

How Data Quality Issues Impact B2B Marketers and Modelers

John F. Hood, President
MCH, Inc.

Executive Summary

The failure to recognize or account for key differences between businesses and institutions can severely affect the reliability of statistical analysis of the business-to-business market at the location level. When properly handled, a different view of the institutional segment can lead to significant financial gains.

Institutions present both a challenge and a largely untapped opportunity for B2B marketers and modelers. These large, non-commercial organizations like local government offices, hospitals, prisons, libraries, schools, and churches account for about 12% of the records on a typical business-to-business database, but they represent approximately 33% of the U.S. economy. In other words, institutions have more potential buying power than commercial businesses. As a result, institutions are vitally important customers for many B2B marketers such as marketers of furniture, office supplies, packaging materials, apparel, and motivational items.

Unfortunately for marketers and modelers, institutions are not businesses, but they are treated like businesses in most databases. This leads to several shortcomings in handling institutional information:

- Many business metrics (attributes) are not appropriate for institutions.
- The SIC system, which is almost universally used to classify businesses, is out-of-date and inappropriate for institutions. NAICS, less widely used, is an improvement but still misses some institutional nuances.
- Because of their inter-relationships, many institutions share physical and/or mailing addresses and often have similar names. Accurately matching institutional records, a fundamental part of the modeling process, is difficult and presents a challenge to users of industry standard software matching tools.

These issues seriously undermine the potential of models affecting one-third of the U.S. “business” economy.

Solutions exist but they are not easy. Specialist institutional databases are available but they require their own models because of the different types of attributes. Some very important SIC equivalents like Hospitals represent only 7,000 records yet have huge economic value. Serious modelers will find it pays to make the effort to locate appropriate institutional data and incorporate it in their models.

Overview

GIGO, garbage in-garbage out, is one of the oldest maxims in the IT world. It also applies to the world of statistical modeling. Prospecting models make predictions based on mathematical formulas that depend on the relationship between external data such as a B2B database and *internal data* such as customer sales. If one of the datasets or the relationship between the datasets is inaccurate, the integrity of the model will suffer.

A prospecting model as discussed here attempts to predict which organizations are the best targets for new sales. The modeler accomplishes this by comparing a set of customer records (or observations) to an external database that represents the broader target universe. Because the external file also contains descriptive attribute data (“number of employees”, “state”, etc.), it is possible to identify a combination of attributes which correlate most strongly to predict responses and buying activity among prospects. Those findings are used to generate an algorithm that scores every prospect record with a prediction of its likely behavior.

In this discussion, we assume that internally supplied data is accurate because it comes from internal sources such as sales records. This is not necessarily true in all cases, and it is extremely important that the limits of internal data are understood. If internal information is not accurate, users must make judgments about the value of any resulting model.

It is much more difficult to assess the accuracy of externally supplied data. Even the compiler of the database may only have a sketchy idea of its quality. Some of these databases are so large and complex—millions of records—that it is impossible to make a truly accurate assessment. There is no accepted independent industry organization that measures and reports on data quality.

Regardless, there are general problems or weaknesses that can be found in all compiled databases including: out-datedness, incompleteness, inaccurate values, imputed values, missing values, abbreviations, typographical and spelling errors, “doing business as” names, classification errors, inconsistencies, and field truncations. The list seems daunting, but if the errors occur in a small percentage of the database, the results of models can be highly useful. Statisticians also have workarounds to mitigate or eliminate some of these issues. This process, called data transformation, is where art meets the science of modeling.

Aside from data issues, there is another source of error that is even more important to understand: record matching. The statistical processes that generate the predictive algorithm are *wholly dependent* on the relationship between the customer information (like sales dollars) on the internal file and the descriptive attributes on the matched external file record. When based on properly matched records, the resulting mathematical formula should generate a reasonable set of predictions for the prospect

records. *But what happens if a match is incorrectly made?* When this occurs, the wrong external data is linked to the wrong customer, and that error is integrated into the mathematical formula which then propagates the error throughout the model. If there are many matching errors, the faulty projections become serious.

Both types of data errors occur to some degree in all prospecting models, but these types of errors are even more difficult to deal with in the institutional segment of the B2B world. Of course that would not really be of concern if institutions were a small factor in a B2B company's market. In reality, the institutional segment is an essential portion of the customer files of most B2B marketers—even though many marketers don't realize it.

