

The ABCs of Data Dictionaries

This overview of the basic components of a data dictionary is designed to educate and inform IDEA Part C and Part B 619 state staff about the purpose and benefits of having up-to-date data dictionaries for their data systems.

What Is a Data Dictionary?

A data dictionary is a document containing descriptions of the data elements in a data system. Examples are the data element name, length of allowed characters for storing the data element, definition, and other technical information (e.g., how and where the data were collected, data format, how the data relate to other data). A data dictionary should be easily accessible to users and managers of a data system. Just as a typical dictionary helps convey the meaning of words, a data dictionary translates data elements into real-world terms. A data dictionary also helps users prepare data elements to be compared or merged with data from other systems.

Why Is a Data Dictionary Needed and How Can It Help?

The data dictionary is critical for ensuring the collection of high-quality data. It is a resource for comparing the meaning of data elements, reviewing reports, and transferring or merging data from other data systems. Without a data dictionary, interpreting data accurately is very difficult. Names, descriptions, and characteristics of data within the system are often abbreviated or undefined, requiring the user to know the system intimately to effectively use the data.

The data dictionary also helps with identifying irregular data and standardizing reported data. For example, a state Part B data manager has been asked to run a report or analyze data for Part B 619 using race. Districts commonly use multiple race variables with various codes to list race. Without a data dictionary, the Part B data manager may be faced with trying to align all the various codes, which can be a daunting task when some districts use 7 codes for race and others use more than 40 race codes. The data dictionary becomes a funnel of sorts, communicating to district/jurisdiction data specialists or managers how race data must be collected to ensure compliance. In general, a data dictionary promotes consistent collection of a data element across the state, helping ensure proper interpretation and use of the information.

A data dictionary is a particularly valuable reference for newly hired Part C/619 coordinators and data managers. It lessens the learning curve by helping them identify and understand the data items used for reporting, analysis, and data sharing requests. Data dictionaries provide a common language and understanding of the data elements in a system so program staff can communicate effectively with each other, with IT staff, and with other early childhood programs and stakeholders. Data dictionaries should be accessible in read-only format to internal and external users who contribute, manage, and/or use data to conduct their work.

Audiences for Data Dictionaries

- Part C/619 coordinators and data managers
- IT staff (who support Part C/619 program staff)
- Other state agencies receiving Part C/619 data
- Local district/jurisdiction Part C/619 program data system users
- Researchers

What Are the Essential Components of a Data Dictionary?

Although they vary in detail and complexity, data dictionaries should contain at least the following information about each data element:

- Name
- Definition/description
- Type, format, and length (e.g., numeric/alphanumeric; number of characters)
- Location/position (e.g., where in the data system the element resides)
- Option sets—values assigned to categorical data elements (e.g., F= female, M= male)

Example of Data Dictionary Entry

| Element ID* | Element* | Report Cycle | Field Type* | Format* | Field Length* | Starting Position | Ending Position | Description* | Edit | Edit # |
|-------------|----------|--------------|-------------|---------|---------------|-------------------|-----------------|----------------------------|--|--------|
| FTE032 | GENDER | 1, 3 | A | X | 1 | 63 | 63 | Indicates student's gender | GENDER must be one of the following valid values: 'F' Female 'M' Male | E215 |

Adapted with permission from Georgia Department of Education (FY2014 Data Collection)

*Designates a required value that must be entered for the data element.

How Can a Data Dictionary Be Maintained and Enhanced?

Effective creation and maintenance of a data dictionary involve state Part C/619 and IT staff working together to define information needs and the technical requirements to meet those needs efficiently. A data dictionary can serve many different purposes and audiences, so it is important that it be maintained through input and attention from both the program and the IT staffs. While the design and upkeep of the data dictionary should be a collaborative endeavor, only authorized individuals should have permission to create and edit its content to preserve the integrity of the entries.

Using a data dictionary effectively, particularly when transferring data to external users, will ensure high-quality data collection and accurate use of data, both internally and externally. Thus, Part C/619 program administrators must integrate use of the data dictionary into policies and procedures that cover the transfer and exchange of data. Program administrators need to ensure the data dictionary is up to date, accurate, easy to interpret, and usable for internal and external users when transferring data.

States can use the **Common Education Data Standards (CEDs)** to create data dictionaries that ensure consistent vocabulary, formatting, and processes across their data systems, thus improving their ability to exchange and compare data across agencies/programs. For more information on CEDs and how its use can improve the alignment of data systems, visit <https://ceds.ed.gov/>.

Part C/619 state agency personnel who have questions about data dictionaries or would like help in developing or revising a data dictionary should contact their DaSy state liaison or write us at dasycenter@sri.com.

About Us

The contents of this report were developed under a grant from the U.S. Department of Education, #H373Z120002. However, those contents do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the Federal Government. Project Officers, Meredith Miceli and Richelle Davis.

The DaSy Center is a national technical assistance center funded by the U.S. Department of Education, Office of Special Education Programs. The DaSy Center works with states to support IDEA early intervention and early childhood special education state programs in the development or enhancement of coordinated early childhood longitudinal data systems.

To learn more about the DaSy Center, go to <http://www.dasycenter.org/> or call 650-859-3881.

