*The*

WTB

Data Warehouse

*Databases and Tables at the*

*Workforce Training and Education Coordinating Board*

*for evaluating*

*Workforce Development System Programs*

*Terje Atle Gjertsen, April 2017*

This page intentionally left blank

Table of Contents

[Main concepts 5](#_Toc515442593)

[Statistical business process and data by state 5](#_Toc515442594)

[Data by content 6](#_Toc515442595)

[Production processes 6](#_Toc515442596)

[General business rules 7](#_Toc515442597)

[Participant data 8](#_Toc515442598)

[Outcome data 11](#_Toc515442599)

[The request 11](#_Toc515442600)

[The delivery 11](#_Toc515442601)

[Outcome production files 13](#_Toc515442602)

[Statistical data and published data 14](#_Toc515442603)

[Participant record level 14](#_Toc515442604)

[Ratios 15](#_Toc515442605)

[Medians 17](#_Toc515442606)

[Naming conventions 18](#_Toc515442607)

[Key business processes 19](#_Toc515442608)

[Appendix 20](#_Toc515442609)

[Table 1 Processes (selected tables) 20](#_Toc515442610)

[Table 2 Tables by state (selected tables) 20](#_Toc515442611)

[Table 3 Tables by database (all tables) 21](#_Toc515442612)

[Process at SBCTC 26](#_Toc515442613)

[Joins and subsets 27](#_Toc515442614)

[Outcome Types 29](#_Toc515442615)

[Footnotes 30](#_Toc515442616)

[Measures 30](#_Toc515442617)

[Index 31](#_Toc515442618)

This page intentionally left blank

# Main concepts

We want to store all data used for evaluating workforce programs on the sql server. How best do we do this? Organizing data that is part of multiple statistical business processes forces us to visit with these processes. This document will therefore be about processes as much as about data structure.

We start by spelling out some key concepts relating to data and statistics. We consider (and name) data either by reference to its content or in reference to its state in the production process.

## Statistical business process and data by state

A statistical business process has the same major component as other processes:

1. Input
2. Production
3. Output

For statistics the main components of the production process are:

The sub processes

… concludes with data in one of several states

Raw data is data as received from data provider. This might include keeping the format (xlsx or txt or csv) but most importantly it means keeping the information in the data unaltered, i.e. no editing.

Production data covers a wide range of states for the data, from raw data simply altered to fit the requirement of the production system up to data fully controlled and edited.

Statistical data is data ready for tabulating or for making statistical inferences. Statistical data is published data at its most granular level.

## Data by content

What is meant by “program participant data” and “outcome data”? Throughout this document the terms are used to mean the following;

Program participant data is data listing participants in a given program. Elements in a table with participant data includes identifying fields (ssn, name), demographic info (gender and race) and other attributes specific to the individual like veteran status and disability. Program and program year (start and exit date) must be indicated as well as other info from program administrator like completer and skills attainment.

Outcome data holds information on employment and education of program participants. For either outcome, and for ease of analysis, there should be one record per individual per outcome quarter (In a raw state this will not always be true).

Static data will sometimes also be referred to as reference data and is the kind of data delegated to look-up tables. It is considered static here because it does not change between programs or program year. Examples are NAICS and CIP.

A fourth type of data, when considered in terms of content, does not fit any of the groups above. This is price indexes used for inflation adjustment. We will refer to this as external data.

## Production processes

In the production of WTR/Matrix we have the following processes.

1. Collect participant data: output = raw program participant data
	1. Request data from administrative entity
	2. Upon reception – check that fields and records are in line with previous years.
	3. Keep the raw data in folder[[1]](#footnote-1) designated to incoming data.
2. Clean participant data: output = program participant production data
	1. Read raw participant data into sas files
	2. Populate standardized core data elements
	3. Validate cohort, check for duplicates and select rows to include in outcome request
	4. Store in program participant database on sql server
3. Collect outcome data: output = raw outcome data
	1. Collate all available programs and request outcome match through SBCTC[[2]](#footnote-2) (Retain table with request on the sql server)
	2. Upon delivery store raw outcome data (as received) on sql server
4. Clean outcome data: output = outcome production data
	1. Read into sas (selecting one or more outcome period)
	2. Validate and standardize core data elements, duplicate check
	3. Retain as a temporary file[[3]](#footnote-3) in SAS
5. Assemble: output = statistical data
	1. Pull cleaned participant data (from sql)
	2. Pull cleaned outcome data (temporary file in step 4 above)
	3. Merge participant and outcome
	4. Store in pub/disseminate database on sql server
6. Disseminate: output = published data
	1. Pull statistical data (from sql)
	2. Run desired transformations, tabulations, and reports
	3. Communicate (Analysis, commentary, Graphics, Distribution)
7. Document: current work and change in processes

## General business rules

In setting up a production system and a data structure we need some general guidelines. The business rules below are such guidelines. Some of these are more relevant in a development phase while some are more useful in a production phase.

* Don’t do irreversible changes (like deleting or editing raw data) without securing copy of the original.
* Store data at its most granular level
* Know the subject of the table
	+ Is the record an object or an event
	+ What unifies the records (program/programyear/participant/exiter/completer)
	+ Is there a unique field for each record
* Know the data elements (columns/fields/variables) in the table
	+ Numeric (continuous or discrete) or categorical (nominal or ordinal). And remember – a categorical variable might very well be denoted by numbers.
	+ How to interpret missing values
* Avoid making tables where many records have many non-relevant fields (stacking tables with few fields in common)
* Avoid making tables at the end of long (time consuming) processes. It will be costly to run again.
* Keep the production modular – if you have to retrace upstream then you should not have to every time seek the headwaters. Store procedures when these require little or no input between different executions.
* Standardize when cost effective.
* All output should be easily reproduced
* If a variable is limited to values given in a list (limited value set) and this variable is used extensively in reports then missing (null as a value) should be avoided. The reason is that reports and graphics in SAS is easier to handle when there are no missing values. Race is an example of a finite set of values while dates or ssn is not. In a production file that has columns with finite value sets an unknown is given the value Z (not provided in data). This way one can distinguish between a true missing (not provided in data) and a tardy type missing (data wrangler missed it when reading the data).

# Participant data

Participant data describes participants in a given program. There are twenty programs on the Matrix of January 2016. Fourteen of these programs have outcome measures of which twelve are also found in the Workforce Training Result (WTR).

