Voorburg Group on Service Statistics

12th Meeting

Copenhagen, 15 - 19 September 1997

Developing a Product Coding System for Industries that Cross Sectors

Pamela Powell-Hill
U.S. Bureau of the Census

Session 3

Abstract: This paper addresses some of the conceptual issues related to developing a product coding system. As a way of examinning, which concept might be the best to use in developing a product coding system, the paper discusses the varied and diverse needs of users. Tourism, information technology and the environment as used as examples. Further, the papers stresses the need for flexibility in a product coding system.

1. Introduction

The United States, Canada, and Mexico have adopted the North American Industry Classification System (NAICS). NAICS is a Asupply-side@ or Aproduction oriented@ industry classification system. This industry classification system was designed to provide information for productivity analysis, input output analysis, and other economic studies. However, there is a persistent and growing need of another set of users that is not met by the new industry system.

These users need a system that will provide more detailed data that can be aggregated in different ways. A product coding system that will support various aggregation schemes could be used by these data users. This could include industries that cross sectors such as tourism, or provide the ability to look at specific market segments. The United States is planning to develop a product coding system for use in the 2002 Economic Censuses. We are using product as a generic term covering both goods and services and are interested in a product coding system to cover both.

The key characteristic needed in a product coding system is flexibility because data users want to be able to select products from various industries or sectors to build their own definition of an industry or segment of a market. The flexibility feature poses challenges for the product classification system developers for it not only may require a new conceptual basis for a product coding system, but it may also require new approaches to data collection.

In this paper, we focus on the some of the conceptual issues associated with designing a product coding system and then discuss what might be needed if you were looking to use that system to define industries that cross sectors.

2. General Product Coding Concepts

When designing a product coding system it is necessary to consider many things. Perhaps the most important is who will the users be and how will they use it. It is safe to say that no two users will have exactly the same needs and that various user needs may have conflicting requirements.

Some users would like a product coding system that is based on industry of origin, much like the Numerical List of Manufactured Products that is used by the U.S. Bureau of the Census. Some would like a system that reflects market oriented groupings.

Some would prefer a system based on physical properties, and others may want to see a system based on mode of transport. There is no end to the options and requirements. The challenge then becomes designing a system that satisfies most of users most of the time. So, where does one begin?

The Central Product Classification (CPC) system recognized and addressed some of these issues. The designers of the CPC tried to balance some of these conflicting needs by identifying the industry of origin (International Standard Industrial Classification (ISIC) code) and the Harmonized Commodity Description and Coding System (HS) associated with each subclass. The CPC also recognizes that the way products are produced is not necessarily the same as the industry of origin, although they are often the same. Sometimes an industry produces goods of a totally different nature, for example meat and hides, which are both produced in slaughterhouses. The CPC chose to place the hides, which are considered raw animal materials, in with agricultural products, while the meat is classified with food products.

The Harmonized Commodity Description and Coding System (HS) of the Customs Cooperation Council was designed to facilitate the collection, comparison and analysis of statistics on international trade for transportable goods. It is a detailed hierarchical system often based on physical properties, i.e. plastics and articles thereof; rubber and articles thereof. Again, this system meets the needs of some but does not address the industry of origin needs of other users. The hide/meat example is treated the same as in the CPC but there is no link to show the relationship to industry of origin.

The Numerical List of Manufactured Products used by the U.S. Bureau of the Census is based solely on industry of origin. So, using the example above, both hides and meat are classified with food. Further, you will find different product codes for sausage made in a meat packing plant and sausage not made in a meat packing plant. Most would argue that these are still the same product (sausage) even though the industry of origin is different. This then may be viewed as a data collection issue.

It becomes more apparent as you try to work through the needs and uses of a product coding system that flexibility is extremely important. It makes you begin to think in terms of a matrix that would show various important relationships. In order for this not to get to unwieldy, one would have to decide on the most important links and work with them. As the U.S. begins to formulate the concept for a product coding system it will need to take all of these issues into consideration. A frequent request that comes from data users concerning product coding is to have

the ability to create industries that cross sectors. Some would like to create their own cross sector industries and others would like to have a given set of industries. Let=s examine what might be needed to fill this request.

