

**Attachment P
Department of Commerce
United States Census Bureau
OMB Information Collection Request
Annual Integrated Economic Survey
OMB Control Number 0607-1024**

Census Bureau NAICS Implementation Action Plan for OMB

Census Bureau NAICS Implementation Action Plan for OMB

October 2024

The development of this action plan was spurred by the terms of clearance from the Office of Management and Budget (OMB) to the Census Bureau for the new Annual Integrated Economic Survey (AIES) and subsequent OMB and Census Bureau discussions on NAICS implementation across statistical programs and agencies raised in various responses to data collection clearances. An outline of this plan was developed by staff in the Census Bureau's Economic Statistical Methods Division and presented to OMB in May 2024, with agreement to deliver a full plan that will remain open for updates over time as components of the plan are put into action.

AIES Terms of Clearance

The Census Bureau received OMB clearance for AIES on February 7, 2024. An excerpt from the terms of clearance states:

"In addition, in light of the Census Bureau's finding in Supporting Statement Part B "that NAICS classifications can be unnatural or challenging for some businesses," the Census Bureau within 1 year of this clearance shall provide OMB a research plan (and relevant research updates) to address such NAICS classification issues. This research plan should include ways the Census Bureau plans to estimate the percentage of respondents across collections that select an incorrect NAICS code; how the Census Bureau plans to estimate the extent to which differences in NAICS code assignments by the Census Bureau and the Bureau of Labor Statistics for the same establishments are due to misclassifications in the Census Business Register; and possible approaches the Census Bureau could take to reduce NAICS misclassification."

NAICS Overview and Implementation Challenges

OMB Statistical Policy Directive Number 8, North American Industry Classification System; Classification of Establishments, directs Federal statistical programs to use NAICS "to classify reporting establishments by types of industrial activity in which they are engaged" and prescribes review of the system every five years "in calendar years ending in 2 and 7" for revision.

As stated in the 2022 NAICS Manual, "NAICS is erected on a production-oriented or supply-based conceptual framework that groups establishments into industries according to similarity in the processes used to produce goods or services. ... Though it inevitably groups the products of the economic activities that are included in the industry definition, it is not solely a grouping of products; put another way, an industry groups producing units."

Producing units classified in NAICS are establishments, with industry of establishment determined by its primary economic activity. Two concepts, **establishment** and **primary economic activity**, need further clarification.

An establishment is defined in the 2022 NAICS Manual as “a single physical location where business is conducted or where services or industrial operations are performed (for example, a factory, mill, store, hotel, movie theater, mine, farm, airline terminal, sales office, warehouse, or central administrative office) Exceptions to the single location exist for physically dispersed operations, such as construction, transportation, and telecommunications. For these activities the individual sites, projects, fields, networks, lines, or systems of such dispersed activities are not normally considered to be establishments. The establishment is represented by those relatively permanent main or branch offices, terminals, stations, and so forth, that are either (1) directly responsible for supervising such activities, or (2) the base from which personnel operate to carry out these activities.”

Many businesses operate across industry lines. For example, a hotel operates a full-service restaurant that is conveniently located within the hotel. Hotels and restaurants are two distinct NAICS industries. To classify the combined hotel and restaurant establishment in its primary NAICS industry, the 2022 NAICS Manual states:

“An establishment is classified in an industry when its primary activity meets the definition for that industry. Because establishments may perform more than one activity, it is necessary to determine procedures for identifying the primary activity of the establishment. In most cases, if an establishment is engaged in more than one activity, the industry code is assigned based on the establishment's principal product or group of products produced or distributed, or services rendered. Ideally, the principal good or service should be determined by its relative share of current production costs and capital investment at the establishment. In practice, however, it is often necessary to use other variables such as revenue, shipments, or employment as proxies for measuring significance.”

The Census Bureau often uses revenue and shipments as proxies for measuring the largest share of production costs and capital investments. In the case of the hotel and restaurant, assuming hotel revenue predominates, appropriate classification of the establishment is NAICS industry 721110, Hotels (except Casino Hotels) and Motels.

Industry coding depends on the makeup of the reporting unit. In the hotel and restaurant example, industry coding is simpler when the hotel and restaurant are separately operated establishments. In this case, the hotel is classified in NAICS industry 721110, and the restaurant is classified in NAICS industry 722511, Full-Service Restaurants.

