Child Nutrition Programs Operational Support Branch Issue 13| Winter 2017

Direct Certification (DC) determines student eligibility for National School Lunch Program (NSLP) benefits based on Supplemental Nutrition Assistance Program (SNAP), Temporary Assistance for Needy Families (TANF), or Food Distribution Program on Indian Reservations (FDPIR) eligibility (or through several other allowable categorically eligible designations). As part of a special technical assistance effort, the Food and Nutrition Service (FNS) has been visiting States to identify innovative and promising practices for improving direct certification rates.

This newsletter presents information gathered during these direct certification visits that may help other States to identify and implement mechanisms to improve automatic access to free school meals for eligible children, as well as increase certification accuracy and save valuable time and resources for school districts.

SY 2014-2015 Report to Congress Highlights Continued DC Performance Improvement

The most recent installment of the *Direct Certification in the National School Lunch Program Report to Congress (RtC)*, assessing the effectiveness of State and local efforts to directly certify children for free school meals, provides readers with detailed information on continued positive results for both States individually, and nationally overall. SY 2014-2015 highlights include:

- In SY 2014-15, an estimated 91 percent of school-age SNAP participants were directly certified for free meals - a four point improvement from SY 2013-2014.
- Twenty-four State agencies achieved direct certification rates at or above HHFKA's 95-percent performance target - doubling the SY 2013-2014 total.
- States agencies provided insight on system improvements including employing probabilistic matching, automated email reminders, and address validation applications to improve source data quality.

For access to the full SY 2014-2015 RtC, please visit the USDA FNS Website at: https://www.fns.usda.gov/report-finder .

INSIDE THIS ISSUE

Match Algorithm Improvement: Developing an Approach to Improving Match Logic on a Continuous, Iterative Basis

Part 2 (of 2 Part Series) on Improving Your Match Algorithm

In our last issue of Match to Meal we discussed the importance of (and approach to) continuously evaluating the effectiveness of your match algorithm. In this issue we provide insight into some of the more advanced techniques States can use to improve the recall and precision of their algorithm.

Since its beginning in 2012, the FNS Direct Certification Training and Technical Assistance (TTA) team has visited over 40 States— all with interest in some shape or fashion in improving their Direct Certification match rate. And regardless of a State's varying need the question most commonly asked is, "What advanced techniques are other states using to get such high match rates, and how do we implement them?" While not directly stated, the question often implies a request for a "magic-bullet." Unfortunately, as outlined in the Part 1 of this series — the ability to apply a single answer is not possible as data characteristics (e.g. available attributes, quality) and processes (e.g. frequency) that influence the potential ROI of any given revision to the State's current algorithm technique are not uniform. That said, experience gained through the site visits has introduced multiple advanced techniques which have demonstrated to be advantageous in multiple and varying systems, and may be well-suited to your State. In Part 2 of this series we introduce the following three opportunities for your consideration: 1) Phonetic Matching; 2) Nickname Matching; and 3) String Matching.

1 Phonetic Matching: Many names in the English language have more than one spelling and are frequently recorded incorrectly in a database due to manual processes (e.g., Cathy vs. Kathy, Marc vs. Mark, Jeff vs. Geoff, etc.). If the "incorrect" (or alternative) spelling of a name is entered into one database (e.g. the Student Information System) an exact comparison function will (most likely) not yield a match from another database (e.g. the SNAP participant file). Phonetic Matching, as the term implies, is a technique that aims to solve this issue as it matches records based on the way they would sound if spoken. Specifically, when a phonetic encoding function is used, the computer can recognize the two as the "same" name, making the proper match. Below are a few examples of popularly-employed phonetic matching methods and how each method works:

- Soundex: Soundex is the most widely known of all phonetic algorithms (in part because it is a standard feature in many popular database software such as DB2, PostgreSQL, MySQL, Ingres, MS SQL Server and Oracle). The function works by encoding characters in a name as a numeric digit. Specifically, it retains the first letter of a name, removes vowels, assigns numbers to the remaining letters, and then removes repeated digits. The code is then truncated (or zeros are added to the end) to create a code containing the first letter followed by three digits. The two encoded values are then compared. The simplicity of the technique makes it effective — however there are limitations associated with non-English pronunciations, and situations in which two names begin with different letters (e.g., Kris and Chris). In these situations other approaches are needed to supplement the process in order to match records.

- NYSIIS: The New York State Identification and Intelligence System (NYSIIS) has been proven to be slightly more accurate than Soundex. NYSIIS incorporates a number of rules to create a code made up of all letters. The complicated set of rules was carefully derived to ensure similar codes for similar sounding words and/or names. The algorithm logic can be found on Wikipedia: *https://en.wikipedia.org/wiki/New_York_State_Identification_and_Intelligence_System*. However, similar to Soundex it is only applicable to words and names in the English language.

- Double Metaphone: Similar to NYSIIS, Double Metaphone follows a complex set of rules to transform names into an all-letter code. However, unlike NYSIIS, both a primary and a secondary code are created using different sets of rules to account for the ancestry of the name. The first code follows rules for the English pronunciation of the name, while the second follows rules based on the name's native dialect. This can help to account for errors in translating foreign names to their English-spelling variation — allowing Double Metaphone to be especially accurate when accounting for European and Asian names, and resolving differences of the same name with different spellings.

Continued on page 3...

Continued from page 2…

2 String Matching: String Matching treats two data entries as "strings" of letters or numbers and determines the physical similarity between them. It is most useful for rectifying spelling or typographical errors. Although there are numerous String Matching techniques, two of the more popular and widely used (and encountered during the TTA team's site visits) are Levenshtein Edit Distance and Jaro Winkler.