Institutions are the not-for-profit sector of the economy. Institutions are all around us: hospitals and nursing homes, churches, schools and colleges, government offices, museums and libraries, and many more. At MCH we call institutions the "purpose-driven" sector of the economy to contrast them with for-profit businesses. Institutions account for about one-third of the economy or over \$4 trillion in GDP. They have been growing faster than businesses for at least the past 50 years and will continue to do so. If you think about it, institutions (such as your local police station) are intrinsically more stable than businesses (such as a local restaurant). Because of their funding and accounting methods, institutions are also more creditworthy than businesses. The average institution is larger than the average business. These factors make institutions ideal B2B customers.

To get the most out of the institutional sector, you need to pay attention to the ways in which institutions are different than businesses and how those differences impact your models and other decision support tools.

Issues with Institutional Data

Compiled marketing databases come in two varieties: broad, general files and specialized files. The broad B2B databases attempt to compile data about every entity doing business in the U.S., approximately 16,000,000 or more records. On the other hand, specialty databases focus on a narrowly defined group of businesses or institutions.

The broad B2B databases attempt to collect every entity doing business in the U.S., so the compilers use sources that are relatively universal such as telephone yellow page databases, credit reports, and government filings for licenses, liens, and other purposes. A weakness, however, is that these sources fail to identify many entities that don't advertise and don't use credit, namely institutions. Once the organizations have been collected for inclusion in the general database, they are assigned to SIC groupings using a variety of methods—some more accurate than others.

In contrast, the specialty databases do not have as comprehensive a mission. Typically they are focused on one type of business or institution, and they have to be very thorough to be viable. An example is a database of hospitals. The compiler will use many methods to make sure all hospitals and nothing but hospitals are on the database which makes SICs irrelevant.

A second key difference between the two types of databases is the type of data that is collected about each business or institution. On the broad databases, the most common indicators besides the SIC itself are number of employees at the location and the sales volume. These attributes are used for every entity

whether appropriate or not. The one-size-fits-all approach is necessary due to the broad-based nature of the database. The specialty databases, however, collect attributes unique to their target business type. In the hospital example above, the file is likely to include the number of beds, the presence of a 24-hour emergency room, and many other services provided by hospitals.

The SIC System and Missing SICs

Most B2B databases use the Standard Industrial Classification (or SIC) system as the basic way to classify or differentiate among all the records. In this discussion, I assume readers are familiar with this system. A newer system, the North American Industry Classification System or NAICS, is a big improvement. However, it is not as widely used and, for example, does not correct the important deficiency in differentiating between schools and school district offices, as below, among others. If NAICS is available, it should be used instead of SICs.

While SICs are long established and widely used, the system has many inadequacies. SICs originated in the 1930s when manufacturing was the dominant sector of the U.S. economy. Even though they have been updated periodically, the government SIC classifications still don't represent institutions well. For instance, institutions represent only about 5% of the defined SIC codes. In contrast, institutions comprise about 33% of the economy. Clearly, institutions are not nearly as finely segmented by SIC codes as businesses.

Churches and non-church religious organizations, for example, are both classified in the same SIC category. From a marketer's point of view, these two types of organizations are completely different from one another with different needs, buying influences, seasonal patterns, etc. Likewise, schools and school district offices are grouped into a single SIC yet they have totally dissimilar functions. There is no SIC defined for diagnostic imaging centers which are popping up all over. There is no SIC for assisted living facilities, a fast-growing sector of nursing and retirement homes. There are numerous other examples of SICs being out of alignment with our current economy

One last example: Ambulance services are grouped in the same SIC as limo services and sightseeing busses: Local Passenger Transportation, Not Elsewhere Classified. Aside from the absurdity of associating a ride in an ambulance with a sightseeing tour, think about the description of the category for a moment—*Not Elsewhere Classified*. *Not Elsewhere Classified*, or "nec," is a term that occurs throughout the SIC definitions. Here are just a few examples:

- Amusement and Recreation, *Not Elsewhere Classified*
- Building Maintenance Services, *Not Elsewhere Classified*
- Health and Allied Services, *Not Elsewhere Classified*
- Miscellaneous Personal Services, *Not Elsewhere Classified*
- Nursing and Personal Care, *Not Elsewhere Classified*
- Social Services, *Not Elsewhere Classified*

The Government's official definitions use *Not Elsewhere Classified* categories as catchalls for various types of organizations that don't fit neatly into bigger categories. While that alone can negatively impact the value of SICs in modeling, the problem is even more extensive in actual practice. General database compilers are forced to use these categories as repositories for many of the organizations where they do not have enough information to assign a precise SIC.

