June 28, 2013

The Honorable John Kline
Chairman
Committee on Education and the Workforce
House of Representatives

The Honorable Richard Hanna
House of Representatives

Status of the Department of Education's Inventory of Its Data Collections

The Department of Education (Education) administers numerous data collections to a variety of entities. Some of these collections are statistical in nature and are primarily used for research purposes and others are associated with grant program administration or program monitoring. Recipients of federal funds, such as states, school districts, and postsecondary institutions, must comply with federal requirements, many of which involve collecting and reporting data to Education for several purposes, including accountability, monitoring, and compliance with civil rights laws. Education's data collections generally must be approved by the Office of Management and Budget (OMB), in accordance with the requirements of the Paperwork Reduction Act of 1980, as amended.[1]

In 2012, we reported that states and school districts cited data collection and reporting among the most burdensome elementary and secondary education program requirements, due in part to being unnecessarily duplicative.[2] Accordingly, you asked us to determine the breadth of Education's data collection efforts, including identifying the individual data elements collected from respondents. Our initial work shows that Education is compiling an inventory of its data collections; thus, this report addresses the following questions:

1. What information will the inventory of Education's data collections contain and when will the inventory be completed?

2. What process is Education using to catalog its data collections, and to what extent does that process include internal controls to ensure the accuracy of the information collected?

3. What are Education's plans to make its data collection inventory publicly available?

---

[1]Paperwork Reduction Act of 1980, Pub. L. No. 96-511, 94 Stat. 2812 (codified as amended at 44 U.S.C. §§ 3501-3521). In this report, we use the term "data collection" to mean a "collection of information" as defined by the Paperwork Reduction Act.

[2]See GAO, *K-12 Education: Selected States and School Districts Cited Numerous Federal Requirements As Burdensome, While Recognizing Some Benefits*, GAO-12-672 (Washington, D.C.: June 27, 2012).

To determine what information the inventory will contain as well as Education's process for cataloging its inventory, we reviewed relevant documentation on the inventory's structure, contents, data fields, and source documents used to populate the inventory and interviewed Education officials and contractors responsible for designing, implementing, and maintaining the data inventory. In addition, we applied relevant federal internal control standards for information processing to assess the extent to which Education's process ensures the accuracy of its inventory. We also interviewed Education officials responsible for the inventory about who they consulted in designing the inventory, and officials from three Education offices: the Office of Elementary and Secondary Education, the Office for Civil Rights, and the Performance Information Management Service, about any input they provided on the inventory. We selected these offices because their data collections are among the largest and most complete of those incorporated into the inventory to date. To obtain an understanding of Education's plan for making it publicly available, we interviewed Education officials and the department's representative on the interagency Data.gov team.[3] In addition, we interviewed OMB officials about any existing or forthcoming guidance relevant to the inventory.

We conducted this performance audit from December 2012 to June 2013 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

In summary, we found that:

- Education's inventory is expected to include descriptive information (referred to as metadata)[4] about the context of each data collection, as well as the specific data elements reported by respondents for each collection. Education officials said that the inventory will contain the department's major data collections and will be ready for public release by November 2013. The inventory is to eventually include all statistical and grant administration collections that meet Education's definition of a data collection.[5]

- Education has followed a reasonable process to populate its inventory and has designed appropriate internal controls to ensure the accuracy of information included, such as reviewing and verifying data entered into the inventory. Education solicited input on its data inventory design from internal stakeholders through a team formed to address data coordination efforts across the various program offices. Officials from the three Education offices we spoke with stated that they participated in high-level meetings to discuss the inventory.

---

[3]Launched in 2009, Data.gov is a centralized website for storing and accessing publicly-available data from federal agencies. Data.gov was developed and has been operating under inter-agency leadership with support from OMB. Education's inventory is a separate effort from Data.gov.

[4]Metadata describe a data collection, including the name of the collection, collection dates, and respondent characteristics. For the purposes of our report, we distinguish metadata from the specific data elements that Education collects from respondents, such as enrollment type or attendance rates.