Participant data for the 14 programs can be found in eleven databases on the sql server.

|  |  |
| --- | --- |
| Database | Worksource Program |
| Apprenticeships | APP | ‘Apprenticeship’ |
| BEA | BEA | ‘Basic Education for Adults’ |
| DLOA | PPT | ‘Postsecondary Professional-Technical Education’ |
|  | WR | ‘Worker Retraining at Community and Technical Colleges’ |
| DSB | DBS | ‘Vocational Rehabilitation for the Blind’ |
| DVR  | DVR | ‘Division of Vocational Rehabilitation’ |
| PCS  | PCS | ‘Private Career Schools’ |
| SCT  | SCT | ‘Secondary Career and Technical Education’ |
| TAA  | TAA | ‘Trade Adjustment Assistance Program’ |
| TB | TB | ‘Training Benefits Program’ |
| WIASRD | WIA\_A | ‘Workforce Investment Act Title 1-B, Adult’ |
|  | WIA\_DW | ‘Workforce Investment Act Title 1-B, Dislocated Workers’ |
|  | WIA\_Y | ‘Workforce Investment Act Title 1-B, Youth’ |
| WPA  | WPA | ‘Wagner-Peyser Act’ |

Three programs (PPT and WR from the DLOA as well as SCT) arrives matched with outcome data that we use. The WIASRD also has outcomes but we have always sought our own matching.

For programs APP, BEA, DBS, DVR, PCS, SCT, TAA, TB, WIA\_A/DW/Y, and WPA program participant data is stored in their respective databases as production files:

* Having standard columns (production data) along with unaltered raw data. Raw data field names are prefixed with rd\_ or left as is when system of field names prevent conflicting field names with production file (WPA and WIA\_A/DW/Y).
* Having one table per program per programyear (Exception WIA\_A/DW/Y – all 3 in same table).
* Keeping all records from data source.
* Flagging records used for outcome match (WTR=’Y’)

For the two programs from the DLOA tables where we use pre matched outcomes we go straight to the statistical file and therefore store this on [WorkforceTrngResults]. More on what makes a statistical file in later chapter.

**Participant production file**

A production file with participant data will look like this:

* ssn = ‘Social Security Number’
* lastName = ‘Last name’
* firstName = ‘First name’
* mi = ‘Middle initial’
* dob = ‘Date of birth’
* address = ‘Address – participant’
* zip = ‘Zip code – participant’
* city = ‘City – participant ‘
* fips = ‘County Fips (last2dig) – participant ‘
* countyName = ‘County Name – participant ‘
* wda = ‘Workforce Development Council Area – participant’
* state = ‘State – participant’
* race = ‘Race’
* eth = ‘Hispanic, Y/N’
* raceEth = ‘Race/Ethnicity’
* gender = ‘Gender’
* disabled = ‘Disabled, Y/N’
* lrnDisablty = ‘Learning Disability, Y/N’
* english = ‘English Language Learner, Y/N’
* priorEd = ‘Prior Education’
* publicAssist = ‘Public Assistance, Y/N’
* econDisadv = ‘Economic Disadvantaged, Y/N’
* veteran = ‘Veteran, Y/N’
* secEdAtExit = ‘In secondary education (high school) at exit, Y/N’
* wsprgm = ‘Workforce System Program’
* startDate = ‘Program Start Date’
* exitDate = ‘Program Exit Date’
* schoolCode = ‘Code identifying training provider’
* programCode = ‘Code identifying instructional program’
* cip = ‘Classification of Instructional Program’
* wdaTrngPrvdr = ‘Workforce Development Council Area – training provider’
* completer = ‘Completion, Y/N’
* award = ‘Group of award types’
* wsprgmYear = ‘ProgramYear f(exitDate)’
* startYearQ = ‘YearQuarter of Start f(startDate) ‘
* endYearQ = ‘YearQuarter of Exit f(exitDate) ‘
* ageAtReg = ‘Age at Registration f(startDate and dob)’
* ageAtExit = ‘Age at Exit f(exitDate and dob)’
* months = ‘Months in training f(startDate and exitDate)’
* uniqueReq = ‘Is record unique across all columns Y/N’
* dupRecChoice = ‘Choice of record after uniqueRec check 1/0’
* missingSSN = ‘Is SSN missing Y/N’
* validSSN = ‘Is SSN valid Y/N’
* uniqueSSN = ‘Is SSN unique to program and programyear Y/N’
* dupSSNChoice = ‘Choice after uniqueSSN check Y/N’
* missingFLD = ‘Are Firstname or Lastname or Dob missing Y/N’
* validFLD = ‘Are Firstname or Lastname or Dob valid Y/N’
* uniqueFLD = ‘Is combination of non-missing FLD unique to prgm and prgmyear Y/N’
* WTR = ‘Record used for WTR Match Req Y/N’
* RowID = ‘Unique id on record, generated upon storage’
* Followed by columns of data as it arrived, aka raw data (for most programs given by prefiks rd\_ )

In this presentation we have separated the fields into segments of data similar in kind. The first three segments speak for themselves. In the fourth segment the attributes of the program can be found. In the fifth segment we have derived new fields, the argument of which is given in brackets - f(·), while the sixth segment has fields borne out of our production process. In the back the raw data is found with a name altered with the prefix rd\_.

# Outcome data

We consider two outcomes - employment and further education. The main source for employment data is in state unemployment insurance systems while a national registry of student records provides us with data on further education. We do not approach these sources ourselves but use SBCTC as a conduit.

## The request

In prior production cycles match requests were made per program. Starting with the 2013-2014 cohort (2015 outcomes) request are made for several programs combined, giving fewer tables to manage. The following discussion considers this new situation.

The match request process is simply selecting programs and records from participant tables. The amount of data in the request is limited to what SBCTC needs for their processes. The request consists of two files:

* a data file with ssn, name, dob, and start- and exit date
* a text document serving as a “packing slip” describing the request

Files are exchanged with SBCTC through the *SBCTC Upload Engine* (SUE). Staff at WTB all use the same username and password on SUE.

A table with all the data of the match request is retained in the database Outcomes on the sql server.

* [Outcomes].[dbo].[Request\_20160825]

## The delivery

The distinction between states of data (raw and production) is also used when dealing with outcome data. The need for data wrangling is considerably less with respect to columns and their content but a close attention to row selection is crucial.

The table with outcomes that comes back from SBCTC has multiple rows per employee (or student) but should have only one per individual per quarter. A person might well attend more than one school or have more than one employer in any given quarter but in our measurement of outcome we will consider only one of each. SBCTC will have addressed any duplicates on quarter (if more than one records per person per quarter). We need:

*-One record (outcome) per ssn per outcome-quarter*

So the key columns in an outcome table are ‘ssn’ and ‘outcome quarter’ which makes a join with participant data possible. In addition the outcome tables have other outcome attributes such as wage and earnings for employment outcome and college and CIP for educational outcome.