Applying NAICS to reporting units other than establishments is a known NAICS implementation issue that creates challenges not addressed by the classification. For example, a reporting unit that is a multi-establishment company with a diverse scope of activities, with shifts over time due to mergers and acquisitions, magnifies the problem of assigning a single NAICS industry. Multi-establishment reporting units grouped by common activity for use in some statistical programs (such as ‘kind of activity units’, or KAUs, in AIES) simplify industry coding, but may not be recognized units with readily available records maintained by respondents.

Businesses that operate in the virtual rather than the physical space also add to ambiguity in NAICS classification. Intermediation services are rapidly expanding, with noticeable growth in online intermediaries employing platforms, websites, and applications, such as ride-sharing arrangement and food order and delivery services. A single platform can provide intermediation services for a distinct

activity, such as passenger transportation via ride sharing, or a range of activities, such as home improvements, furniture assembly, and dog walking. The broader the range of activities that are intermediated, the more difficult it is to classify in NAICS.

A major benefit of NAICS is the level of comparability across the United States, Canada, and Mexico in the presentation of economic statistics. With every five-year revision, new trilateral agreements are obtained at varying levels within the NAICS structure. The time required to obtain these agreements is built into the revision cycle but can be viewed as slowing down the process and relegating some revisions to the national (6-digit) NAICS level, which is detail that is not consistently implemented across U.S. statistical programs.

Statistical programs implement each new NAICS vintage according to internal needs, capabilities, and resources. Lags of a year or more in statistical programs' adoption of the current NAICS vintage lead to outdated, less relevant data, which is a particularly pronounced problem for sectors where technological change drives new business models. Sector 51, Information, is a sector where industries and structure have been revised significantly with every vintage of NAICS.

In summary, NAICS works well for classification of:

- Single unit companies
- Undiversified multi-unit companies with establishments in the same industry (e.g., locations of a fast-food franchisee)
- Physical establishments that equate to recognizable NAICS industries (e.g., retail convenience stores, bed-and breakfast inns, etc.)

Alternatively, NAICS classification is challenging for:

- Diversified multi-unit companies treated as single units for classification, with ever-changing operations due to mergers and acquisitions
- Industries affected by rapid changes in technology
- Establishments with digital activities that cross industries (such as digital intermediation)

NAICS has been in use for over 25 years. Development of each new vintage provides opportunities to clarify the scope of sectors and industries where there are gray areas. Given the long timespan of implementation experience, the Census Bureau has identified sectors and industries with implementation assignment difficulties and differences across statistical agencies. Some examples of difficult sectors and industries and differences across agencies are listed below:

- Sector 23 Construction

As cited on the Bureau of Labor Statistics (BLS) Quarterly Census of Employment and Wages (QCEW) website, 'For the specialty trade contractor industries (NAICS 238), BLS uses six-digit NAICS codes ending in "1" for residential construction units and "2" for nonresidential construction units, instead of "0.'" This doubles the number from 19 to 38 2022 NAICS industries in subsector 238 maintained by BLS on its Business Register.

- NAICS 334413, Semiconductor and Related Device Manufacturing

The Census Bureau includes semiconductor manufacturing in NAICS industry 334413, Semiconductor and Related Device Manufacturing, based on providing physical transformation and owning physical inputs. BLS and other agencies use a broader set of criteria to classify establishments in NAICS industry 334413, including those not providing physical transformation or owning physical inputs but solely owning associated intellectual property (e.g., designs of physical goods). The concept of factoryless goods production has received much attention over the years, but cognitive testing reveals that respondents at the establishment level do not have the information to appropriately classify such activity. Even with several vintages of questions on economic census instruments to break down the required components of such classification, respondents were found to misinterpret key questions in post-collection interviews. All related questions were dropped in the 2022 Economic Census, and this issue remains a persistent problem for classifying establishments where factoryless goods production is common.

- Sectors 42 Wholesale Trade, 44-45 Retail Trade

For some trade industries, such as selling of office furniture and building materials, the distinction between the two sectors is blurry. In response to cognitive testing, the Economic Census collects various inquiries and evaluates responses using implementation rules to delineate the sector classifications, ensuring consistency in automated edits and manual NAICS classification. Other agencies and programs solely use class of customer criteria that were relied upon prior to NAICS in the Standard Industrial Classification (SIC). Companies may consider themselves part of one sector or the other, but this self-designation does not consistently align with implementation rules.