Classification Issues with SICs

If one examines individual records in a typical broad database organized by SICs, it is easy to pick out numerous suspect classifications. When you think of it, it's not at all surprising. Do you know your own SIC? It is not a question you can ask in a phone survey and get reliable answers. Other techniques are similarly problematic. The definitions in the SIC manual are not that easy to understand especially if you are making a lot of decisions quickly. Note the following definitions:

8051: Skilled Nursing Care Facilities

"Establishments primarily engaged in providing inpatient nursing and rehabilitative services to patients who require continuous health care, but not hospital services. Care must be ordered by and under the direction of a physician..."

8052: Intermediate Care Facilities

"Establishments primarily engaged in providing inpatient nursing and rehabilitative services but not on a continuous basis. Staffing must include 24-hour per day personnel with a licensed nurse on duty full-time during each day shift..."

8059: Nursing and Personal Care Facilities, Not Elsewhere Classified

"Establishments primarily engaged in providing some nursing and/or health related care to patients who do not require the degree of care and treatment that a skilled or intermediate care facility is designed to provide..."

To assign nursing homes to the proper categories requires much more information about each entity than you can find in a yellow page listing. Would you like to be responsible for correctly parsing these distinctions and the innumerable others imbedded in the other 1,000 SIC categories? In fact, for a user who is not fully aware of the choices, there are several other SICs that might plausibly cover nursing homes such as 8361: Residential Care.

The difference with a specialty nursing home database is readily apparent. By ignoring the SICs the specialist is able to focus on important differences that are the product of our current economy. They might include nursing homes, nursing homes providing for Alzheimer's patients, nursing homes with assisted living facilities, etc. These distinctions are more relevant to what is happening at the location and also what the needs of the location are, fundamental attributes that drive buying behavior.

Missing Institutions

The customary compiling data sources used by the B2B compilers are not sufficient for identifying many institutions, making the databases largely incomplete in the institutional segments. When institutions are not available in the universe file, buyer records cannot be matched to them, creating gaps in the information available for generating a predictive model. In contrast, specialist database compilers find other sources to build their more complete databases.

Some of the most widely used business files on the market are by-products of credit-reporting agencies. The compilers earn the bulk of their revenue by providing credit scores. They identify and classify businesses based on credit reporting. That method breaks down in the non-profit, non-commercial world of institutions. Many institutions simply don't use credit. They receive funds from intergovernmental transfers, grants, endowments, contributions, and taxes and only spend those funds once they are received.

Institutional accounting focuses on controlling the funds so they will be used only for the stipulated purposes. When a purchase order is released, funds are encumbered meaning the money is reserved

for the purchase and can't be spent on anything else. The P.O. is only released if the money is available, so there is no need for credit. The B2B databases that use credit measurement as a prime purpose have little interest in institutions.

Likewise, while many institutions advertise in the yellow pages, many others don't. Most public schools don't. Some churches do, some don't. Yellow pages advertising databases can identify a wide variety of organizations, but it is far from a complete source of institutional records.

Appropriate Attributes

All viable enterprises have employees, but only businesses have sales and profits. Because broad B2B databases describe all organizations—even non-businesses—with business attributes, they are inadequate for measuring the buying power of institutions. When the mission is to identify and classify every entity in the U.S., a consistent compiling approach is mandatory. For the general database compiler, this creates another crucial dilemma. Until the entity is classified by SIC its status as an institution may not be known, but by then it is typically too expensive to go back and collect specialty data.

In many cases a business attribute such as "Number of Employees" may seem to apply to an institution just as accurately it would to a business. A bit of reflection, however, reveals potential problems with that assumption. For example, churches and many other institutions are largely staffed by volunteers. They appear under-sized in evaluations based solely on an employee count. Employee size measures are also distorted for schools and colleges. They have many employees, but they also serve far more people on a full-time basis than their employees, namely their students. At a school or college there are 5 to 10 times more people in the institution than are captured in an employee count. When an institution is required to provide full-time services for these extra people, the economic impact of the number of employees is surely understated.