[5]For the purposes of its inventory, Education defines a data collection as data it collects for program administration or statistical purposes, and for which it was required to request approval from OMB under the Paperwork Reduction Act.

- Education plans to make its inventory publicly available through a searchable web database by November 2013. The web database may be launched through the department's website or through Data.gov.

Background

Education collects data from various entities on topics associated with pre-kindergarten through higher education. Some information Education collects is required by federal statute or regulation. For instance, both the Elementary and Secondary Education Act of 1965 (ESEA) and the Individuals with Disabilities Education Act (IDEA) contain provisions that require states and school districts to collect data or other information and report those data to Education. While the department uses reported information for a variety of purposes, it primarily uses it to provide support and oversight of states and school districts. To assist states in reporting required data and the department in managing the data, Education created a data system, known as EDFacts.[6] States use this system to report to Education certain data required by ESEA and IDEA, among other required information. Recipients of federal funds under these and other programs must also comply with other federal data collection requirements. For example, since 1968, Education has administered a survey—known as the Civil Rights Data Collection—that helps it administer and enforce federal civil rights laws, which generally prohibit discrimination on the basis of race, color or national origin, sex, disability, or certain other characteristics. In addition, Education conducts other data collections primarily used for research purposes, such as the National Household Education Survey and National Postsecondary Student Aid Study.

Under the Paperwork Reduction Act, federal agencies, including Education, are generally required to obtain OMB approval prior to requiring or requesting the disclosure of information.[7] To seek OMB review and approval of a proposed data collection, an agency submits an Information Collection Request package to OMB. The package includes a description of the collection, its planned use by the agency, and other information and supplemental materials. The agency must certify that, among other things, the collection is necessary for the proper performance of the agency's functions, and does not unnecessarily duplicate otherwise accessible information. OMB can approve data collections for up to 3 years.[8] Agencies must re-submit a collection request package at the end of the approved period or when making changes to a collection.

---

[6]The EDFacts system centralizes state-reported K-12 educational performance data into one federally coordinated data repository. Through the EDFacts collection, states submit K-12 education data to Education on approximately 180 data groups at the state, district, and school levels. EDFacts data include information on schools, services, staff, students, and educational outcomes.

[7]44 U.S.C. § 3507(a). The Paperwork Reduction Act requires that agencies take specified steps before conducting or sponsoring certain data collections, including reviewing the collection and soliciting public comment, as well as obtaining OMB approval. Agencies must follow these requirements before requiring or requesting information from (1) 10 or more non-federal persons, including state and local governments; or (2) federal agencies or employees for general statistical purposes. 44 U.S.C. § 3502.

[8]In general, OMB may approve the Information Collection Request with or without change, disapprove it, or indicate that the request was improperly submitted. An agency may also withdraw its request at any time.

In September 2010, Education formed its Data Strategy Team[9] to address a perceived lack of coordination on data collections across its program offices, according to Education officials. In February 2011, the Data Strategy Team first discussed creating an inventory of the department's data collections as a logical step to coordinate efforts. Shortly after, the Data Strategy Team formed a working group for the inventory, chaired by the National Center for Education Statistics (NCES), and invited all program offices to name a representative to the working group. That group now has about 20 core members and regularly briefs the Data Strategy Team on the status of the inventory project, according to Education officials.

Education's Inventory Is Designed to Include Descriptive Information and Specific Data Elements on Its Major Data Collections

Education's inventory will contain over 300 fields of descriptive information (metadata) about the context of each data collection. According to Education officials, the majority of these metadata fields are taken directly from the supporting documentation for the relevant Information Collection Request package that Education submitted to OMB for approval[10] and these fields cover a wide range of topics, including:

- the relevant program within Education and program description;

- type of respondent (e.g., state, postsecondary institution, etc.);

- purpose of the collection listed in the Information Collection Request package;

- whether the collection is voluntary or mandatory (i.e., the respondent is required by law to provide the information);

- whether the collection is part of program monitoring and/or grant monitoring compliance;

- frequency and most recent year of the data collection;

- abstracts for the data collection and any follow-up collections; and

- estimated burden time for respondents to comply with the collection.