- Sectors 44-45 Retail Trade, 48-49 Transportation and Warehousing

Sector 44-45, Retail Trade, includes companies that maintain warehouses for the products they own and sell at retail via e-commerce. The primary activity of warehousing is in NAICS Industry Group 4931, Warehousing and Storage. Respondents often do not understand the distinction.

- Sectors 51 Information, 54 Professional, Scientific, and Technical Services

Classification for establishments of diversified companies that publish software but also engage in software development, software consulting, systems design, provision of IT infrastructure, etc. is a challenge, particularly when products are either not reported or are combined in write-in responses.

- NAICS 517122, Agents for Wireless Telecommunications Services

While the Census Bureau follows the NAICS Manual to classify agents for wireless telecommunication services in NAICS 517122, Agents for Wireless Telecommunications Services, BLS classifies them in Sector 44-45, Retail Trade, specifically in NAICS industry 449210, Electronics and Appliance Retailers. The BLS QCEW website states these establishments 'meet the established guidelines for retail trade establishments.'

- NAICS 324110, Petroleum Refineries; NAICS 452210, Department Stores; NAICS 521110, Monetary Authorities-Central Bank; NAICS 622110, General Medical and Surgical Hospitals

These NAICS industries are examples where BLS QCEW and Census Bureau County Business Patterns data on number of establishments are very different. Research on these differences is ongoing.

Government policy analysts, academics and researchers, the business community, and the public use data by NAICS to analyze the size and structure of the U.S. economy over time. Thus, it is incumbent upon suppliers of NAICS data to ensure minimal levels of NAICS misclassification. The Census Bureau cannot verify the accuracy of NAICS codes or data published by outside entities, including other government agencies and private research companies. The Census Bureau understands the importance of accurate NAICS classification and aims for constant improvement, particularly as it consolidates seven separate annual surveys into the single AIES to provide comprehensive national and subnational economy-wide data at the 6-digit NAICS industry level. Through research and innovative remedies such as the use of machine-learning coding applications, the Census Bureau is confident it can continue to deliver high-quality NAICS data for use in critical decision-making in government and elsewhere.

Census Bureau NAICS Research and Remedies

The following sub-sections describe the Census Bureau's NAICS code assignment practices, including its development of innovative machine-learning classification tools, as well as research into respondent NAICS self-coding difficulties and remedies.

- NAICS Assignment Sources

The Census Bureau assigns NAICS codes to businesses using multiple sources, including its own data collections, BLS NAICS codes, Internal Revenue Service (IRS) tax filings, and Social Security Administration (SSA)-assigned NAICS codes for business births. For single-unit establishments with missing or partial NAICS codes, the Census Bureau's 'best administrative NAICS' algorithm generally gives BLS NAICS codes a higher ranking than IRS and SSA NAICS codes to assign Census NAICS codes. BLS and the Census Bureau maintain a Memorandum of Understanding that allows BLS to share NAICS codes for Employer Identification Numbers on its Business Register with the Census Bureau every quarter. There is no reciprocal agreement for the Census Bureau to share NAICS codes with BLS. BLS relies on administrative data from state unemployment insurance (UI) programs, QCEW's Multiple Worksite Report (MWR) and Annual Refiling Survey (ARS), and other collections to identify and update NAICS codes in its Business Register. For more information on BLS methods, see [Concepts : Handbook of Methods: U.S. Bureau of Labor Statistics \(bls.gov\)](#).

Approximately 80% (5.1 million) of Census Bureau single-unit establishments can be linked to a BLS record based on Employer Identification Number. Of these, BLS is the NAICS source for approximately 48% (2.4 million), and BLS and Census Bureau NAICS codes match for an additional 31% (1.6 million).

Measurement objectives and other differences between BLS and the Census Bureau contribute to differences in NAICS implementation. As shown in the table below, the Census Bureau assigns NAICS codes primarily based on output (sales, revenue, receipt classification differences). For example, an e-commerce retailer with a warehouse may end up coded by BLS as a warehouse because more employees are designated as warehouse workers rather than employees performing sales or other retail operations.

| CENSUS BUREAU | BLS |
|---|--|
| Output focused (Sales, Receipts, Revenue, Shipments) | Labor focused (Employment Type) |
| Federal Government based scope (companies that pay federal taxes or file other federal tax forms) | State Government based scope (companies that pay unemployment insurance, excluding certain non-profits, nonemployers, ...) |
| Multiunit company defined as more than one establishment in the US | Multiunit company defined as more than one establishment within a single state |

- Misclassification of NAICS in the Census Business Register

To understand how differences in NAICS code assignments by the Census Bureau and BLS are due to misclassifications in the Census Business Register, the Census Bureau will continue ongoing efforts to investigate specific NAICS with large differences in establishment counts. This involves selecting a sample of cases in the identified sectors that have large discrepancies and manually reviewing for misclassification in the Census Bureau Business Register, using third-party data when available.

