

What is a wiki?

According to Google, a wiki is “a website that allows collaborative editing of its content and structure by users.” That is precisely the purpose of this wiki, in which we will eventually invite users to help us provide metadata information and definitions for the data available through UAccess Analytics.

What is metadata?

Metadata is “a set of data that describes and gives information about other data.” Another word or phrase we could use might be data dictionary.

Therefore, the **Analytics Metadata Wiki** is a dictionary of all of the thousands of data columns available in Analytics. Each definition page also provides a list of the Analytics subject areas that contain each data column. There’s also a Data Warehouse/EPM information section on each page.

How do I use the Wiki?

You can access the Wiki by opening a web browser and navigating to <http://wiki.bi.arizona.edu/>. That link will take you to the Main Page. There are some useful links on the left, and a search function in the upper-right corner. Let’s start with the Search function, as it’s the likely starting point for most Wiki use.



Click in the Search field and just start typing the name of the data column you want. The Search function is not case sensitive. As you type, the autofill function will attempt to find the specific data column you’re looking for. The more characters you enter, the better. Select the data column you’re looking for, once it appears in the list. You could also opt to use the *containing* function at the bottom of the selection list, which will return a list of all data column names that contain whatever string of characters you entered in the field.

What will I see in the Wiki?

There are three sections to every definition screen in the Analytics Metadata Wiki. Those section are the Summary section, the Analytics Information section, and the Data Warehouse/EPM Information section. Each section contains different information pertinent to the data column you selected.

Summary

The Summary section contains the **Definition** of the data column, the **Source System(s)** for that data, and the **Modules** within each system that contain the data column you selected. This section provides you information about what the data means when it’s used in your analysis and, often, what you can expect to see if you were to include the data column in that report. This is the furthest most people need to delve into the Wiki.

In this example, we’ve searched for information about the **Term Code** data column. Here is the information gleaned from the Summary section:

Definition: *This is the code that identifies an academic term, e.g. 2194. The format is CYYT, where C=Century, YY=Year, and T=Term. Term Designators are 1=Spring, 2=Summer, 4=Fall, and 5=Winter. Therefore, 2194 is the Fall 2019 term.*

Source System: UAccess Student, UAccess Employee

Module(s): Student Financials, Set Up SACR, Records and Enrollment, Campus Community, Recruitment Proposal and Offer, DSDS, Curriculum Management, Term and Session, Financial Aid

Analytics Information

This section shows a table identifying which subject areas contain the requested data column and where within the subject area that data column exists. Here’s a sample of the information from the Analytics Information section for **Term Code**, sorted alphabetically by Subject Area:

Subject Area	Folder within Subject Area	Sub Folder	Presentation Field Name	Warehouse Info Row
CSW - Advising - Active Major GPA	Dates and Terms	-	Term Code	3
CSW - Student Academic Profile	Term	-	Term Code	17
CSW - Student Academic Profile Snapshot	Term	-	Term Code	18
CSW - Student Financials	Class - Course Fees	-	Term Code	13
CSW - Student Records - Class Enrollment	Class	-	Term Code	13
CSW - Student Records - Class Enrollment	Dates and Terms	-	Term Code	5
CSW - Student Records - Class Enrollment	Session	-	Term Code	15
CSW - Student Records - Class Instructor	Session	-	Term Code	15
CSW - Student Records - Class Instructor	Term	-	Term Code	17
RCM - Responsibility Center Management	Class	-	Term Code	13

This section is useful in the earliest stages of creating a report. If you’re not sure if a particular data column exists in a specific subject area, or want to know in which subject areas some data column might exist, and where, the Analytics Information section of the Wiki provides that information.

In the sample, you can see that the Term Code data column exists in many different subject areas and, sometimes, in multiple folders within the subject areas. But in every case, the name of the data column stays the same and always means the same thing in the same format.

The numbers in the **Warehouse Info Row** column in this table pertain to the **Data Warehouse/EPM Information** section of this document and the Wiki.

Data Warehouse/EPM Information

This section reveals where the data originated. The numbered rows refer to the *Warehouse Row Info* data column in the previous section of the screen. See the Analytics Information table above.

This is only an example of the full table for the **Term Code** data column. Only the rows identified above were included here.

Row Number	Data-warehouse Schema	Data-warehouse Table	Data-warehouse Field	Data Type	EPM Table	EPM Field	Transformation Logic
3	SYSADM	PS_UA_MAX_SP_CF	STRM	VARCHAR2(4)	-	-	-
5	SYSADM	PS_UA_SP_CURFUT	STRM	VARCHAR2(4)	-	-	-
13	SYSADM	PS_D_CLASS	TERM_CD	VARCHAR2(4)	-	-	-
15	SYSADM	PS_D_SESSION	TERM_CD	VARCHAR2(4)	-	-	-
17	SYSADM	PS_D_TERM	TERM_CD	VARCHAR2(4)	-	-	-
18	SYSADM	UA_SNAPSHOT_CONTROL	TERM_CD	VARCHAR2(4)	-	-	-

In the table above, there are two common values in two different columns. SYSADM, in the **Data-warehouse Schema** column, identifies this as a PeopleSoft schema. You might also see KUALI_ADMIN in this column, if you're looking at Financial data. Neither is material to this discussion.

The other common value here is the VARCHAR2(4) value from the **Data Type** column. This value is more important, in that it identifies both the type and length of the values you'll see in the actual data column in the various subject areas.

VARCHAR2 means that the data is alphanumeric in nature and the (4) tells us that the data has a maximum length of four characters. There's a bit more to it than that, but that's the information that's pertinent for our use here.

Other data types you might see in the Data Type column would be:

Value	Definition
CHAR	fixed-length alphanumeric character strings
VARCHAR/VARCHAR2	variable-length alphanumeric character strings
NUMBER	fixed and floating-point numbers
DATE	point-in-time values (dates and times)
LONG	variable-length character data containing up to two gigabytes of information

The values in the **Data-warehouse Table** column are even more important. They identify the specific transaction system from which the data originated. The table below will help define the first two values in most of the information in the Data-warehouse Table data column.

Let's use **PS_D_SESSION** from the **Data Warehouse/EPM Information** table above as an example.

Position	Value	Definition
1	PS	Table from PeopleSoft, the original developers behind UAccess Student and UAccess Employee
1	KF	Table from Quali Financials, the original developers behind UAccess Financials
1	KC	Table from Quali Coeus, the original developers behind UAccess Research
1	UA	UArizona-created table
2	F	Fact table – the primary keys of the dimension and associated facts or measures
2	D	Dimension table – provides descriptive information for all the measurements recorded in the fact table
2	UA	UArizona-created table
3	Various	Identifies the module where the data originated

You can decipher that example to understand that the data comes from **PeopleSoft**, that it's **dimensional** data, and that it originated in the **Session** module.

The **Data-warehouse Field** column provides the name of the data column as it exists within the table. This often corresponds to the user-facing field name in the source system. This information is most valuable to those who have direct access to the data in the data warehouse, of course, but can also be useful when tracking down suspected discrepancies within the data.