The metadata fields are largely taken from the Information Collection Request package, so certain metadata fields that are not in the request packages will be excluded from the inventory, according to Education officials. For example, while the inventory includes a metadata field citing the statutory authorization for a data collection, it does not include citations for any regulatory authority. Although several fields relate to a collection's purpose, there are no metadata on how, specifically, Education uses the individual data elements collected to monitor a program or grantee.

In addition to metadata, each collection in the inventory will be linked to a list of the specific data elements collected from respondents. For example, the list of data elements collected under the

---

[9]The Data Strategy Team is co-chaired by officials from Education's Office of Planning, Evaluation, and Policy Development; National Center for Education Statistics; and Privacy, Information and Records Management Services.

[10] Information is also obtained from the program website and in consultation with program officials when it is not available in the Information Collection Request package.

School Improvement Grants study within EDFacts includes student attendance rate and the number of students who complete advanced coursework.

Education officials estimate that the metadata and data elements for some of its major data collections will be finalized and included in the inventory by June 2013. These include EDFacts, the Civil Rights Data Collection, and some statistical data collections conducted by NCES. Another group of collections is expected to be finalized and included in the inventory by November 2013, including the Common Core of Data and Integrated Postsecondary Education Data System.[11] (See enc. I for a list of data collections and projected time frames for completion.)

Education officials said that other data collections, including those pertaining to federal student aid programs, will be added as work continues on the data inventory. While officials told us that they do not yet know the total number of collections to be included in the inventory, they anticipate that the inventory will eventually include all data collections conducted across the department that require OMB approval, including those conducted for program administration or statistical purposes. NCES officials told us that the inventory working group will meet in June 2013 to begin identifying the remaining data collections to add to the inventory. These officials said that most of the remaining collections are annual performance reports, such as those provided by grantees. In contrast to the collections currently being entered into the inventory, which produces data files that can be manipulated by users, these remaining collections largely collect information in text form, which is not as easily manipulated as data. Consequently, the working group will need to decide which of these annual reports will be considered "data collections" for the purposes of the inventory.

Education Has Followed a Reasonable Process to Populate the Inventory Based on Input from Internal Stakeholders

We found that Education has implemented a systematic process for populating the inventory that includes appropriate control policies and procedures. According to federal government standards for internal controls, management should establish control mechanisms and activities to ensure that internal control objectives are achieved, such as ensuring the accuracy and validity of transactions during information processing.[12] Education has implemented a variety of control activities for entering, reviewing, editing, and verifying the information about the data collections included in the inventory (see fig. 1). NCES staff conducted initial data entry for the inventory, and contractors have been assigned to continue populating the inventory.[13] Metadata entered into the database by contract staff are independently reviewed and verified by other contract staff, contract project managers, and NCES staff. After necessary edits are made, a report containing the metadata fields is generated for review and verification by program staff overseeing the particular data collection. Together, NCES and program staff discuss the report and further edits are made to the metadata as needed. Similarly, the list of data elements

---

[11]The Common Core of Data and the Integrated Postsecondary Education Data System are both NCES programs. The Common Core of Data collects annual fiscal and non-fiscal data from state educational agencies about public elementary and secondary schools in the United States. The Integrated Postsecondary Education Data System gathers institution-level data in such areas as enrollment, program completions, faculty, staff, and finances from primary providers of postsecondary education.

[12]See GAO, *Standards for Internal Control in the Federal Government*, GAO/AIMD-00-21.3.1 (Washington, D.C.: November 1999).