The complete delivery from SBCTC consists of 7 tables

1. Participant data: *Participant\_Training*
2. Employment outcome: *Participant\_Employment*
3. Educational outcome: *Participant\_Higher\_Education*
4. Adjustment factors: *Quarterly\_Factors*
5. Look up table with activity based classification of enterprise: *SIC\_NAICS*
6. Look up table with classification of instruction program: *Trade\_CIP\_Codes*
7. Look up table with educational institutions: *Training\_Providers*

We commit only 3 of these tables to the database Outcomes. The first table on participants is the same as the one we sent off, the two outcome tables are saved as is (raw data) with a name in accordance with the name of the corresponding match request, the adjustment factors is also secured on the database. Expanding on the example from the chapter on the request:

* [Outcomes].[dbo].[Request\_20160825]
* [Outcomes].[dbo].[Request\_20160825\_Qtrly\_Factors]
* [Outcomes].[dbo].[Request\_20160825\_Raw\_Empl]
* [Outcomes].[dbo].[Request\_20160825\_Raw\_HiEd]

The three look up tables are not committed to the same database. We have the two classification standards already in the database *Lookup* while training providers are not stored anywhere. There is limited information in the training\_provider table and none are used in the current production. Presumably there are few changes in the table’s content between match requests but a system for storing this table should nevertheless be put together.

There are several reasons to store the two outcome tables as raw data and not as production data. Raw outcome data tables are big (millions of rows) and doubling up the number of columns with standardized ones will take up a lot of space on the sql server. In addition there is the chance that the standard in ‘standardized columns’ might change (from one to two outcome quarters). Finally, bringing big amounts of data back and forth between sql and sas takes time. We do this one time with the raw data. After this we limit traffic back and forth the sql server to a given period (like third quarter past exit), apply a production file standard to a subset only, hold this table a temporary file (in SAS) until it is merged with a participant table.

## Outcome production files

Below is a description of production files for both outcomes. You wont find these files on the server as stand alone tables but through joins with participant tables the columns will appear in the statistical files.

**Employment production file**

* ssn = ‘Social Security Number’
* EmplQtr = ‘Employment Quarter’
* EmplFullTime = ‘Employed Full time, Y/N’
* hWage = ‘Hourly Wage (dollars current to EmplQtr)
* qEarngs = ‘Quarterly Earnings (dollars current to EmplQtr)
* yEarngs = ‘Annual Earnings Estimated (dollars current to EmplQtr)
* qHours = ‘Quarterly hours’
* Naics = ‘Naics’
* EmplSource = ‘Source of Employment data’
* hWageAdj = ‘Hourly Wage inflation adjusted’
* qEarngsAdj = ‘Quarterly Earnings inflation adjusted ‘
* yEarngsAdj = ‘Annual Earnings Estimated inflation adjusted ‘
* InflAdjPer = ‘Period of inflation adjustment’ \*
* Followed by columns of data as it arrived, aka raw data (prefiks rd\_ )

*\*A note on inflation and inflation adjustment. Employment outcome data in match requests made to SBCTC come back with current dollars and dollars adjusted to 2009. Employment outcome data in DLOA do not have current dollars but have dollars adjusted to 2009 (suffix 2009) and dollars adjusted to the latest first calendar quarter (given in documentation but not apparent from column name). Employment outcome data from ERDC has only one outcome quarter (latest first calendar quarter ) for all participants with dollars current to this period. More on the logic of inflation adjustment see the document ‘Real Wages’ and consult sas script.*

**Higher Education production file**

* ssn = ‘Social Security Number’
* HiEdQtr = ‘Education Outcome Quarter’
* HiEdCip = ‘Field of study’
* HiEdCollege = ‘College Code’
* HiEdSource = ‘Source of Education data’
* Followed by columns of data as it arrived, aka raw data (prefix rd\_ )

# Statistical data and published data

When all input data is cleaned and lined up, a series of joins will produce the file making the basis for disseminated data. This is published data in its most granular form and the result from the step in the statistical business process named “Assemble”.

## Participant record level

Statistical data, ***statfiles***, is defined in terms of a given output. WTR is different than outputs in the ETP process and statistical data will therefore be different.

Statfiles for the WTR are

* limited to records with program participants were match was attempted (i.e. where we had valid ssn)
* limited to columns that are part of standard program participant attributes (except the DLOA based tables were rd\_ columns were kept for ease in ad hoc analysis)
* limited to a number of outcome periods given by current state mandate (currently only third quarter)

Statfiles for the WTR are stored in its ***own database*** in [WorkforceTrainingResults]. Collating the DLOA based statfiles with statfiles coming from match requests (JntReq) gives us a statfile holding all programs stripped for all non standard columns. The latest one when writing this is: [WorkforceTrngResults].[dbo].[WTR1415\_PrtcpntRecords]

Published data are derived from statistical data, from plain counts to statistical measures like medians and averages. These data are found in tables and charts (sometimes embedded in text only) of the WTR. We do not need to store all our published data (making pubfiles), particularly when making tables and graphics with only one programyear. These presentations can be pulled directly from the statfile. However, when assembling multiple years worth of published data in time series it saves time and secures consistency having previous published numbers stored. So we make pubfiles for any figures stacked into time series and store also this on [WorkforceTrngResults].

Specifying the structure of these tables has been demanding. Variables that look similar are found to vary considerably after careful reading of footnotes. Two measurements with similar variable name will between programs come from different sub sets of the statistical file (e.g. completer vs non-completer or students excluded vs included).

On shape we considered both wide and narrow for the pubfiles.

A wide form table will have a column structure like this: *wsprgm;wsprgmyear;secEdAtExit;completer;hiEd;measure1value;measure2value;measure3value ….*

A long form table will have a column structure like this: *wsprgm;wsprgmyear;secEdAtExit;completer;hiEd;measure;value*

The first five are categorical variables needed for the main measures in the WTR.

On scope of the pubfiles we chose to limit the pubfiles by current time series used in the WTR but this can be expanded going forward.

The current system is based on dividing metrics used in time series in two separate tables:

* ratios
* medians

## Ratios

Ratios are based on counts – one for numerator and one for denominator. The pubfile with ratios prior to the 1314 cohort, *WTR1213\_Ratios*, was gathered from archived xls workbooks which did not include the counts behind the ratios. The workbooks also had more ratios than what appeared in publications and to simplify later retrieval all ratios used in past publications are flagged through the fields [wtrSeries] and [matrix].

These are the fields in the table ‘*WTRYYYY\_Ratios*’:

[wsprgm], [wsprgmYear], [secEdAtExit], [completer], [Empl\_EmplFT\_HiEd], [numerator], [denominator], [measure], [ratio], [wtrSeries], [matrix]

The table holds only aggregated data and hence the use requires careful selection of rows (another reason for flagging of past use).