3. Industries that Cross Sectors

In varying degrees of detail we know about a few industries that cross sectors of interest to users: the environment, tourism, geographic mapping systems (GPS/GIS), and information technology. As you can see from the list, some of these industries, such as information technology and GPS/GIS are emerging and changing at this very moment. For emerging industries, we do not have a lot of data that tells us sector contributions to the industry such as annual growth revenue, and employment. But for others, such as tourism and the environment, we do have revenue, number of employees, and other historical data available to guide us in understanding the products and the sectors that might contribute to it. We can use our understanding of them to design the product coding system.

We see three possible approaches to defining a cross-sector industry and its products:

1. An **all-encompassing definition** that includes every possible product that contributes to the industry including machining nuts and bolts for the machines that build taxi fare meters, assembly-line robots, and so on.

2. A **limited definition** that is narrowed to products more directly linked to the cross-sector industries. For example rather than selecting products across sectors from the root level of manufacturing assembly-line tools, we would select products from the end of the assembly line process such as axles, tires, propellers, rotors, spark plugs, computer chips and so forth.

3. A definition confined to the **major contributors** to the industry that entails selecting only those products directly linked to sector activities which significantly contribute revenue and major employment to the industry. For example, the tourism industry and the transportation sector are very closely linked: without airplanes, ships, trains, cars, vans, buses, taxis, bicycles, motorcycles and so on, there would be no tourism much beyond walking or horse-riding distance.

A good place to start thinking about this complex set of issues is to consider an industry we know right now. So, we will start with tourism because you and I can easily relate to it. This industry illustrates the issues involved in developing a product

coding system. We know that the NAICS sectors that contribute to the tourism are:

1. Transportation (passenger transportation)
2. Accommodation and food services (hotels/motels/bed and breakfasts, RV parks, restaurants, cafeterias, bars)
3. Administrative and support services (travel agents, tour operators)
4. Arts, entertainment, and recreation (performing arts, sports, museums)
5. Retail (gasoline stations; clothing, luggage/leather goods, and sporting goods stores) and,
6. Finance and insurance (travel insurance; banks; credit bureaus).

All-encompassing definition. If we define the tourism industry to encompass any activity related to tourism and tourists, we open the industry coding system to a chain of infinite and extreme possibilities. First, we would have to add other source sectors to our contributors: agriculture, manufacturing, and possibly construction.

We could start moving along the chain of products that define the tourism industry with the food that goes into meals in restaurants. Think about what that means. We would need to gather data at the product level for sectors that contribute vegetables, meat, milk, seafood and so on to the tourism industry. Then we would need to gather data on wholesale and retail marketing of the food. We would then account for all the restaurant activity to prepare meals and serve the food to the customer. Finally, we would include the customer=s payment for the whole chain of activities which, if the customer uses a credit card, opens tourism to another set of products from the finance and insurance sector.

The way today=s economy functions, collecting data that link meals to tourism, for example, is not always possible. In fact we know it is difficult if not impossible to collect data in a way that would help define the tourism industry. For instance, while data users would like to know whether diners in a restaurant were tourists, >locals=, or business people, it would be impossible for business owners to report data that way.

Next let us move to the construction sector--building of hotels, motels, lodges, campground sites, and et cetera. The first thing we would need is product codes that would allow us to identify things like hotel/motel construction, restaurant, amusement park, movie theater, convention center, theme park, airport, train station, and maybe even road construction. This sounds as if it might be more feasible from a data collection standpoint

rather than from product coding. However, it needs to be explored from the latter, product coding, perspective.

In the next piece of the chain, tourism related to manufacturing products would need to include manufacturing of food, golf clubs, transportation equipment, recreation equipment including RV=s, tents, camping equipment; and so on. One might even argue to include all the parts that go into these things along with the robots and other machines that make these parts. Some would argue that this is going too far and might have us including the whole economy.