[13]NCES officials assigned this task and other inventory work under an existing contract.

collected from respondents is prepared by Education staff, uploaded to the inventory by contract staff, and then reviewed by program staff for accuracy.[14]

Figure 1: Education's Process for Populating Its Data Inventory
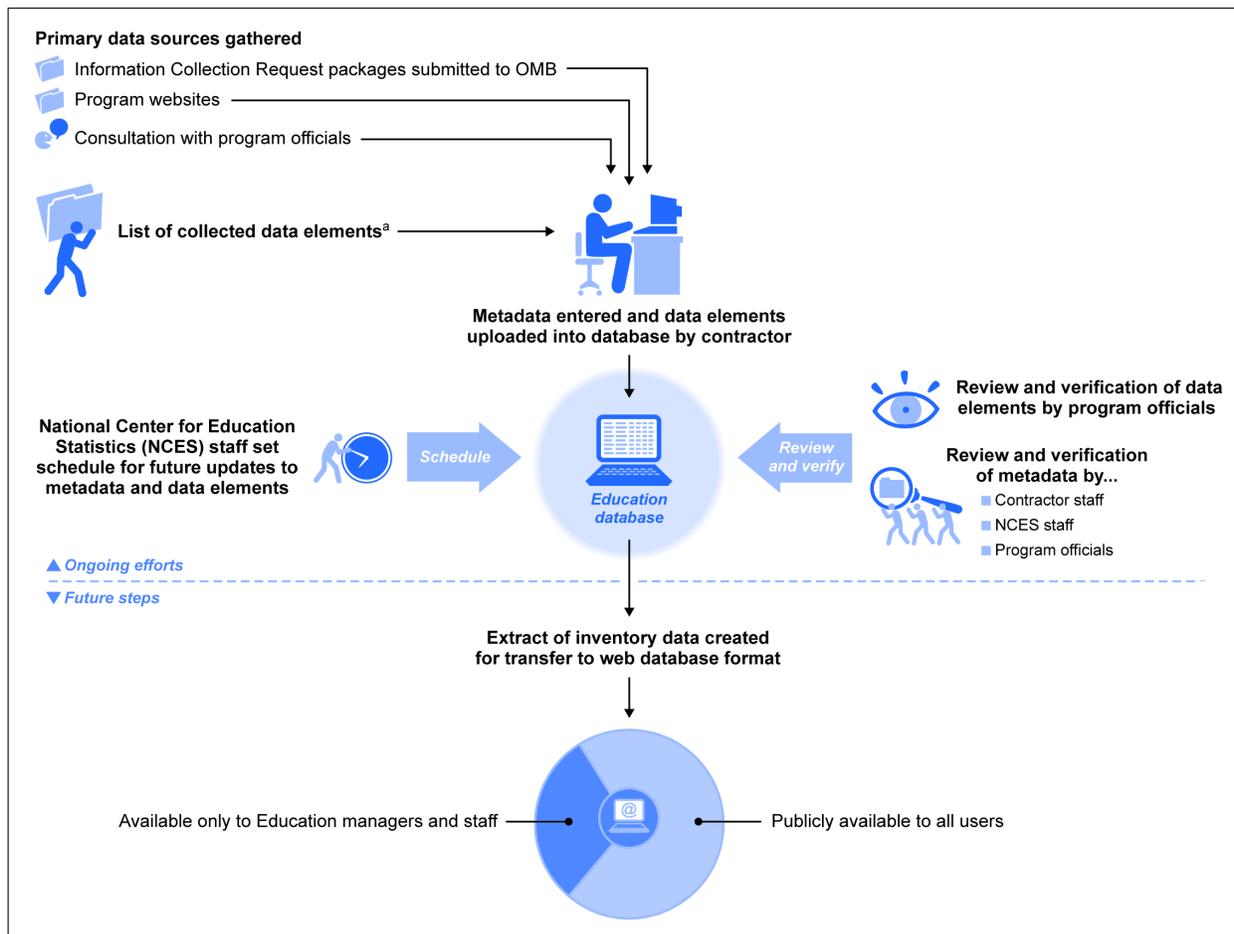
Source: GAO analysis of information provided by Education.