The first two fields specify the cohort. The next two fields [secEdAtExit] and [completer] are based on a binary (Y=yes, N=no) fields where an A (=all) suggest the binary has not been used.

Every value in [Empl\_EmplFT\_HiEd] has one corresponding value in [measure]. The reason both are included is that while the first field gives us the sets[[4]](#footnote-4) used in finding [numerator] and [denominator] the second field gives us a more legible measure[[5]](#footnote-5).

Given the above and the 10 ratios (measures) calculated for every combination of wsprgm, wsprgmYear, secEdAtExit, and completer the total number of rows in the table is found by the following:

* The programs BEA, TAA, TB, WIA-A, WIA-DW and WPA are all mute on *secEdAtExit* and *completer* which along with 10 measures give us a total of 60 ratios (6programs·1secEdAtExit·1completer·10measure=60).
* PCS is mute on *secEdAtExit* but has different values in the *completer*. The 3 values A, Y and N gives us a total of 30 ratios for this program (1programs·1secEdAtExit·3completer·10measure=30).
* WIA-Y is mute on *completer* but has different values of the variable *secEdAtExit*. The 3 values A, Y and N gives us a total of 30 ratios for this program (1programs·3secEdAtExit·1completer·10measure=30).
* The remaining programs APP, DSB, DVR, PPT, SCT, and WR are mute on *secEdAtExit* but have the 3 values A, Y and N on *completer* giving us 180 ratios (6programs·1secEdAtExit·3completer·10measure=180).

This gives us 14 programs and 300 ratios for one year worth of exiters. For the matrix we use 14 of these ratios. For the remaining WTR products we use upwards of 36 ratios.

In the table *WTR1415\_Ratios* the program DVR becomes unmuted on *secEdAtExit* where not all subsets have full time employees and the tally is changed to 355.

A note on [wtrSeries], [matrix]. It is not obvious that one should keep these two fields beyond the table with the old series. Still choosing to do this is reasoned by a desire to keep the tables consistent over time. When a requested range for time series exclude the older cohorts (1213 and older) then we will have much more flexibility in choosing ratios to display.

## Medians

The structure for medians is identical to ‘ratios’ with four exceptions.

* There is an additional categorical in [HiEd]
* Instead for [numerator], [denominator], [ratio] we have [count], [median]
* Medians past the 2012-2013 cohort also has the flagging of use in [wtrSeries] and [matrix][[6]](#footnote-6)
* We need to indicate inflation adjustment period

The general table structure for ‘medians’ is like this:

[wsprgm], [wsprgmYear], [secEdAtExit], [completer], [HiEd], [measure], [median], [count], [InflAdjPer], [wtrSeries], [matrix]

There are three variables for which medians are measured. These are annual earnings, hourly wage, and quarterly hours.

One program (WIA-Y) has three values in *secEdaAtExit* and mute on *completer* giving us 27 medians (1program · 3secEdaAtExit · 1completer · 3Hied · 3measure=27).

Seven programs (APP, DSB, DVR, PCS, PPT, SCT,WR) are mute in *secEdaAtExit* but not in *completer* giving us 189 medians (7programs·1secEdaAtExit·3completer·3hied·3measure=189).

Six programs (BEA, TAA, TB,WIA-A, WIA-DW, WPA) are mute in both *secEdaAtExit* and *completer* giving us 54 medians (6programs·1secEdaAtExit·1completer·3hied·3measure=54)

This gives us 270 medians for one year worth of exiters.

An important note on [wtrSeries] is that in the past medians used earnings and wages for all where used regardless of student status. This is not correct according to State Core Measures and the earnings for non–students are used starting with the 1314 cohort. For most programs this change is of minor magnitude and within one percentage point.

# Naming conventions

The name of a table carries a lot of information. The components indicate worksource program, programyear, state in the production process, and content.

The codes used for ***program*** can be found in the chapter *Participant data*.

Values used for ***worksource program year*** will either include both years (2013-2014 or 1314) or only starting year (PY2013 for PY13, in line with DOLETA[[7]](#footnote-7)). Note that the annotation using first year differs from IPEDS, State Fiscal Year, and Federal Fiscal Year where the latter year is used.

For ***state in production*** we have four values. Assign name by the highest (by value added) level if there is data in more than one state of the production. The ***numeral*** was inserted to make tables sort in the same direction as the overall process. The numeral idea was thought of before making the decision to locate all statistical files on its own database – WorkforceTrngResults. The result is that the only database where this desired sorting function comes to play is SCT where we receive participant and outcome from same source - ERDC.

* 1Raw – raw data
* 2Prod – data in production
* 3Stat – statistical data
* 4Pub – published data

Locating all statistical files in WorkforceTrngResults furthermore removes the need for the numeral in this database since this database holds only the end product of the overall process.

For ***content*** we have three values. If a table has both participant data and outcome then name it as participant.

* Partcpnt – program participant
* Empl – Employment outcome
* HiEd – Continued/Higher education outcome

Note that for statistical and published data content (and the product) is given. In our case the end product is the workforce training results and matrix. The database WorkforceTrngResults holds multiple statistical files for a given cohort as well as a statistical and publication tables with WTR in the name. The reason for this is that there is a different process behind the stat-files. We have three different pipe lines:

* Statfiles coming from the joint requests and
* Statfiles coming from the DLOA tables and
* Statfiles coming from ERDC (SCT) and
* Statfiles coming from the occasional single program match request (WPA1415)

For a standard database with program participant data see the appendix and *Table 3 Tables by database (all tables)* and database BEA. For the odd case, where the numeral actually servers a function, see the database SCT.