In general, these techniques should be used later in a matching strategy, and after other techniques have been used to decrease the number of strings that need to be compared (i.e., first, last, DOB; nickname, DOB modifications, phonetic matching, etc.) as they can be taxing on your system because each string must be compared to every other string — and with millions of strings in your data set, that is a tremendous number of comparisons.

- **Levenshtein Edit Distance:** This technique involves evaluating the similarity between two names (i.e., strings) by comparing the number of single character insertions, deletions, and substitutions that it takes to convert one string into the other. Programming functions for Levenshtein are readily available on open source platforms such as *StackOverflow.com* and *Github.com*, (both of which are sites designed to be communities of programmers helping each other in their programming efforts). For a deeper dive into this technique, an example of an optimized SQL implementation can be found on the web at: *http://blog.softwx.net/2014/12/optimizing-levenshtein-algorithm-in-tsql.html* .

- **Jaro Winkler:** Jaro Winkler uses a combination of rules and methods (including Levenshtein) designed and determined based on experience and knowledge gained from long-term data matching studies done by the US Census Bureau. Jaro Winkler is unique (from Levenshtein) in that it puts added emphasis on the *beginning* of the word, rather than the end — making the assumption that a typist is likely to be more focused (therefore accurate) at the beginning of the word, than the end. Additionally, Jaro Winkler incorporates a list of 36 similar character pairs programmed to account for letters easily mistaken for one another. For example, "Henry" and "Hemry" would have a high "similarity" score because "m" and "n" is considered to be a "similar pair." Another incorporated rule provides that mistakes in names of five letters or more be scored lower, as names of greater length experience increased probability for typographical errors. Programming functions for Jaro Winkler are easily found on open source sites such as *StackOverflow.com* and *Github.com*. An example of an optimized SQL implementation can be found on the internet at: *http://stackoverflow.com/questions/16448297/fuzzy-grouping-in-sql* .

3 Nickname Tables: Even more difficult than "name spelling" variations are "nicknames." *"Why is this?"* one may ask. Simply, and unfortunately for your match engine, nicknames often have no relationship to the individual's actual legal name. Nicknames can be shortened versions of a name such as "Liz" for "Elizabeth," a pseudonym such as "Bob" for "Robert," a variation of a person's last name such as "Mic" for "McCarthy," or even after a physical or behavioral trait such as "Ginger" for someone with red hair, or "Tank" for a powerfully-built second-grader. Additionally, many times common nicknames are also used as legal names. For example, there are 42 accepted derivatives of the name Amelia that are used as both nicknames for those named Amelia, and legal names on their own (e.g., Mia). To address these issues which may confound your matching efforts, there are several potential strategies available for inclusion into your system. Two popular and effective techniques include:

- If nicknames are stored in the Student Information System (i.e. collected as a separate data attribute), they can be incorporated into the match process (in place of the legal name) and compared against the benefits file on a dedicated pass.

- Utilize open source nickname tables. For example, the Old Dominion University - Web Science and Digital Libraries Research Group created an open source Nickname and Diminuitive Name Lookup which can be found at: https://github.com/carltonnorthern/nickname-and-diminutive-names-lookup. The table contains a list of ~1,600 U.S. first names and their associated nicknames or diminutive names. There are Java, Perl and Python parsers provided.

CONTACT INFORMATION

Operational Support Branch
Child Nutrition Programs
USDA Food & Nutrition Service
CNStateSystems@fns.usda.gov

If you would like to submit a promising practice or lesson learned for possible inclusion in a later edition, please send suggestions to CNStateSystems@fns.usda.gov.

Pilot Expansion to Further Test DC Medicaid

FNS recently announced the selection of **15 States** (*see Table below*) to participate in new demonstrations to evaluate the ability and effectiveness to connect kids to both free and reduced price school meals through direct certification utilizing Medicaid participant data. Although FNS has conducted similar demonstrations since SY 2012-2013 to evaluate DCM for free meal eligibility, the new demonstrations will expand direct certification activities to also include reduced price meals.

As with other matching and direct certification efforts incorporating other means-tested programs, using Medicaid data for this purpose has the potential to improve student access to healthy school meals, reduce administrative burden for schools, and improve certification accuracy – as well as increasing an LEA's or school's identified student percentage for participation in the Community Eligibility Provision. FNS will work closely with the selected States to evaluate the impact of the pilots, share best practices, and provide technical assistance — and M2M looks forward to highlighting State successes and learning experiences in upcoming issues.

FNS announced a Request for Applications (SP 23-2016) to participate in the demonstrations in January 2016. Seven States were selected to begin the pilots in SY 2016-2017 and eight were selected to begin in SY 2017-2018.

DCM for Free & Reduced Price Pilot States Beginning SY 2016-2017	DCM for Free & Reduced Price Pilot States Beginning SY 2017-2018	DCM for Free Meals only Pilot States since SY 2012-2013
California*	Connecticut	Illinois
Florida*	Indiana	Kentucky
Massachusetts*	Iowa	New York
Nebraska	Michigan	Pennsylvania
Utah	Nevada	
Virginia	Texas	
West Virginia	Washington	
	Wisconsin	

*California, Florida and Massachusetts transitioned from previous DCM pilots for free meals only

New York State Demos Exciting DC Improvements

With the goal of improving eligibility determinations and overall direct certification system effectiveness and efficiencies, NYSED has undertaken an aggressive project to move from a local-level match approach, to a State-level match process. The new system will provide several improvements and opportunities for New York LEAs — including a highly effective record-matching software solution. Be sure to watch for a future M2M in-depth story on NYSED's project and associated achievements and lessons learned.

Key members of New York's DC System Improvement Team display their favorite issues of M2M. From Left to Right: Tom Myers, New York State Technology Enterprise Corporation; Sandra Hadley, NYSED CNP IT; Jennifer Knapp, NYSED CNP; and Brian Hart, CGI.