[a] The list of collected data elements is provided in a file or online source by program office staff, or NCES staff extracts the list from existing data codebooks.

NCES officials told us they had not yet established a firm plan regarding who will be responsible for updating the inventory in the future. They said one option would be asking program offices to update their respective data collections.

As members of Education's Data Strategy Team, representatives from various program offices provided significant input to the inventory's design, according to Education officials. Officials we interviewed from three Education offices said that they were aware of the data inventory, and

---

[14]According to Education officials, all data collections included in the inventory to date have existing electronic data files from which a list of data elements can be generated by program staff. For other data collections that Education has yet to identify for inclusion in the inventory, staff may have to use a different process to identify data elements and enter them in the inventory.

representatives from two of these offices confirmed that they participated in the inventory working group, which included attending meetings to discuss the purpose and design of the inventory. Officials from the third office said their office chose not to participate in the inventory working group but received regular updates at Data Strategy Team meetings. Officials from all three offices said they anticipate using the inventory to identify any areas of overlap and inconsistent data definitions among the various data collections across program offices. Officials from two offices said their staff would also use the inventory to avoid duplication before undertaking any new collections. Education officials told us they have not yet articulated in their written policies and procedures how the inventory might be used, but plan to do so once the inventory is in place.

Education officials told us that because they originally intended the inventory to be used by internal stakeholders only, they did not solicit input from stakeholders outside the department during the development process. However, they plan to solicit comments from the public once the inventory is made publicly available. OMB officials told us they were not involved in Education's inventory initiative, but said they considered the creation of a data collection inventory to be a best practice for federal agencies. OMB officials added that they were not aware of similarly comprehensive inventories at any other agencies, noting that Education was a leader in this practice. In May 2013, OMB released a memorandum establishing a framework for effective information management at federal agencies. OMB officials we interviewed prior to the memorandum's release said that Education's inventory initiative appeared to be in line with the requirements in the memorandum.[15] After the memorandum's release, Education officials said they believed their inventory was generally in line with the memorandum, although they plan to review the memorandum more thoroughly and make adjustments as needed.

Education Plans to Launch a Publicly Searchable Web Database

Education plans to make its data collection inventory publicly available through a searchable web database in November 2013. Officials said that many of the inventory's metadata fields and all lists of data elements collected will be available to the public, while other fields, such as those Education uses to track its collections internally, will be for internal use only. According to Education officials, the web version of the inventory will allow users to search for data collections by keyword, including searching for data elements collected, or to browse data collections by topic.

Once the inventory is populated with metadata on the initial set of data collections, NCES officials said they had planned to move the inventory to an updated version of the Data.gov website.[16] However, Education officials said that until a prototype of the updated Data.gov web interface is released, Education plans to develop its own web interface prototype. According to these officials, whether the inventory is launched through the department's website or through Data.gov will depend on whether the new Data.gov interface is available in time for Education's

---

[15]OMB, *Open Data Policy: Managing Information as an Asset,* OMB Memorandum M-13-13 (Washington, D.C.: May 9, 2013). Issued pursuant to the Executive Order of May 9, 2013, *Making Open and Machine Readable the New Default for Government Information*, this memorandum directs agencies to manage information as an asset throughout its life cycle to promote openness and interoperability, and properly safeguard systems and information. The memorandum establishes specific requirements and provides implementation guidance to help agencies comply with the new policy.

[16]Education has a representative on an interagency team working to expand the Data.gov website under the White House's Open Data policy. In the fall of 2012, Education's representative on this team began working with the inventory working group on plans to make the inventory public through Data.gov.

planned launch of the inventory in fall 2013. Education officials expect their prototype to be completed by mid-summer 2013, at which time Education will conduct usability testing within the department and possibly with select outside users. Officials said that regardless of which interface they choose, they will solicit public comments on the inventory once it becomes publicly available.

