling frame includes LEAs in the 50 states, District of Columbia, and Puerto Rico, as well as Department of Defense Education Schools and territories.<sup>2</sup>

Based on these criteria, we estimate a final frame of approximately 16,394 LEAs. The preliminary LEA-level CCD file had 18,883 records, of which ineligible districts were deleted based on the criteria above.

We propose to select a stratified random sample of LEAs based on the district frame described above. As such, we will conduct descriptive analyses on the sample based on geographic location, LEA level, and LEA size, but we do not expect to be able to make representative claims disaggregated at these levels in the annual reports. However, we do expect to be able to do so in the final report, which will be based on the full set of 1,504 LEAs.

## Sample Size

Determination of an appropriate sample size is typically based on four factors:

1. **Number of units in the target population.** As described above, the target population was calculated as 16,390 districts.
2. **Desired level of precision.** Based on standard practice, we are setting the desired confidence level at 95% and margin of error at 5%. Because the data collection has not been conducted yet and so we do not have an estimate for the standard deviation of responses, we are assuming a standard deviation of 0.5.
3. **Key analysis variables.** The variables used in the analysis will be the percentage of LEA websites that include various data transparency best practices for providing information on federal student privacy law compliance.

---

<sup>2</sup> Territories include American Samoa, Guam, Marshall Islands, Micronesia, Northern Marianas, Republic of Palau, and Virgin Islands.

- 4. Important reporting domains (i.e., subpopulations).** For the final report (across all four years), results may be broken out by various demographic characteristics such as % poverty, % minority, % ELL, locale type, and enrollment size. For each individual data collection year, the analysis will focus on overall results and not subgroup patterns, due to the relatively small sample sizes. Similarly, the trends analysis will also focus just on the overall patterns and not the subgroups.

Based on these four parameters, a minimum of 376 LEAs will need to be sampled annually, resulting in a total sample size of 1,504 LEAs across all four years of data collection.

### Sample Selection

We propose to select a single sample of 1,504 LEAs at the outset of the study, and randomly dividing that sample across the four data collection years. In addition, we will draw a replacement sample of 100 LEAs that may be used if any of the original sample are not feasible to include at the time of data collection (e.g., if LEA status changes from “open” to “closed,” “inactive,” or “future”, or if the LEA’s website is inactive). In such cases, an LEA from the replacement set will be randomly selected from the same stratum as the removed LEA. This process will ensure that an LEA is selected only once over the course of the study; the selected LEAs will be generalizable to LEAs that were in operation in 2017-18.

We propose to select a stratified random sample based on five enrollment size categories, as shown in Table 1.

Table 1: Proposed sample stratification, by enrollment size strata, 2017-18

Enrollment size	Total # of LEAs*	# of LEAs in sample	% sampled	Sample LEAs in each year	% of each year’s sample	Replacement sample
20 largest LEAs	20	20	100%	5	1%	NA
10,000 or more (excl 20 largest)	899	356	40%	89	24%	25
2,500 to 9,999	3,006	376	13%	94	25%	25
1,000 to 2,499	3,475	376	11%	94	25%	25
1 to 999	8,990	376	4%	94	25%	25
<b>Total</b>	<b>16,390</b>	<b>1,504</b>	<b>9%</b>	<b>376</b>	<b>100%</b>	<b>100</b>

Source: National Center for Education Statistics, Common Core of Data, 2017-18.

Oversampling larger districts, as shown above, will enable the project to provide feedback regarding privacy protection practices to school districts covering a relatively large percentage of the nation’s students. The 20 largest districts, which will be sampled with certainty, enroll 12% of the nation’s students and the remaining 1,484 enroll an estimated 27% of students, for a total coverage of 39% of all students.



In addition, oversampling larger districts will help to provide a relatively high concentration of districts in the sample that have higher concentrations of students from low-income families (i.e., eligible for free or reduced-price lunches (FRPL)), students from racial-ethnic minorities, and English language learners (ELLs). Table 2 shows the number of high-poverty (50-100% FRPL), majority-minority, and high-ELL (20-100% ELL) districts in each sample stratum.

Table 2: Number of LEAs that are high-poverty, majority-minority, and high-ELL, for each stratum of the sampling frame, 2017-18

Enrollment size	Number of LEAs in sampling frame			
	Total	High-poverty (50-100% FRPL) <sup>3</sup>	Majority-minority (50-100% non-white) <sup>4</sup>	High-ELL (20-100% ELL) <sup>5</sup>
20 largest LEAs	20	16 (80%)	20 (100%)	3 (15%)
10,000 or more (excl 20 largest)	899	449 (50%)	537 (60%)	174 (19%)
2,500 to 9,999	3,006	1,250 (42%)	953 (32%)	272 (9%)
1,000 to 2,499	3,475	1,390 (40%)	745 (21%)	178 (5%)
1 to 999	8,990	4,066 (45%)	2,687 (30%)	598 (7%)
<b>Total</b>	<b>16,390</b>	<b>7,171 (44%)</b>	<b>4,942 (30%)</b>	<b>1,225 (7%)</b>

Source: National Center for Education Statistics, Common Core of Data, 2017-18.

For each LEA, we will pull the following CCD information to support the sample selection, data collection, and descriptive analyses:

- LEA ID - NCES Agency Identification Number
- LEA Name - Education Agency Name
- LCity - Location city
- LState - Two-letter U.S. Postal Service abbreviation of the state in which the LEA is located
- LZIP - Location 5 digit ZIP code
- WEBSITE – The Uniform Resource Locator (URL) for the unique address of a Web Page of an education entity
- UPDATED\_STATUS- The updated operating status of the LEA
- LEA\_Type - LEA type (the "Categorical Values" column shows the actual code with the text equivalent)
- LEVEL - LEA level
- Locale code
- Fall enrollment
- % FRPL
- % Minority
- % ELL

<sup>3</sup> 9.2% LEAs having missing information for FRPL. Of the 16,390 eligible LEAs, 1,501 do not valid information for at least 1 school.

<sup>4</sup> Only 133 LEAs do not have valid information.

<sup>5</sup> Over a fifth of LEAs do not have valid information (3,888).