- Machine learning classification application – BEACON – current and future uses

BEACON is a machine-learning tool developed by the Economic Statistical Methods Division of the Census Bureau for the 2022 Economic Census. BEACON helps respondents select a NAICS industry if they do not find a pre-listed NAICS industry that fits their business activity on the Census questionnaire. BEACON is invoked when the respondent provides a write-in description of their business. Based on that write-in description, BEACON displays a list of likely NAICS codes and corresponding descriptions, in an order ranked by the application, from which the respondent selects its primary industry.

Overall, BEACON led to a substantial decrease in the number of NAICS write-ins that needed to be manually coded in the 2022 Economic Census, compared to those in the 2017 Economic Census:

| | |
|--------------------------------------|---------|
| 2022 Economic Census NAICS write-ins | 166,000 |
| 2017 Economic Census NAICS write-ins | 500,000 |

Given the accuracy rate (further described below) and success of BEACON in reducing the number of uncoded write-in industry descriptions in the 2022 Economic Census, **plans are to update BEACON to the appropriate NAICS vintage for use in the 2027 Economic Census. Additionally, the Census Bureau will investigate the use of BEACON for AIES in the future.**

- Assistance to the NAICS user community via the Dr. NAICS Large Language Model (LLM)

The Dr. NAICS LLM, currently under development at the Census Bureau, will automate ingest of inquiries and generate responses to general NAICS inquiries that are fielded daily from the NAICS user community by classification analysts through email and phone calls. Early testing indicates a high rate of accuracy in automated responses (over 87%), with concurrent testing and updates to train the model in the coming months before internal release to production. Ultimately, **the Census Bureau will investigate releasing a public-facing Dr. NAICS LLM application on the NAICS website after proving the success of the application for internal use.**

- Identifying, quantifying, and correcting respondent difficulties and errors in self-assigned NAICS codes in Census Bureau collections

The Census Bureau’s most comprehensive and detailed collection of NAICS codes for establishments is the economic census. Historically, during and after each economic census, metrics are calculated for the number of establishments (and percent of the mailed universe) with a NAICS code that is not equal to the mailed NAICS code, the most current NAICS code maintained on the Business Register. By NAICS industry, these classification differences are reviewed and validated for the largest weighted cases and to detect suspicious patterns of misclassification. Macro-level data review and targeted searches coupled with micro-level record review is broadly one of the ways the Census Bureau mitigates respondent assignment of incorrect NAICS codes.

A new measure enabled by the BEACON NAICS classification enhancement in the 2022 Economic Census is the percentage of respondents using BEACON to select a NAICS code. The assumption is that if a respondent engaged with BEACON, they disagreed with the pre-determined NAICS code assigned to them (the mailed NAICS code). This could indicate the degree to which NAICS codes maintained on the Business Register may not be accurate, with a subset of these cases assigned incorrect NAICS codes from respondents in past collections or from administrative sources (BLS, IRS, SSA).

Below are the number and percentage of establishments selecting a pre-listed NAICS code versus number and percentage using BEACON to select a NAICS code in the 2022 Economic Census, as of August 2, 2023. Additional detailed data split out by NAICS sector and multiunit versus single-unit establishments were calculated and reviewed.

| NAICS Self-Coding in the 2022 Economic Census | Number (000) | Percent |
|---|-----------------|---------|
| Establishments selecting a NAICS code (either method) | 2,585 | 100.0 |
| Establishments selecting a pre-listed NAICS code | 2,260 | 87.4 |
| Establishments selecting a NAICS code using BEACON | 325 | 12.6 |