# Key business processes

# Appendix

## Table 1 Processes (selected tables)

|  |  |  |
| --- | --- | --- |
| **Process** | **Input** | **Output** |
|  |  |  |
| *Collect program participant data* | *Description of cohort and data structure (To program administrator)* | *Raw data, as is.* |
|  |  |  |
| *Clean program participant data* | *Raw data, as is.* | *BEA1314\_2Prod\_Partcpnt* |
|  |  |  |
| *Collect outcome data* | *BEA1314\_2Prod\_Partcpnt\_Request**As part of**Request\_20151121* | *Request\_20151121\_Raw\_Empl* |
|  |  |  |
| *Clean outcome data* | *Request\_20151121\_Raw\_Empl* | *temporary-file* |
|  |  |  |
| *Assemble* | *Outputs above* | *Statfile, Measures, and MeasuresSeries* |
|  |  |  |
| *Disseminate* | *Statfile* | *Various reports* |
|  |  |  |
| *Document* |  | *This document* |

## Table 2 Tables by state (selected tables)

|  |  |
| --- | --- |
| **Raw data** |  |
|  |  |
| *tables with only raw data* | *is usually not on the sql-server but combined with prod-data* |
| *SCT1314\_1Raw\_Empl* | *on the sql-server because needed transformation b4 join with participant table* |
|  |  |
| **Production data** |  |
|  |  |
| *TAA1314\_2Prod\_Partcpnt* |  |
| *SCT1314\_2Prod\_Empl* |  |
|  |  |
| **Statistical data** |  |
|  |  |
| *JntReq1314\_Stat* | Statfile with all participants in program part of joint request in summer 15 |
| *SCT1314\_3Stat* | Statfile for SCT that includes all fields from production – on SCT base |
| *SCT1314\_Stat* | Statfile for SCT w only WTR fields – on WTR base |
| *PPT1314\_Stat* | Statfile for one of DLOA programs, includes all fields from program participant file– on WTR base |
| *WTR1415\_PrtcpntRecords* | *Statfile with all programs - 1415 cohort* |
|  |  |
| **Published data** |  |
|  |  |
| *WTR1213\_Medians* | *Medians used in published data - program years 0708 through 1213* |
| *WTR1213\_Ratios* | *Ratios used in published data - program years 0708 through 1213* |
| *WTR1415\_Medians* | *Medians used in published data - program year 1415* |
| *WTR1415\_Ratios* | *Ratios used in published - program year 1415* |
|  |  |
| **Lookup** |  |
|  |  |
| naics2012 |  |
| county |  |
| wda |  |
|  |  |