Again, the specialty database may or may not include an attribute like number of employees. If the specialty database is of churches, schools, libraries, museums, or colleges, employee size may not be available because typical users of specialist files want more relevant information such as the number of patrons or students. The key is that the attribute will be appropriate for the type of institution. That benefit is a penalty for the modeler, however, because different types of attributes are difficult if not impossible to deal with in a single prospecting model for a B2B marketer.

Summary of Data Issues

Broad B2B databases do not handle the institutional one-third of the economy well because of fundamental problems with the SIC code structure, because of the difficulty of correctly classifying institutions within that structure, because the source data underlying their compiling is incomplete, and because standard business metrics are inappropriate to describe institutions. Specialty databases resolve those defects but provide a new challenge to the modeler, namely that the modeler must identify when and which specialty databases are needed and then acquire them. In order to use the data in a model, the modeler must somehow convert specialty attribute data to business equivalents like employee size, SIC, and sales volume. Alternatively, the modeler can generate a model for each specialty database. However, that might mean generating individual models for, say, 7,000 hospitals or 300,000 churches when the overall objective is to model the broad universe of 12,000,000+ organizations. Attempting to develop a B2i predictive model by solely relying on typical compiled firmographic databases is a compromise that will lead to inferior/ineffective models.

Issues Matching to Institutions

As noted above, matched records (or observations) are the basis of prospecting models. Any mismatched observations are embedded in the mathematical formulas and are then propagated throughout the resulting model. Accurately matching records among two or more databases involves arcane and daunting issues.

In a mailing application the matching standards can be quite loose. The objective is simply to eliminate duplicate addresses to save postage, and the consequences for a mismatch are not severe. If you fail to identify a match, a duplicate mail piece reaches the recipient. If you make a match where there is none, someone doesn't get the mail piece. For mailers, it is possible to strike a balance that allows the matching process to be simple, fast, and inexpensive.

When your objectives require a high degree of matching accuracy, however, the tendency is to tighten the matching standards. The result, of course, is fewer matches. Fewer observations generate a model with a lower level of confidence. This built-in conflict between matching accuracy and quantity of observations is a potential pitfall for a modeler. It is tough to determine matching error rates without extensive scrutiny of the results. B2B matching is more difficult than B2C matching because of the company name or the 4th address line. A simple example of different versions of a company name that commonly prevent matches includes IBM versus International Business Machines. At the same time, the use of suite numbers can prevent matches or generate mismatches depending on how they are handled by the matching software. Your matching resource or vendor can review how they handle these types of problems, and you should ask.

As difficult as it is to match "business records" to a B2B database, the problem is even more challenging when "institution records" enter the mix. There are a variety of issues related to organization names, addresses, and phone numbers.

Similarity of Names

Entities with similar names cause one of the most common institutional data matching challenges. Many institutions are publicly funded and are part of the identity of their communities. As a result, dozens of institutions within a community frequently include their city name as part of their institutional name.

Many standard company name matching algorithms allow for less than perfect name matching which is normally very effective at identifying the same name when spelled or abbreviated slightly differently. In the institutional world, however, one must be very careful in the use of those software features. In the example below, the records represent a wide variety of services with widely divergent budgets, purchasing trends, and product needs. Matching errors could introduce severe distortions, especially if SIC codes are a key component of the model's algorithm.

ZIPCODE	INSTITUTION NAME	ADDRESS	CITY
77338	HUMBLE ACADEMY CHILDCARE	107 GRANBERRY ST	HUMBLE
77338	HUMBLE ANIMAL HOSPITAL	114 S HOUSTON AVE	HUMBLE
77338	HUMBLE CARDIOLOGY ASSOC	9950 MEMORIAL BLVD # 201	HUMBLE
77338	HUMBLE ELEMENTARY SCHOOL	20252 FIELDTREE DR	HUMBLE
77347	HUMBLE EVEN START FAMILY LEARNING PAT	611 HIGGINS	HUMBLE
77338	HUMBLE FIRE DEPARTMENT	108 WEST MAIN STREET	HUMBLE
77338	HUMBLE HEAD START	110 ISAACKS RD	HUMBLE
77338	HUMBLE HEALTHCARE CENTER	93 ISAACKS RD	HUMBLE
77338	HUMBLE HIGH SCHOOL	1700 WILSON RD	HUMBLE
77338	HUMBLE INFANT ACADEMY	311 E MAIN ST	HUMBLE
77338	HUMBLE MEDICAL CENTER	9810 FM 1960 BYPASS RD W STE 115	HUMBLE
77346	HUMBLE MIDDLE SCHOOL	11207 WILL CLAYTON PKWY	HUMBLE
77338	HUMBLE MRI AND DIAGNOSTIC CENTER	19007 HIGHWAY 59 N	HUMBLE
77339	HUMBLE PEDIATRIC ASSOC	611 ROCKMEAD DR # 600	HUMBLE
77338	HUMBLE POLICE DEPARTMENT	310 N BENDER AVE	HUMBLE
77338	HUMBLE PRAYER HALL	7211 FM 1960 RD W STE 180	HUMBLE
77338	HUMBLE SHERIFFS OFFICE-DIST 2 SUBSTATION	7900 WILL CLAYTON PKWY	HUMBLE
77338	HUMBLE SLEEP DISORDER LAB	18955 N MEMORIAL DR # 400	HUMBLE