From October 2021 through February 2022, the Census Bureau conducted the Economic Census 2021 Industry Classification Report pilot test. This was used to test new Economic Census questionnaire features, including BEACON, with respondents in a production environment. The sample for the Industry Classification Report consisted of 37,000 single-unit establishments. By design, a third of the sample had a reliable NAICS code of record. These units comprised the so-called “truth deck” and were used to evaluate BEACON’s accuracy. In this analysis, a successful NAICS self-classification is defined as two actions occurring: (1) BEACON returns the correct 6-digit NAICS code as one of its search results and (2) the respondent selects the correct NAICS code. Using the approximately 7,000 truth deck responses received, the following probability estimates for these components of accuracy were calculated:

| | | | | |
|--|---|--|---|--|
| BEACON returns correct NAICS code (P_Return) | | Respondent selects correct NAICS code given that it is a search result (P_Select) | | Successful NAICS self-classification (P_Success) |
| 90.1% | × | 83.7% | = | 75.5% |

Classification staff at the Census Bureau analyzed the use of BEACON for the subset of respondents receiving classification forms solely to collect NAICS codes for small single units in the 2022 Economic Census. **This analysis will inform future collections** and included:

- Identification of the top 20 write-ins left uncoded by respondents
- Identification of the top 20 write-ins coded by respondents
- Respondent-coded NAICS codes matched to mailed NAICS codes on the Business Register, at 2- to 6-digit NAICS levels, for all responses, pre-listed NAICS code responses, and BEACON responses
- Identification of top 20 NAICS codes where self-assignment didn’t match the mailed NAICS code on the Business Register at the 2-digit sector level
- Findings of cognitive testing of NAICS classification questions across collections

Improving respondents’ understanding of NAICS codes and the use of these codes in collecting and publishing data is an ongoing effort. The Census Bureau conducts pre- and post-collection cognitive interviews, often including questions about respondents’ self-classified NAICS coding.

Generally, respondent confusion with assigning NAICS codes is particularly noticeable when the reporting unit is not in sync with the respondents’ records. Below are a few summarized respondent sentiments:

- NAICS is not salient to how businesses keep their records and feels artificial for participants; that is, because NAICS is a standardized classification system, and

businesses often needed more or different details in their chart of accounts, mapping records to the corresponding NAICS is challenging for some and impossible for others.

- Census surveys do not match internal reporting, and they are uncomfortable making decisions on how to manipulate their data to match our requests.
- Participants had trouble understanding their NAICS classification, and then struggled to think of how their business units might be related to their NAICS classification.

The Census Bureau will continue to use the findings of cognitive testing to improve communication and transparency of NAICS definitions, concepts, and methodologies to respondents in survey materials. We will also continue to test the use of novel NAICS coding inquiries and prompts in collection instruments to guide respondents to assign appropriate NAICS codes.

- Disseminate information on machine learning NAICS coding applications and research at conferences, on the Census Bureau website, and in white papers

It is important for the Census Bureau to broadly communicate our work to others on improving NAICS coding using machine learning and our continued development of machine-learning informed models of self-classification. **We will continue to disseminate such information at national and international conferences, on the Census Bureau website, and in white papers as we increase and improve usage of artificial learning tools, such as BEACON.**

Recommendation for Cross-Agency ITWGs

Currently the Economic Classification Policy Committee (ECPC) is a cross agency group chartered by OMB to develop and maintain NAICS. Implementation of NAICS by agencies' statistical programs may be shared and discussed with the ECPC, but adjudication of major implementation differences is out of scope for the ECPC, nor is there any central governance providing cross-agency coding adjudication. An Interagency Technical Working Group for NAICS Implementation Adjudication led by OMB could potentially tackle ECPC limitations in adjudication and harmonization of NAICS coding practices across agencies and statistical programs. Such a group could:

- Research the scope of and motivation for NAICS implementation differences across agencies, adjudicate significant differences for similar data products, and explore options to reconcile differences
- Research and understand definitional differences in units across agencies
- Consider best practices and case studies in NAICS classification implementation across agencies. As described above, BLS modifies national NAICS codes using a unique 6th digit to collect additional detail. The Census Bureau's Economic Census uses unique 7- and 8-digit NAICS-based codes for this purpose. Numerous agencies and programs create custom combinations of NAICS codes for data summaries. Reviewing these and other cases to inform best practices could strengthen the entire federal statistical system.

Additionally, as broad classification topics with economy-wide importance continue to emerge (e.g., bioeconomy, environment, space economy, etc.), OMB could coordinate investigation into such topics

through devoted Interagency Technical Working Groups. Recommendations on NAICS revisions could then be funneled to the ECPC during public comment periods, and technical support provided to the ECPC as it reviews related recommendations for revisions.