Another link in the tourism product chain, the distribution network, would include products from transportation, wholesale and retail trade sectors. Transportation actually reaches into the transport of goods to the final users, tourists and all those involved in tourism in this case; and into transportation of these people.

As you can see, even if the economic data were collected at these levels and linked to the tourism industry as well, the branching effect contributes to enormous complexity that could lead to questionable, unmanageable, and unreliable data.

Limited Definition. The first thing we must do is define the industry, tourism.
A dictionary is a good source:

Tourism is: 1. The practice of traveling for recreation. 2. The guidance or management of tourists. 3. a. The promotion or encouragement of touring. b. The accommodation of tourists. **(Webster=s Collegiate Dictionary Tenth Edition 8 1993).**

AThe practice of traveling for recreation@ includes all modes of transportation: cars, buses, planes, trains, motorcycles and so on. And these can be traced down to the very detailed level of nuts, bolts, rubber, rivets through the intermediate level: chassis, tie rods, tires and axles to the final product --a motor vehicle. From there, the finished product is shipped to a delivery site and from there to a dealer who retails it to someone who ultimately uses it for traveling for recreation. Just in this phase, we=ve crossed manufacturing, transportation, and retail sectors.

AThe guidance or management of tourists@ involves tour agents, guides, transportation, roadways, signals, signs, maps, schedules, passes, exchange of money and information. And in this phase of tourism, we have crossed manufacturing, transportation, information, retail, and financial sectors.

AThe promotion or encouragement of tourists and accommodation of tourists@ adds another sector to the list of limited-definition sectors: arts, entertainment, and recreation.

So by this definition of the tourism industry, we would need to begin aggregating data at the detailed product level from manufacturing, retail, transportation, finance, arts, entertainment, and recreation sectors. Notice that in gathering data for tourism, we are talking about aggregating data *across* sectors not hierarchically upward.

Major contributors. An alternative technique to defining an industry as an interaction of all conceivable products is to define an industry as the >sum= of its major sector contributors. Data on receipts/revenue/value added and employment could be the defining factors for determining major contributors. This method cuts out reaching back to the most remotely-connected detailed product data (such as bolts) thus reducing the dimensions of the industries= activities and increasing the possibility of a more reliable and controllable set of data for the tourism industry.

As we have noted earlier, transportation of people is a prime example of a major sector activity in tourism. Airplanes, trains, ships, cars, RV=s, bicycles, motorcycles, taxis, shuttle buses are how people get to the places they are touring. Additionally, trams, aerial lifts, scenic railroads, rafts, canoes, float planes, and helicopters transport tourists to see the sights and experience the recreation they traveled to obtain. This second list can be closely attributed to tourism. Much of the first list except perhaps cars, bicycles and other personal modes of transportation we think might be able to yield data that reflect their contribution to tourism.

Another major activity generated by tourists and contributing to this definition of the industry, tourism, is retail trade. People rent, buy, and repair RV=s, cars, bicycles and all the other transportation modes that we earlier considered. Additionally people purchase gasoline and other fuels to run, cool and heat the vehicles. They replace tires and other automotive parts while >on the road= to and in their destination site.

As we noted in our earlier discussions, tourists eat food on their way to and at their chosen touring and recreation site. They indulge in specialty foods like beignets, local seafood, fruits, vegetables and confections like salt water taffy. They wash it down with sodas, beer and wine. They purchase souvenirs, art, videos, jewelry, postcards, and newspapers. While data users would like to know how much of these products are sold to

tourists versus everyone else, it would be impossible for the merchants to provide this information.