Shared Physical Addresses

Two examples of shared physical addresses are shown below. The first are two separate examples of churches and affiliated schools located at the same place. This is very common and other institutions co-located with churches include childcare centers and senior centers. City halls, police departments, fire departments, and other city facilities commonly share the same address.

ZIPCODE	INSTITUTION NAME	ADDRESS	CITY
76248	ST ELIZABETH ANN SETON CATHOLIC SCHOOL	2016 WILLIS LN	KELLER
76248	SAINT ELIZABETH ANN SETON CHURCH	2016 WILLIS LN	KELLER
76248	NEW HOPE CHRISTIAN SCHOOL	2105 WILLIS LN	KELLER
76248	NEW HOPE LUTHERAN CHURCH	2105 WILLIS LN	KELLER

The doctor and dental practices shown below demonstrate another variation of co-location. The suite numbers differentiate between the practices, but not every database compiler or matching resource will account for suite numbers well.

Matching software often employs fuzzy logic to account for abbreviations and minor spelling differences, and that software could easily treat minor differences in suite numbers as if they were the same address. Again, mismatched medical and dental records will distort statistical results. This issue is compounded by the many other types of health services such as labs, pharmacies, and diagnostic imaging offices are also often located in these types of medical office complexes.

ZIPCODE	INSTITUTION NAME	ADDRESS	CITY
75007	AHMAD MEDICAL PRACTICE	3730 N JOSEY LN # 100	CARROLLTON
75007	CAMPBELL DENTAL PRACTICE	3730 N JOSEY LN # 102	CARROLLTON
75007	MEDICAL DOCTORS CLINIC	3730 N JOSEY LN # 104	CARROLLTON
75007	MC QUADE & ROSSMANN	3730 N JOSEY LN # 110	CARROLLTON
75007	SUSTER MEDICAL PRACTICE	3730 N JOSEY LN # 111	CARROLLTON
75007	TRAMMELL DENTAL PRACTICE	3730 N JOSEY LN # 126	CARROLLTON

Shared Mailing Addresses

Mailing addresses generate additional matching issues. The example below shows an instance where a school system’s district office receives and redistributes all the mail for every public school in the community. As a result, every institution has the same mailing address, and the high school and the school district office even share the same physical location. This situation is fraught with mismatching potential.

In this situation, all records except LITTLE BUCKS DAYCARE share the same SIC classification. However, as we noted in the discussion about SICs above, schools and school districts should absolutely not share an SIC because they serve very different functions. This example – with all of its uncertainty for a modeler— is very common in public school databases.

INSTITUTION NAME	PHYS ADDRESS	CITY	MAIL ADDRESS
FREER HIGH SCHOOL	905 S NORTON ST	FREER	PO BOX 240
FREER JUNIOR HIGH SCHOOL	615 S NORTON ST	FREER	PO BOX 240
NORMAN M THOMAS ELEMENTARY SCHOOL	1404 S NORTON ST	FREER	PO BOX 240
FREER INDEPENDENT SCHOOL DISTRICT	905 S NORTON ST	FREER	PO BOX 240
LITTLE BUCKS DAYCARE	906 E AVE	FREER	PO BOX 240

The issue is not limited to public schools and district offices. The example below shows three different parts of city government—with three different SICs—sharing a mailing address even though they are located in three different places. These institutional address landmines highlight the need for a modeler’s matching logic to account for both mailing and physical addresses.