Agency Comments

We provided a draft of this report to the Department of Education for its review and comment and relevant excerpts to the Office of Management and Budget for technical review. In its written comments, reproduced in enclosure II, Education said that it appreciated GAO's efforts to review and understand the breadth of the Department's data collection efforts. Education also recognized the importance of compiling a comprehensive catalog of data collections, using it to help reduce reporting burden where possible, and promoting open and transparent government by making the catalog publicly available.

_____

As agreed with your office, unless you publicly announce its contents earlier, we plan no further distribution of this report until 30 days from its issue date. At that time, we will send copies of this report to the Secretary of Education, relevant congressional committees, and other interested parties. In addition, this report will be available at no charge on GAO's website at http://www.gao.gov.

If you or your staff members have questions about this report, please contact me at (202) 512-7215 or scottg@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this report. Key contributors to this report were Meeta Engle, Assistant Director; James Bennett; Susan Chin; Sarah Cornetto; Lauren Gilbertson; Kirsten B. Lauber; and Kathleen van Gelder. Also contributing to this report were Deborah Bland and Julie DeVault.
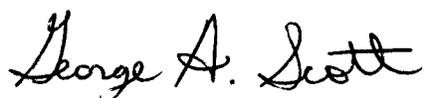
George A. Scott

George A. Scott
Director
Education, Workforce,
  and Income Security Issues

Enclosures (2)

Enclosure I: List of Data Collections in Education's Data Inventory by Projected Date of Completion

	Collections Education Projects to Be Included in the Inventory by June 2013
1.	EDFacts Data Collection[a]
	1.1 EDFacts Accountability 2011-12
	1.2 EDFacts Assessment 2011-12
	1.3 EDFacts Career and Technical Education 2011-12
	1.4 EDFacts Charter Schools 2011-12
	1.5 EDFacts Elementary and Secondary Education Act (ESEA) Flexibility 2011-12
	1.6 EDFacts Financial Data 2011-12
	1.7 EDFacts General Information 2011-12
	1.8 EDFacts Graduates and Dropouts 2011-12
	1.9 EDFacts Homeless, Neglected, or Delinquent 2011-12
	1.10 EDFacts Migrant Education Program 2011-12
	1.11 EDFacts Race To The Top 2011-12 (2 data groups)
	1.12 EDFacts Safe and Drug-Free Schools 2011-12 (7 data groups)
	1.13 EDFacts School Improvement Grant (SIG) 2011-12 (15 data groups)
	1.14 EDFacts Special Education/Individuals with Disabilities Education Act (IDEA) 2011-12
	1.15 EDFacts Teacher Quality 2011-12 (3 groups)
	1.16 EDFacts Title I 2011-12 (18 data groups)
	1.17 EDFacts Title III/Limited English Proficiency 2011-12 (13 data groups)
2.	School Survey of Crime and Safety (SSOCS)
3.	Private School Universe Survey (PSS)
4.	Civil Rights Data Collection (CRDC)
5.	Schools and Staffing Survey (SASS)
6.	Teacher Follow-Up Survey (TFS) after administration of the SASS
7.	Principal Follow-Up Survey (PFS)
8.	National Public Education Financial Survey (NPEFS)
9.	High School Longitudinal Study (HSLS)
10.	Early Childhood Longitudinal Study Kindergarten Cohort (ECLS-K)
11.	Early Childhood Longitudinal Study Birth Cohort (ECLS-B)
12.	Early Childhood Longitudinal Study Kindergarten Cohort 2011 (ECLS-K: 2011)
13.	Fast Response Survey System Arts in Public Schools (FRSS-Arts)
14.	National Household Education Survey (NHES)
15.	Education Longitudinal Study of 2002 (ELS)
16.	IDEA Part C – State Performance Plan (SPP) and Annual Performance Report (APR)
17.	Teacher Compensation Survey (TCS)
18.	Consolidated State Performance Report (CSPR)
19.	IDEA Part B State Performance Plan (SPP) and Annual Performance Report (APR)