If you wanted to try to take this to an extreme, we could consider parts of the local infrastructure that also support tourists when they are there such as, hospitals, police, and doctors. However, most users would argue that this also goes too far. Barbers, beauty shops, florists, banks, fishermen, grocery stores, and realtors also provide services sometimes used by tourists. Can we measure the tourism here? To implement a major contributors definition of an industry most likely we will have to more narrowly define some area than others in order to maintain the integrity of the industry definition.
Even if we limit the definition of the tourism industry to the major activities of transporta- tion and accommodations, we have narrowed the dimensions of the industry, tourism, but have not solved the problem of availability of data. Much could apply to business travelers, too. This is an issue that architects of the product coding system must address once we have a viable approach to defining the areas for which we need economic data. A central problem with this approach is: What do we do if the industry is emerging and changing so that we do not know what the major components are? This will become more evident as we examine the possibility of creating product codes for the information technology and space industries.

Indeed, the impetus motivating developing a product coding system is that past U.S. classification systems were only designed for the manufacturing sector and were not standardized. Additionally, they were often outgrowths of industry coding systems rather than being designed to stand alone.

The idea of a matrix approach for a product coding system will mean that the architects will have to consider both the detailed level and the possible aggregation schemes of interest to data users. The major contributors definition of the tourism industry discussed above says this:

TOURISM=Transportation+Accommodation & Foodservices+Support Services+Retail+ Arts, Entertainment, and Recreation.

On the face of it, that looks simple enough to code for we merely need to add the detailed economic data for each of those areas and the result is the tourism industry. But it is not clear that the detailed product data relating to tourism for each of these areas is actually collectable.

Optimally, data users of a product coding system require the flexibility to define their own components of an industry and perhaps their own industries to meet their analytical and

research needs. To achieve that end in database terms, they need relational micro data from which they may select and manipulate any data however they wish: flexible aggregation.

These requirements do not necessarily conflict with institutional needs though they pose disclosure and sampling design problems that are not in this paper=s scope. Statistical agencies do need to use the same standardized methods to measure economic activity. To accomplish this they would need to agree on predefined industries and their components. However, this would not necessarily preclude other data users from developing their own industries in addition.

A product coding system that is usable by this diverse set of data analysts is possible if every item input to the process were given an unique code. This means that we will have a very large number of codes. To help manage the product codes, they also need to be gathered into a framework such as groups, classes, sets, sections, or clusters (or all of these). The codes could be numbered in a sequence yet may not have to be used hierarchically. In the context of the products and industries that we have considered here, for example, the general framework for building a product coding system for tourism could look like this:

TOURISM
  Transportation
    Air
       Rail
    Water
    Transit and ground
    Scenic and sightseeing
    Support activities **(for maintaining planes, trains, ships, roads, trams...)
  Accommodation and Food Service
    Hotels and motels
    Bed and Breakfast inns
    RV parks and recreational camps
    Rooming and boarding houses
    Full-service restaurants
    Limited-service eating place
    Cafeterias
    Snack and nonalcoholic beverage bars
    Drinking places
   Administrative and Support
    Travel agencies
    Tour operators
  Arts, Entertainment and Recreation
    Dance companies
    Musical groups and artists

- Sports teams and clubs
- Race tracks
- Museums
- Historical sites
- Zoos and botanical gardens
- Amusement and theme parks
- Casinos
- Golf courses and country clubs
- Skiing facilities
- Marinas
- Bowling centers

Retail
- Gasoline stations
- Clothing and clothing accessories stores
- Shoe stores
- Jewelry, luggage, and leather goods stores
- Sporting goods stores

Finance and Insurance
- Insurance carriers
- Credit union

This framework would then need to be fleshed out to provide more product detail as needed by users.

Let=s turn our attention to two industries with less clear boundaries: the environment and information technology. The environmental industry has been recognized since the mid-1960's as an important and ever-growing industry. These sectors directly contribute to the environmental industries:

1. Manufacturing (pollution abatement systems and facilities, catalytic converters, chemicals, machinery, measurement instruments)
2. Professional, scientific, and technical services (research and development);
3. Waste remediation systems (oil spill clean-up systems; and sanitary systems).
4. Public administration (administration of environmental quality programs).

It also includes activities involved in conserving and protecting endangered species and studying and finding solutions to global warming and other whole-earth environmental issues.