## Table 3 Tables by database (all tables)

|  |  |  |  |
| --- | --- | --- | --- |
| **Apprenticeships** | **Content** | **State** |  |
|  |  |  |  |
| *APP\_0910* |  |  | *Joined with outcomes, not by new standard* |
| *APP\_1011* |  |  | *Joined with outcomes, not by new standard* |
| *APP\_1112* |  |  | *Joined with outcomes, not by new standard* |
| *APP\_1213* |  |  | *Joined with outcomes, not by new standard* |
|  |  |  |  |
| *APP1314\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *APP1314\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *APP1314\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
| ***….*** |  |  |  |
| *APP1516\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *APP1516\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *APP1516\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
| *programs* |  |  |  |
| *providers*  |  |  |  |
|  |  |  |  |
| **BEA** | **Content** | **State** | Basic Education for Adults |
|  |  |  |  |
| *BEA0809\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *BEA0809\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *BEA0809\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
| ***….*** |  |  |  |
| *BEA1516\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *BEA1516\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *BEA1516\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
| **CareerBridge** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **Census** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **CTC** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **DataLoad** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **DataLoadTest** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **DLOA** |  |  | Data Linking Outcome Assessment |
|  |  |  |  |
| *PPT0910\_Stat* | *Partcpnt* | *Prod/Stat* | *WTR stat file retrieved and conformed to standard w outcomes in 2015Q1 dollars* |
| *PPT0910\_Stat\_Labels* | *-* | *-* | *Metafile* |
| *PPT0910 \_Stat \_Request\_IWPR* | *-* | *-* | *Process file* |
|  |  |  |  |
| *JP1011-JP1314* | *Partcpnt* | *-* | *Pre matched - PPT (Post Sec Prof TecH)* |
| *WR1011-WR1314* | *Partcpnt* | *-* | *Pre matched - WR (Worker Retraining)* |
| *XFR1011-XFR1314* | *Partcpnt* | *-* | *Pre matched – XFR (Transfer)* |
|  |  |  |  |
| *CIP-SOC* |  |  |  |
| *RaceCodes* |  |  |  |
| *SBCTCCollegeCodes* |  |  |  |
|  |  |  |  |
| **DSB** | **Content** | **State** | Department of Services for the Blind |
|  |  |  |  |
| *DSB0809\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *DSB0809\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *DSB0809\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
| ***….*** |  |  |  |
| *DSB1617\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *DSB1617\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *DSB1617\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
| **DVR**  | **Content** | **State** | Division of Vocational Rehabilitation |
|  |  |  |  |
| *DVR0809\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *DVR0809\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *DVR0809\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
| **….** |  |  |  |
| *DVR1617\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *DVR1617\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *DVR1617\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
| **ETP** |  |  | Eligible Training Provider |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **IPEDS** |  |  | Integrated Postsecondary Education Data System |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **Lookup** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **OFM** |  |  | Office of Financial Management |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **Outcomes** | **Content** | **State** |  |
|  |  |  |  |
| *Request\_20151121* | *-* | *-* | *Process file* |
| *Request\_20151121\_Qtrly\_Factors* | *-* | *-* | *As arrived from SBCT* |
| *Request\_20151121\_Raw\_Empl* | *Empl* | *Raw* | *Check for dups & derive outcome fields b4 join* |
| *Request\_20151121\_Raw\_HiEd* | *HiEd* | *Raw* | *Check for dups & derive outcome fields b4 join* |
|  |  |  |  |
| *Request\_20160825* | *…* | *…* | *…* |
| *Request\_20161128* | *…* | *…* | *…* |
|  |  |  |  |
| **PCHEES** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **PCS**  | **Content** | **State** | Private Career Schools |
|  |  |  |  |
| *PCS1314\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Not identical to initial request – off by up to 87 records* |
| *PCS1314\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *PCS1314\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
| **….** |  |  |  |
| *PCS1516\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *PCS1516\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *PCS1516\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
| **PIRL** | **Content** | **State** | Participant Individual Record Layout |
|  |  |  |  |
| *PIRL\_2016\_Q1* |  |  |  |
| *PIRL\_2016\_Q1\_Labels* |  |  |  |
| **….** |  |  |  |
| *PIRL\_2017\_Q2* |  |  |  |
| *PIRL\_2017\_Q2\_Labels* |  |  |  |
|  |  |  |  |
| **PUMS\_DW** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **ReportServer** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **ReportServerTempDB** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **Research** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **ResearchTest** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **SCT**  | **Content** | **State** | Secondary CTE |
|  |  |  |  |
| *SCT1314\_1Raw\_Empl* | *Empl* | *Raw* | *As arrived, w 1+ outcome/ person/quarter* |
| *SCT1314\_2Prod\_Empl* | *Empl*  | *Prod* | *Transformed: 1 outcome/ person/quarter (ready for join with participant) and with standard outcome fields derived.* |
| *SCT1314\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *No SSN but ResearchID from ERDC* |
| *SCT1314\_2Prod\_Partcpnt\_Lables* | *-* | *-* | *Metafile* |
| *SCT1314\_3Stat* |  | *Stat* | *This file has all participants in the cohort, also the ones where ERDC could not find an ssn (missingssn=Y).* |
| *SCT1314\_3Stat\_Labels* | *-* | *-* | *Metafile* |
|  |  |  |  |
| *SCT1415\_1Raw\_Empl* | *Empl* | *Raw* | *Same as prior year* |
| *SCT1415\_2Prod\_Empl* | *Empl*  | *Prod* | *Same as prior year*  |
| *SCT1415\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Same as prior year* |
| *SCT1415\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *SCT1415\_3Stat* |  | *Stat* | *Same as prior year* |
| *SCT1415\_3Stat\_Labels* | *-* | *-* | *Metafile* |
|  |  |  |  |
| *SCT1516\_\** |  |  | *Same as 1415 ( kept school district data on statfile)* |
|  |  |  |  |
| **SSISDB** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **TAA**  | **Content** | **State** | Trade Asjustment Assistance |
|  |  |  |  |
| *TAA1314\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *TAA1314\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *TAA1314\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
| **….** |  |  |  |
| *TAA1516\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *TAA1516\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *TAA1516\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
| *TAPR\_FY2015\_Q3* | *Partcpnt* | *Raw* | *Trade Adjustment Assistance (Fiscal) Year 2015* |
| *TAPR\_ FY2015\_Q3\_Labels* | *-* | *-* | *Metatfile* |
|  |  |  |  |
| *TAPR\_FY2016\_Q4* | *Partcpnt* | *Raw* | *Trade Adjustment Assistance (Fiscal) Year 2016* |
| *TAPR\_ FY2016\_Q4\_Labels* | *-* | *-* | *Metatfile* |
| **….** |  |  |  |
| *TAPR\_FY2017\_Q2* | *Partcpnt* | *Raw* | *Trade Adjustment Assistance (Fiscal) Year 2017* |
| *TAPR\_ FY2017\_Q2\_Labels* | *-* | *-* | *Metatfile* |
|  |  |  |  |
| **TB** | **Content** | **State** | Training Benefits |
|  |  |  |  |
| *TB1314\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *TB1314\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *TB1314\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
| **….** |  |  |  |
| *TB1516\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *TB1516\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *TB1516\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
| **Test** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **WBDRS\_DW** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **WBDRS\_ERDC** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **wbdrs\_historical** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **Wendy\_Files** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **WIASRD** | **Content** | **State** |  |
|  |  |  |  |
| *WIASRD\_2013\_Q1*  |  |  | *Pre matched – For the 3 WIA programs (Title I)* |
| ***…*** |  |  |  |
| *WIASRD\_2015\_Q3* |  |  | *Pre matched – For the 3 WIA programs (Title I)* |
|  |  |  |  |
| *WIA0910\_Stat* | *Partcpnt* | *Prod/Stat* | *WTR stat file retrieved and conformed to standard w outcomes in 2015Q1 dollars* |
| *WIA0910\_Stat\_Labels* | *-* | *-* | *Metafile* |
| *WIA0910 \_Stat \_Request\_IWPR* | *-* | *-* | *Process file* |
|  |  |  |  |
| *WIA1314\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *From WIASRD\_2014\_Q4. Select WTR=’Y’ and drop Rd\_ , A0, B\_, C0, and D0 fields 4 particpnt prodfile.* |
| *WIA1314\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *WIA1314\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
| ***…*** |  |  |  |
| *WIA1516\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *From WIASRD\_2015\_Q4. Else same as year b4* |
| *WIA1516\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *WIA1516\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
| **WorkFirst** | **Content** | **State** |  |
|  |  |  |  |
| *CommunityServiceOffice* |  |  | *Name of CSO and croswalk to FIPS* |
|  |  |  |  |
| *WF1213\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *WF1213\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *WF1213\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
| **….** |  |  |  |
| *WF1516\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *WF1516\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *WF1516\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
| **WorkforceBoard\_Final** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **WorkforceBoard\_Staging** |  |  |  |
|  |  |  |  |
| ***…*** |  |  | *Database not described in this document* |
|  |  |  |  |
| **WorkforceTrngResults** | **Content** | **State** |  |
|  |  |  |  |
| *JntReq1314\_Stat* | *Partcpnt* | *Stat* | *Statfile wit all prgms part of Request\_20151121* |
| *JntReq1314\_Stat\_Labels* | *-* | *-* | *Metafile* |
|  |  |  |  |
| *JntReq1415\_Stat* | *Partcpnt* | *Stat* | *Statfile wit all prgms part of Request\_20160825* |
| *JntReq1415\_Stat\_Labels* | *-* | *-* | *Metafile* |
|  |  |  |  |
| *JntReq1516\_Stat* | *Partcpnt* | *Stat* | *Statfile wit all prgms part of Request\_20170922**[[8]](#footnote-8)* |
| *JntReq1516\_Stat\_Labels* | *-* | *-* | *Metafile* |
|  |  |  |  |
| *PPT1314\_Stat* | *Partcpnt*  | *Prod/Stat* | *Drop Rd\_ fields to limit to statfile* |
| *PPT1314\_Stat\_Labels* | *-* | *-* | *Metafile* |
| *PPT1415\_Stat* | *Partcpnt*  | *Prod/Stat* | *Drop Rd\_ fields and set WTR=’Y’ to limit to statfile* |
| *PPT1415\_Stat\_Labels* | *-* | *-* | *Metafile* |
| *PPT1516\_Stat* | *Partcpnt*  | *Prod/Stat* | *Drop Rd\_ fields and set WTR=’Y’ to limit to statfile* |
| *PPT1516\_Stat\_Labels* | *-* | *-* | *Metafile* |
|  |  |  |  |
| *SCT1314\_Stat* | *Partcpnt* | *Stat* | *Statfile with only WTR fields - 1314* |
| *SCT1314\_Stat\_Labels* | *-* | *-* | *Metafile* |
| *SCT1415\_Stat* | *Partcpnt* | *Stat* | *Statfile with only WTR fields - 1415* |
| *SCT1415\_Stat\_Labels* | *-* | *-* | *Metafile* |
| *SCT1516\_Stat* | *Partcpnt* | *Stat* | *Statfile with only WTR fields - 1516 (pluss schooldistrict dt)*  |
| *SCT1516\_Stat\_Labels* | *-* | *-* | *Metafile* |
|  |  |  |  |
| *WPA1415\_Stat* | *Partcpnt* | *Stat* | *Statfile from Request\_20161128. Drop F01-F91 and all skies\_ b4 stack w other statfiles* |
| *WPA1415\_Stat\_Labels* | *-* | *-* | *Metafile* |
|  |  |  |  |
| *WR1314\_Stat* | *Partcpnt* | *Prod/Stat* | *Drop Rd\_ fields to limit to statfile* |
| *WR1314\_Stat\_Labels* | *-* | *-* | *Metafile* |
| *WR1415\_Stat* | *Partcpnt* | *Prod/Stat* | *Drop Rd\_ fields and set WTR=’Y’ to limit to statfile* |
| *WR1415\_Stat\_Labels* | *-* | *-* | *Metafile* |
| *WR1516\_Stat* | *Partcpnt* | *Prod/Stat* | *Drop Rd\_ fields and set WTR=’Y’ to limit to statfile* |
| *WR1516\_Stat\_Labels* | *-* | *-* | *Metafile* |
|  |  |  |  |
| *WTR1213\_Medians* | *-* | *Pub* | *Medians 4 time series - prgmyears 0708 through 1213*  |
| *WTR1213\_Ratios* | *-* | *Pub* | *Ratios 4 time series - prgmyears 0708 through 1213* |
|  |  |  |  |
| *WTR1314\_Medians* | *-* | *Pub* | *Medians 4 time series - prgmyear 1314* |
| *WTR1314\_PrtcpntRecords* | *Partcpnt* | *Stat* | *Statfile with all programs - prgmyear 1314* |
| *WTR1314\_Ratios* | *-* | *Pub* | *Ratios 4 time series - prgmyear 1314* |
|  |  |  |  |
| *WTR1415\_Medians* | *-* | *Pub* | *Medians - prgmyear 1415* |
| *WTR1415\_PrtcpntRecords* | *Partcpnt* | *Stat* | *Statfile All programs - prgmyear 1415* |
| *WTR1415\_Ratios* | *-* | *Pub* | *Ratios - prgmyear 1415* |
|  |  |  |  |
| *WTR1516\_Medians* | *-* | *Pub* | *Medians - prgmyear 1516 (see footnote on WF)* |
| *WTR1516\_PrtcpntRecords* | *Partcpnt* | *Stat* | *Statfile All Programs - prgmyear 1516* |
| *WTR1516\_Ratios* | *-* | *Pub* | *Ratios prgmyear 1516 (see footnote on WF)* |
|  |  |  |  |
| *z\_* |  |  | *Tables to be deleted* |
|  |  |  |  |
| **WPA** | **Content** | **State** | Wagner Peyser Act |
|  |  |  |  |
| *LERS\_PY2015\_Q4* | *Partcpnt* | *Raw* | *From ESD. As given to DOL/ETA* |
| *LERS\_PY2015\_Q4\_Labels* | *-* | *-* |  |
|  |  |  |  |
| *WPA1314\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *WPA1314\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *WPA1314\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
| *WPA1415\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop F01-F91 and Skies\_ fields for particpnt prodfile (Based on LERS PY\_2015\_Q3)* |
| *WPA1415\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *WPA1415\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
| *WPA1415\_Stat* | *Partcpnt* | *Stat* | *Had to run match for this program separately* |
|  |  |  |  |
| *WPA1516\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Based on LERS PY\_2015\_Q4 (same RowId) Grab all rows* |
| *WPA1516\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *WPA1516\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
|  |  |  |  |
| **WTECB\_PROD** | **Content** | **State** |  |
|  |  |  |  |
| ***…*** |  |  | *Database- not described in this document* |
|  |  |  |  |

