

## Assignment and Comparison of SIC & NAICS Codes

### Overview

Data Axle's U.S. and Canadian Business Databases are organized according to the U.S. Government's Standard Industrial Classification (SIC) Coding system that categorizes businesses by general business activity. Maintained are over 17,000 business category titles which net into 10,500 unique 6-digit (typical is 4-digits) SIC codes. An additional Franchise Code can also be included to further identify product brands, professional specialties and major chains; over 3.7 million records have this additional depth.

Whereas the SIC system classified businesses based on the **products or services they produced**, NAICS classify businesses based on the **processes used to create the products or services**. Data Axle has added two additional digits to the 6-digit NAICS codes to better match conversion from 6-digit SIC codes, offering a significant improvement in granularity.

The primary determinant of SIC assignment is Yellow Page heading from local phone directories. In addition to the categorization being self-reported, this source also offers excellent coverage, is regularly refreshed, and is for public consumption. Our 6-digit SIC-based coding offers a suitable match for any and all Yellow Page headings, and can be expanded to account for new headings and industries.

NAICS recognizes the changing and growing importance of service-based enterprises and groups the economy into over 1,170 industries (of which 565 are service related) in 20 broad sectors (of which 16 are service related).

### YE Time Series Assembly

YE Time Series removes the additional two character from the 8-digit Data Axle NAICS assignments so that the more familiar 6-digit NAICS is standardized.

For assembly of both SIC and NAICS, an algorithm runs looking at all empty and unclassified entries for all years and if necessary will either complete an empty value, or change an unclassified value to indicate type of business activity used if previous and/or following years contain a pattern of active industries codes.

YE Time Series currently uses the NAICS 2017 definition for all records in the database.

## **Census Comparison**

It is common to see variance (sometimes significant), between aggregate industry verticals when looking at Data Axle (therefor YE Time Series) data versus Census. There are several reasons for this:

- With Census, businesses are inputting what they believe to be the most appropriate industry code. However more often than not there are multiple codes that are similar and the business may not select what is actually the most appropriate as industry classification is not typically their strong suit.
- Data Axle uses Yellow Page headings as a guide to assign what is actually the most appropriate industry code, and telephone verify the businesses to confirm the accuracy. This helps to ensure the correct code is selected. As a result, by viewing primary and secondary codes, we typically see approximately 90% or better in terms of industry classification accuracy.
- Since a system is used that is SIC to NAICS based it would expand the opportunity for discrepancy. Data Axle may have created a subcategory that doesn't fit neatly into the 4-digit codes business owners/Census would present

A BDRC comparison between the YE Time Series data and Census determined that Census counts for industry segments that contain financial, real estate, and tax reporting entities can be very high when compared to Data Axle and YE Time Series data. This is primarily due to the government data including establishments that never were "in business" nor had any intent to conduct commercial activities.