The environment industry is not necessarily identifiable as Aenvironmental@. For example, all airplanes must have an environmental control system to clean the air and maintain the proper mix of oxygen with other gases. Even some of the higher-end automobiles have Aclimate controls@ which might fall into the realm of environment. Like Atourism@, the broader the product

definition the greater the dilution of definition and the more complex the input-output. Greater complexity, especially when the direct relevance of the data is thin, can make the data unmanageable in the sense that once tabulated, we do not know exactly what product we really have.

Major contributors to the environmental industry are those sectors that are significant as big employers, or produce significant revenue, or add value by cleaning up waters, preserving trees, cleaning the air, and/or by meeting federal and local standards.

We know from the Census Bureau=s ***1995 Manufacturing Profiles*** industrial air pollution control equipment that shipments of industrial air pollution devices after manufacture in 1995 totaled $752.8 million. From the Census Bureau=s 1992 Economic Census, we also know that sanitary services had $15.2 billion in revenue in 1992 and from that same census, we know that consulting and design engineering services related to construction and site work yielded $21 billion in revenue in 1992.

The manufacturing, utilities, and professional, scientific, and technical services (including testing services, building inspection services, environmental consulting services, and research) sectors are major contributors to the industry that we call the environment.

Information technology is an industry that is emerging and changing very rapidly right now. It does not have a precise definition though we understand that the term refers to electronic communication of information using equipment such as: computers, modems, satellites, televisions, VCR=s, telephones (cellular, radiotelephone), pagers, microwaves and lasers. The Internet is often considered the central defining element of the industry, sparking many new information communication businesses and methodologies. But emerging as a close second in its growing presence in our everyday lives is the GIS/GPS (Geographic Information Systems/Geographic Positioning Systems). The GIS is hardware and software tools for acquiring, analyzing, storing, and displaying spatially-referenced information. The GPS is made up of satellites or space vehicles, receivers (antenna, clock, processor, battery, and memory), and ground control. Together, the GIS/GPS are used for environmental risk analysis, hazard analysis, urban and regional planning, emergency response, wildlife management, transportation planning and agriculture.

The GIS/GPS= major contribution is not so much large employment and significant revenue as it in other information technology industries. Rather, it contributes accuracy and immediacy to emergency rescue operations and to hazardous waste spills remediation which results in saved lives and protected wildlife, flora, fauna, and people.

As the information technology industry is defined today, its major contributing sectors are: manufacturing (computer and electronic product manufacturing) information (broadcasting and telecommunications, information services and data processing services); professional, scientific, and technical services (computer systems design and related services; advertising services); and educational services. While today information technology has a strong presence in a number of sectors, as we develop our product coding classification system for the future, we need to continually test the classification system against such emerging industries as these to be certain it has the expansion room to accommodate new products that surely will emanate from the ongoing and constant research and development of electronic technology.

4. Summary

Perhaps we have raised more questions than we have answered in our consideration of issues associated with developing a product classification system. We are committed to developing a system in the near future that systematically classifies products. As we have discussed, it is important for the system to be flexible to allow for various aggregation schemes. These aggregation schemes may be defined by the data user or pre-defined by

statistical agencies (to ensure cross-agency industry data consistency or comparability).

As we have noted, some of the emerging industries like the environment and information technology have sectors and products that are themselves emerging as the industry matures. So, the product classification system must not only support flexible aggregations but also major additions and deletions as old products are supplanted by new ones; old industries supplanted by new ones, and entirely new ones are added.

What will the product classification system architects need to do to start the process? We see several key questions that need to be answered:

1. *Data user.* Who will be using the classification system; what do they need in a coding system?

2. *The best framework for the system.* Given the needs and uses, what is the best framework for a new product coding system? Should the coding system be the same for manufacturing and non-manufacturing products?

3. *Product Definition*. What is a product, particularly in the non-goods producing sectors?

4. *Aggregation Schemes.* How many, if any, aggregation schemes should be predefined? How many should be taken into consideration in developing the product coding system?

Some of you have been through this process before and know the challenges we are facing. We are hoping to learn from the existing classification systems and build a new system that is compatible with them.