INSTITUTION NAME	PHYS ADDRESS	CITY	MAIL ADDRESS
YANTIS CITY HALL	100 N MAIN ST	YANTIS	PO BOX 245
YANTIS SENIOR CENTER	102 N MAIN ST	YANTIS	PO BOX 245
YANTIS VOLUNTEER FIRE DEPARTMENT	117 NORTH MAIN STREET	YANTIS	PO BOX 245

Be Careful Using Telephone Numbers

When telephone numbers are available, they are often used to “improve” the match process. This creates problems similar to the shared address issues above. Instances of common phone numbers across multiple types of institutions abound. If your matching logic uses telephone numbers as a high matching priority, you will get multiple matches or capture only the first record with the matching phone number.

Following are examples where one telephone line is used to serve multiple entities: medical office suites where there multiple practices and services are supported by a single receptionist; various government offices routing calls through the same main number; schools and school districts sharing the district phone number; and many other circumstances.

What You Can Do to Help

It’s possible for a marketer or modeler to help your matching team or matching vendor deal with some of these issues. Customer records may contain data that can make matching more accurate. For example, any classification data that might indicate what kind of institution your customer is should be included on the matching record. It is important that this information comes from your sales force or another internal source rather than from an overlay of a general B2B file. Otherwise, you’re relying on the results of a prior problematic match of data. If you have a hospital sales force and know that all of their customers are hospitals, those records can be flagged as hospitals. Your match vendor can then use that flag to ensure those records are matched only to hospitals.

Another method to improve match results is to prioritize types of institutions based on your knowledge of your company’s sales strategies or product profile. Let’s say you know that both churches and childcare centers are on your customer file. You also know that you spend much more marketing and sales efforts on the churches and that the childcare customers are more or less an afterthought. Your vendor can use that knowledge to prioritize the matching so that churches are matched before childcares. There may still be errors, but the errors will be skewed towards your priorities and therefore will be less damaging.

Summary of Matching Issues

If your company or resource uses standard industry matching software, you must be aware and compensate for the name and address similarities above. Mismatches that cause wrong SICs to be used in the model will undermine its accuracy. Extremely “tight” matching standards will help alleviate this problem but result in much lower match rates and fewer observations. Matching accuracy is extremely important because the matches are the foundation upon which the mathematical formulas rest. Any inaccuracy in the formula is spread throughout the prospecting model when it is used to score the prospect database.

Conclusions

Prospecting models are only as good as the data that is used to generate them. Errors commonly occur in compiling external databases. Users have no control over the accuracy of the external data except in choosing which data sources to use. When using a general B2B database, the institutional one-third of the economy is likely to be more error prone than the business portion because of structural issues within the SIC system and the use of business metrics to describe non-businesses. Specialist databases exist and supply more accurate data for institutions and some business segments. While the specialist datasets may add complexity to the development of individual models, database marketers should strongly consider using them to describe the institutional portion of the universe better.

Matching records is also more difficult in the institutional space. Mismatches are propagated through the model and thus have the potential to substantially undermine its accuracy. If a high-quality source of data is used for the institutional universe, applying different matching standards for institutions and businesses could improve the accuracy of matches.

Modelers need to be aware of the pitfalls caused by institutions in B2B models. Just because they are harder to deal with doesn't mean that institutions should be ignored or avoided. How do you know whether you should acquire specialist databases? Here is one suggestion.

Analyze your customers as best you can. If the institutional sector is as important to you as their contribution to the economy, they should be reasonably prominent in the customer file. Are there a significant number of institutions in the file? Do they have larger than average sales? Is their RFM or lifetime value better or higher than average?

While these questions cannot be answered definitively using general B2B databases to classify customers, clues will certainly emerge. An analysis of several key SICs will be indicative:

8062—General Medical and Surgical Hospitals

8211—Schools and School Districts

8661—Religious Organizations

If any of those SIC segments are above average performers, it's a signal that you need specialist institutional databases to get the maximum accuracy and value from your models. The better those segment perform, the more important specialist data is to you.

When using the specialist databases, keep the matching process separate from the customary business matching process where possible. Work closely with your match vendor to create different and tighter match settings to reduce the potential for matching errors. By reviewing the results of the matching rules, you will be able to fine-tune the process as needed.

Institutions represent significant buying power which gives marketers and modelers much to gain by incorporating institution-oriented practices in their statistical analysis. To generalize, it's best to use business data to match and model business behavior and institutional data to match and model institutional behavior.