	Collections Education Projects to Be Included in the Inventory by November 2013
1.	Academic Library Survey (ALS)
2.	Adult Literacy and Lifeskills Survey (ALL)
3.	Assessment and Teaching of 21st Century Skills Pilot Test (ATC21S)
4.	Baccalaureate and Beyond (B&B)
5.	Beginning Postsecondary Study (BPS)
6.	Beginning Teacher Longitudinal Study (BTLS)
7.	Consolidated Annual Report (CAR)
8.	Common Core of Data (CCD)
9.	School District Finance Survey (F-33)
10.	Fast Response Survey System (FRSS) on various topics
11.	High School Completion Validation Study (HSCVS)
12.	Integrated Postsecondary Education Data System (IPEDS)
13.	National Adult Training and Education Survey (NATES)
14.	National Assessment of Adult Literacy (NAAL)
15.	National Assessment of Educational Progress (NAEP)
16.	National Reporting System (NRS) for Adult Education
17.	National Indian Education Study (NIES)
18.	National Longitudinal Transition Study-2 (NLTS2)
19.	National Postsecondary Student Aid Study (NPSAS)
20.	National Study of Postsecondary Faculty (NSOPF)
21.	Postsecondary Education Quick Information System (PEQIS)
22.	Pre-Elementary Education Longitudinal Study (PEELS)
23.	Program for International Student Assessment (PISA)
24.	Program for the International Assessment of Adult Competencies (PIAAC)
25.	Progress in International Reading Literacy Study (PIRLS)
26.	Public Library Survey (PLS)
27.	State Library Agencies Survey (StLA)
28.	Teaching and Learning International Survey (TALIS)
29.	Trends in International Mathematics and Science Study (TIMSS)
	Other Collections to Be Included in the Inventory After November 2013
	Information collections related to federal student aid programs and other currently unidentified data collections

Source: GAO analysis of Education documents.

[a] The EDFacts system centralizes state-reported K-12 educational performance data into one federally coordinated data repository. Through the EDFacts collection, states submit K-12 education data to Education on approximately 180 data groups at the state, district, and school levels. EDFacts data include information on schools, services, staff, students, and educational outcomes.

UNITED STATES DEPARTMENT OF EDUCATION

OFFICE OF PLANNING, EVALUATION AND POLICY DEVELOPMENT

ASSISTANT SECRETARY

Mr. George A. Scott
Director, Education, Workforce,
 and Income Security Issues
Government Accountability Office
Washington, DC 20548

Dear Mr. Scott,

Thank you for the opportunity to preview the Government Accountability Office (GAO) draft report, *Status of the Department of Education's Inventory of Its Data Collections*, (GAO-13-596R). The U.S. Department of Education (the Department) appreciates GAO's efforts to review and understand the breadth of the Department's data collection efforts. We believe it is of critical importance that the Department compiles a comprehensive catalog of our data collections, uses it to identify further opportunities to reduce reporting burden where possible, and makes it publicly available in accordance with our principles of open and transparent government. We are especially pleased that our work to develop and publish a data inventory was found by GAO to be a "reasonable" process in a number of respects, and led the Office of Management and Budget (OMB) to label us "a leader in this practice."

I am proud of the Department's work to create its Data Strategy Team (DST); which is led jointly by my office, the Office of Planning, Evaluation, and Policy Development, along with the National Center for Education Statistics and the Privacy Information Records Management Service. The DST works across all Department offices to build cohesive processes and data policies as well as to improve transparency in our collection and release of data. Compiling the data inventory was something the DST determined that it needed to do to create a complete picture of the data that the Department collects. We are pleased to learn that in addition to our planned uses of it, GAO, OMB, and Congress are also interested in this tool.

We appreciate GAO's examination of this important issue and note that we have no comments or suggested edits to the draft report. I look forward to the release of the final report and its discussion of our important work in this area.

Sincerely,

Denise Forte
Acting Assistant Secretary

(131220)