## Process at SBCTC

nnn

>>> Monica Peper <mpeper@sbctc.edu> 9/8/2017 1:02 PM >>>

Terje, here are our data match sources.

* In-State employment data. Source: WA State Employment Security Department (ESD). A cohort is submitted at the time of request and could take anywhere from a couple of days to a week to get results back. Results are not retained for future matches.
* Out-of-State employment data. Source: Oregon and Idaho Employment Departments. A large data set is submitted for a match once a year. SBCTC retains these matched records in our system and matches to those records for future employment match requests during the year.
* Federal employment data. Source: FEDES (OPM/DOD). A large data set is submitted for a match once a year. SBCTC retains these matched records in our system and matches to those records for future employment match requests during the year if authorized to include it.
* CTC Higher Ed data (public community and technical college data). Source: SBCTC enrollment records. A cohort is matched at the time of request. Results are not retained for future matches.
* Other Higher Ed data (University and other public or private college data). Source: National Student Clearinghouse (NSC).
	1. Internal: A large CTC data set is submitted for a match once a year. SBCTC retains these matched records in our system and matches to those records for future internal higher ed match requests.
	2. External (WTECB): A cohort is submitted at the time of request and could take anywhere from a couple of days to a couple of weeks to get results back. Results are not retained for future matches.

Thanks,Monica

## Joins and subsets

Joins determines some of the key subsets used in our tables. Information on these subsets are often best delegated to footnotes. If data arrived all cleaned up with no duplicates and with a standardized file structure our job would be reduced to joins and tabulating. The most important joins are illustrated below. From outcome files (employment and education) we select records based on outcome quarter. With the help of ssn and outcome-quarter the two selections of outcomes are combined horizontally with the program participant table.

SSN

OutcomeQtr

…Employment Outcome attributes

SSN

OutcomeQtr

…Program participant attributes

SSN

OutcomeQtr

…Educationl Outcome attributes

The static data can either be horizontally combined and stored along with the statistical data or a static file can be called upon when needed. This latter alternative is more in line with a relational database structure and allows more efficient management of text (is it “Native American” or is it “Indian” – easier to change in one lookup and keep a code in all other tables).

SSN

…Program participant attributes

lookup

code and decode

lookup

code and decode

lookup

code and decode

All records in the match request are found in the statistical file. There are many ways to subset this cohort but 6 of them sums up the result of the joins.

