Complete College America
File Specification
2014 Update

Updated February 20, 2014 for the 2014 Data Collection Managed by SHEEO

Flat File Format Specification

Background

The Complete College America data collection process requires everyone to upload data from a CSV file. Data must be subsequently approved by the state agency's keyholder through the CCA online interface which includes a "one step validation" process, online user data views, and one click approval.

States can create these files internally using the information in this specification or they can enter data into the MS Excel spreadsheet that will create a CSV file to be uploaded. For more details on using the MS Excel spreadsheet, please refer to the Data Collection User Guide and the spreadsheet itself.

Data can be uploaded more than once for a given institution or state agency if it is necessary to add more data elements or make corrections. Both the online interface and the MS Excel application will allow you to refresh the last version of data and the online interface will also allow you to see previous "archived" versions of data. Once data is approved, it can no longer be modified or updated from a CSV file batch upload. This year's online approval system, however, does allow users to undo an approval so that the data can be updated.

Variable Name Construction

Every element of data submitted is uniquely identified by a combination of the following dimensions:

1. Data year or cohort year (5 Characters)
2. Institution Type, Award Type, or Award Seeking Type (3 Characters)
3. Enrollment Type (3 Characters)
4. A specific measure (3 Characters)
5. A demographic group (3 Characters)

A set of mnemonic codes for each of the above dimensions is provided in "2014 CCA Variable Names.XLSX". Each element name is constructed by the concatenation (strung together) of five mnemonic codes in the sequence indicated above for a variable name that is 17 characters long.

CSV File Format

Two optional batch file formats will be allowed. The three-column format is strongly recommended and allows agencies to combine state agency data with multiple sets of institution data in a single file. The two-column format allows entry of any or all data for a single state agency or institution.

The default batch load file format will be a CSV file with three fields per row; the element name as a string followed by a numeric data value followed by the OrgID of the state agency or institution. For Example: [ElementName, 15, OrgID]. This method is preferred and this is the format the MS Excel workbook generates.

The two-column batch load file format will be a CSV file with two fields per row; the element name as a string followed by a numeric data value. For Example: [ElementName, 15].

With only one named element per row:
- Quality assurance is straight forward because each item of data is uniquely identified.
- With a little practice the concatenated mnemonics are 'human-readable.'
- The rows can be assembled in any order and missing data (rows) are ok.
- You can load all or any part of a state agency's or institution's data with a CSV file.
- You don't need to worry about the order of multiple "columns" or missing values.

Variable Name Construction
Context Metric 1 (partial spread sheet view)

CONTEXT METRIC 1: ENROLLMENT (Unduplicated Headcount)		
Relevent Comments:	[INSERT COMMENTS HERE]	
If cell size '10' or less, suppress the data and report 'DS'. If data is not available, report 'NP'. If value is '0', report '0'.		**2006-07** **Full-Time First-Time and Continuing Students**
Two-Year Institutions	**TOTAL Students**	1000
	Race: Hispanic	200
	Race: Black, non-Hispanic	200
	Race: White, non-Hispanic	400
	Race: Asian	100
	Race: Native Hawaiian or other Pacific Islander	20
	Race: American Indian/Alaska Native	20
	Race: Two or More Races	20
	Race: Unknown	20
	Race: Non-resident Alien	20
	Gender: Male	460
	Gender: Female	540
	Age: Directly from High School (17-19 years old)	800
	Age: Age 25 and over	100
	Age: Other (20-24 years old)	80
	Age: Unknown	20
	Income: Received Pell Grant	150
	Remedial: Took Remedial Courses	220

Example CSV File Formats for the above data with three fields per row, a variable name, a value and an OrgID separated by commas. The two-column version does not include the OrgID. Row sequence does not matter.

3-column format	**2-column format**
Y06072YRFTSHDCTOT, 1000 , 12345	Y06072YRFTSHDCTOT, 1000
Y06072YRFTSHDCRHS, 200 , 12345	Y06072YRFTSHDCRHS, 200
Y06072YRFTSHDCRBL, 200 , 12345	Y06072YRFTSHDCRBL, 200
Y06072YRFTSHDCRWH, 400 , 12345	Y06072YRFTSHDCRWH, 400
Y06072YRFTSHDCRAS, 100 , 12345	Y06072YRFTSHDCRAS, 100
Y06072YRFTSHDCRPI, 20 , 12345	Y06072YRFTSHDCRPI, 20
Y06072YRFTSHDCRAI, 20 , 12345	Y06072YRFTSHDCRAI, 20
Y06072YRFTSHDCRMR, 20 , 12345	Y06072YRFTSHDCRMR, 20
Y06072YRFTSHDCRUN, 20 , 12345	Y06072YRFTSHDCRUN, 20
Y06072YRFTSHDCRNR, 20 , 12345	Y06072YRFTSHDCRNR, 20
Y06072YRFTSHDCRGM, 460 , 12345	Y06072YRFTSHDCRGM, 460
Y06072YRFTSHDCRGF, 540 , 54321	Y06072YRFTSHDCRGF, 540
Y06072YRFTSHDCAHS, 800 , 54321	Y06072YRFTSHDCAHS, 800
Y06072YRFTSHDCAAD, 100 , 54321	Y06072YRFTSHDCAAD, 100
Y06072YRFTSHDCAOT, 80 , 54321	Y06072YRFTSHDCAOT, 80
Y06072YRFTSHDCAUN, 20 , 54321	Y06072YRFTSHDCAUN, 20
Y06072YRFTSHDCPGR, 150 , 54321	Y06072YRFTSHDCPGR, 150

A full table of variable names that can be sorted by the five dimensions of the mnemonics as well as a table mapping the mnemonics to individual metrics is provided in a separate Excel file, "2014 CCA Variable Names.XLSX". These tables will also be provided in the Excel Application.

Also note that the **OrgID** represents either a state agency or an institution. For institutions in most cases this is the six-digit IPEDS UnitID. For most state agencies the **OrgID** is the standard two character state code. If you have more than one state agency in your state, please contact customer service or your keyholder for your agency's **OrgID**.

File Management

- If you submit data using the two 2-column format then all data in a single file must be for the same State Agency or Institution. You must provide the Agency or Institution ID at the time the file is uploaded and it will apply to all rows in the file. If you would like to combine multiple State Agencies and/or Institutions in the same file you will need to use the 3-column format which includes the OrgID in the file. WE RECOMMEND THE STATE 3-COLUMN FORMAT
- A file can contain all variables or just some.
- Subsequent uploads for the same institution (or state agency) will replace data where the same element was loaded previously and will add data where new elements are loaded.
- Any functional Windows file name with an extension of .TXT or .CSV is acceptable. We suggest that you name your files and versions of files in a consistent manner for your own purposes.
- Data element names submitted with no data will not be uploaded. If you would like to make row with a value in it null, then you will need to submit the variable with 'NV' in the value field or perform a reset in the approval screen.
- Data element names submitted with zero ('0') as data will be treated as authentic data with the exception of data submitted for Outcome 4, '0' is not an acceptable value for Average Time or Credits to Degree.
- Errors and detailed history about data submitted are available on the online interface.