Empl = N

Empl = Y

Student = Y

Student = N

FT = N

FT = Y

With the order Empl, EmplFullTime, Student we can code the subsets like this:

*… starting on the top and going around with the hands of the clock*

* NNY Student with no employment
* YNY Part time employment and studies
* YNN Part time employment and no studies
* YYN Full time employment and no studies
* NNN No Employment and Not in School

*… and finally the triangle in the middle*

* YYY Full time employment and studies

Exhaustive sets of subtotals (where A = Y or N)

* YAA Employed
* NAA Not Employed
* YYA Employed Full time
* YNA Employed Part time
* NAA Not Employed
* AAY Student
* AAN Not Student
* NNN No Employment and Not in School
* YAY Employed or in School

## Outcome Types

There are multiple ways in which to describe outcomes:

* Emp – Y/N
* HiEd – Y/N
* OutcomeType
	+ '1' = 'Earning Wages' *YAA*
	+ '2' = 'No Earnings, Student' *NNY*
	+ '3' = 'No Earnings, Not Student' *NNN*
* OutcomeType2
	+ '10' = 'Earning Wages, Not Student' *YAN*
	+ '11' = 'Earning Wages, Student' YAY
	+ '01' = 'No Earnings, Student' *NNY*
	+ '00' = 'No Earnings, Not Student' *NNN*
* OutcomeType3
	+ '101' = 'Full time employed only' *YYN*
	+ '100' = 'Part time employed only' *YNN*
	+ '111' = 'Full time employed and Student' *YYY*
	+ '110' = 'Part time employed and Student' *YNY*
	+ '01' = 'Student only' *NNY*
	+ '00' = ' No Earnings, Not Student ' *NNN*
* OutcomeType4
	+ ‘YYA‘ = ‘Full time employed’ YYA
	+ ‘YNA’ = ‘Part time employed’ YNA
	+ ‘NNA’ = ‘Not employed’ NNA

## Footnotes

When presenting metrics from these data sets there are some essential information that should go into either

* a footnote on the graphics
* a footnote on the page
* a methods appendix (or an appendix with definitions)

Essential info

* Participant data: Source and cohort definition
* Outcome data: Coverage and Source
	+ WA – ESD
	+ WA, OR and ID – ESD
	+ WA, OR, ID – ESD & all other states through WRIS
	+ Federal employees (Federal Employment Data Exchange System, or FEDES)
* Measuring quarter
* Inflation adjustment
* Definition of
	+ Exiter
	+ Cohort construction
	+ Rate of Employment
	+ Rate of Full Time Employment
	+ Record selection in median calculations
	+ Annualized earnings

## Measures

1. 'Empl' = 'Employment'
2. 'Empl\_ExclUnemplStud' = 'Employment (2)'
3. 'Empl\_ExclAllStud' = 'Employment (3)'
4. 'FT\_Empl' = 'Full-Time Employed'
5. 'FT\_Empl\_ExclUnemplStud' = 'Full-Time Employed (2)'
6. 'FT\_Empl\_ExclAllStud' = 'Full-Time Employed (3)'
7. 'FT\_Empl\_OfE' = 'Employees In Full-Time Job'

*~~'FT\_Empl\_OfE\_ExclUnemplStud' = 'Employees In Full-Time Job (2)'~~ = 'FT\_Empl\_OfE'*

1. 'FT\_Empl\_OfE\_ExclAllStud' = 'Employees In Full-Time Job (3)'
2. 'EmplOrStud' = 'Employed Or In Further Education'
3. 'Stud' = 'In Further Education'
4. 'yEarngsAdj' = 'Median Annualized Earnings'
5. 'hWageAdj' = 'Median Hourly Wage'
6. 'qhours' = 'Median Quarterly Hours'

\* (2) Excluding Students With No Employment (3) Excluding All Students ;

# Index

Inflation adjustment 13

Outcome data 6

Production data 5

Production file - Employment 13

Production file - Higher Education 13

Production file - Participant 9

Program participant data 6

Program year 17

Raw data 5

Static data 6

Statistical data 5

SUE 11

Last page left blank

1. For data collected in 2016 this was X:\16 RawProgramParticipantData. For data collected in 2017 see L:\Data\_Sluice [↑](#footnote-ref-1)
2. State Board of Community and Technical Colleges [↑](#footnote-ref-2)
3. For the rationale see chapter on outcome data. [↑](#footnote-ref-3)
4. See chapter

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| *LERS\_PY2015\_Q4* | *Partcpnt* | *Raw* | *From ESD. As given to DOL/ETA* |
| *LERS\_PY2015\_Q4\_Labels* | *-* | *-* |  |
|  |  |  |  |
| *WPA1314\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop Rd\_ fields 4 particpnt prodfile* |
| *WPA1314\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *WPA1314\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
| *WPA1415\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Select WTR=’Y’ and drop F01-F91 and Skies\_ fields for particpnt prodfile (Based on LERS PY\_2015\_Q3)* |
| *WPA1415\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *WPA1415\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
| *WPA1415\_Stat* | *Partcpnt* | *Stat* | *Had to run match for this program separately* |
|  |  |  |  |
| *WPA1516\_2Prod\_Partcpnt* | *Partcpnt* | *Raw/Prod* | *Based on LERS PY\_2015\_Q4 (same RowId) Grab all rows* |
| *WPA1516\_2Prod\_Partcpnt\_Labels* | *-* | *-* | *Metafile* |
| *WPA1516\_2Prod\_Partcpnt\_Request* | *-* | *-* | *Process file* |
|  |  |  |  |
|  |  |  |  |
| **WTECB\_PROD** | **Content** | **State** |  |
|  |  |  |  |
| ***…*** |  |  | *Database- not described in this document* |
|  |  |  |  |

## Process at SBCTC

nnn

>>> Monica Peper <mpeper@sbctc.edu> 9/8/2017 1:02 PM >>>

Terje, here are our data match sources.

	* In-State employment data. Source: WA State Employment Security Department (ESD). A cohort is submitted at the time of request and could take anywhere from a couple of days to a week to get results back. Results are not retained for future matches.
	* Out-of-State employment data. Source: Oregon and Idaho Employment Departments. A large data set is submitted for a match once a year. SBCTC retains these matched records in our system and matches to those records for future employment match requests during the year.
	* Federal employment data. Source: FEDES (OPM/DOD). A large data set is submitted for a match once a year. SBCTC retains these matched records in our system and matches to those records for future employment match requests during the year if authorized to include it.
	* CTC Higher Ed data (public community and technical college data). Source: SBCTC enrollment records. A cohort is matched at the time of request. Results are not retained for future matches.
	* Other Higher Ed data (University and other public or private college data). Source: National Student Clearinghouse (NSC).
		1. Internal: A large CTC data set is submitted for a match once a year. SBCTC retains these matched records in our system and matches to those records for future internal higher ed match requests.
		2. External (WTECB): A cohort is submitted at the time of request and could take anywhere from a couple of days to a couple of weeks to get results back. Results are not retained for future matches.Thanks,Monica

Joins and subsets [↑](#footnote-ref-4)
5. See chapter **Measures** [↑](#footnote-ref-5)
6. To make apparrant the change in choice of medain earnings [↑](#footnote-ref-6)
7. Department of Labor, Employment and Training Administration [↑](#footnote-ref-7)
8. Includes 2012-2013, 2013-2014, and 2014-2015 for Workfirst [↑](#footnote-ref-8)