



Direct Certification 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.

## DEVELOPING MATCH ALGORITHMS

A match algorithm is how a computer looks at two records and determines if they are describing the same person. When the comparable fields in both records match exactly, it makes it 'easy' for the computer to decide that it is the same person. However, when certain parts of the record do not match, that decision becomes more difficult.

Defining the rules and methods that give the computer insight is frequently known as "developing the match algorithm." It is one of the key areas in meeting the legislatively mandated benchmark to directly certify 95 percent of children in families receiving SNAP benefits. It's reasonable to predict that if a State's match algorithm does not match records at a rate of 95 percent or greater, it will be nearly impossible for the direct certification rate to meet the benchmark. Simply put, a low match rate will translate into a low direct certification rate.

*"So this sounds important. What should I consider when thinking about this...?"* When it comes to developing the match algorithm, it's important to remember that one size does not, indeed cannot, fit all. It is not about the specific technology or rules used; but

*Continued on page 2...*

### INSIDE THIS ISSUE

*USDA is an equal opportunity  
provider and employer.*

### Page 2

USDA News/Announcements  
Success Stories

### Page 3

Selection of Database Fields  
for Matching

### Page 4

Probabilistic Matching  
Explained

## USDA NEWS/ ANNOUNCEMENTS

### Outstanding Performance Awards

The USDA Food and Nutrition Service announced the 13 recipients of the 2014 Direct Certification Performance Awards. State agencies acknowledged under the *Outstanding Performance* category are Illinois, Florida, Maryland, Minnesota, North Carolina, and Oklahoma. State agencies acknowledged under the *Substantial Improvement* category are Arkansas, Guam, Massachusetts, Missouri, New Hampshire, New Jersey, and Pennsylvania.

The Performance Awards were established by the Healthy, Hunger-Free Kids Act of 2010 (HHFKA) to encourage State agencies that administer the NSLP to ensure that all children participating in SNAP are directly certified for free meals without application. Awards were based upon direct certification data available for SY 2013-2014. Recipients will share amongst the \$2 million in funds made available under each category, and must make use of funds to support USDA school meals programs.

*Continued from page 1...*

rather the *process* in creating the match algorithm.

This is true because match algorithms overcome specific problems in source data and work differently for specific kinds of demographics. Each step in the direct certification process is very important. However, the entire process is dependent upon the ability to obtain and process source data to enable the generation of accurate matches of children eligible for free meals. Therefore, researching the best match algorithm for your State is a process.

When initiating this process, States should take the following necessary steps:

1. Research cause of non-matches.
2. Determine appropriate strategies to overcome.
3. Run tests in test environment to determine the effectiveness of the algorithm.
4. Ensure overcoming one set of false negatives does not produce a set of false positives.
5. Deploy and monitor, build for ongoing improvement.

States that have followed these steps have seen the greatest improvement in their matches.

## SUCCESS STORIES

Although direct certification systems with coinciding matching methods may vary across States, a successful match process requires a high-quality matching algorithm. Arkansas and Montana both employed an extensive research process that was tailored to their specific State needs.

### ARKANSAS

- Determined that a resource in the research department could provide in-house IT expertise
- Researched the causes and potential causes of non-matched records
- Developed a strategy to overcome problems specific to Arkansas, involving a 22-step automated process that made use of:
  - Tailored phonetic matches
  - Custom string matches
  - Nickname and not-nickname tables
- Match rate jumped from 81 to 93 in the span of a few months and the State continues to refine and improve processes

### MONTANA

- Utilizing a COTS rules engine, Montana developed custom match rules while developing a new portal based on a Microsoft Dynamics CRM platform
- Montana has generated a weighted matching algorithm that provides match scores on four data fields:
  - Last Name
  - First Name
  - Date of Birth
  - Gender
- Once two records are determined to be a match, the unique ID from each agency is joined in a lookup table (i.e., bridge file) so records will remain tied for future use, eliminating the need for repeat manual adjudication

**If your State needs guidance addressing direct certification challenges or has a success story that USDA can share, please let us know! We encourage you to contact us for assistance or to share a recent accomplishment.**

# SELECTION OF DATABASE FIELDS FOR MATCHING

## The Basics

The linchpin of direct certification involves matching data, or attributes about a real-world child, in electronic form. In essence, each attribute, field, or data element, describes a single characteristic associated with that child, e.g., first name, last name, date of birth (DOB).

While databases are created for different purposes, and therefore contain different attributes, it is helpful to think of the attributes, or fields available to match on, along a three-part continuum:

### 1 Primary Match Attributes

Primary match attributes for a child are the best descriptors of the real-world child; these attributes are among the highest quality and most stable descriptors found in both databases. They frequently include attributes such as first name, last name, and DOB. They are the highest quality descriptors because they are commonly used in real-world interactions as core characteristics of the person's identity and less likely to change than other attributes (e.g., address).

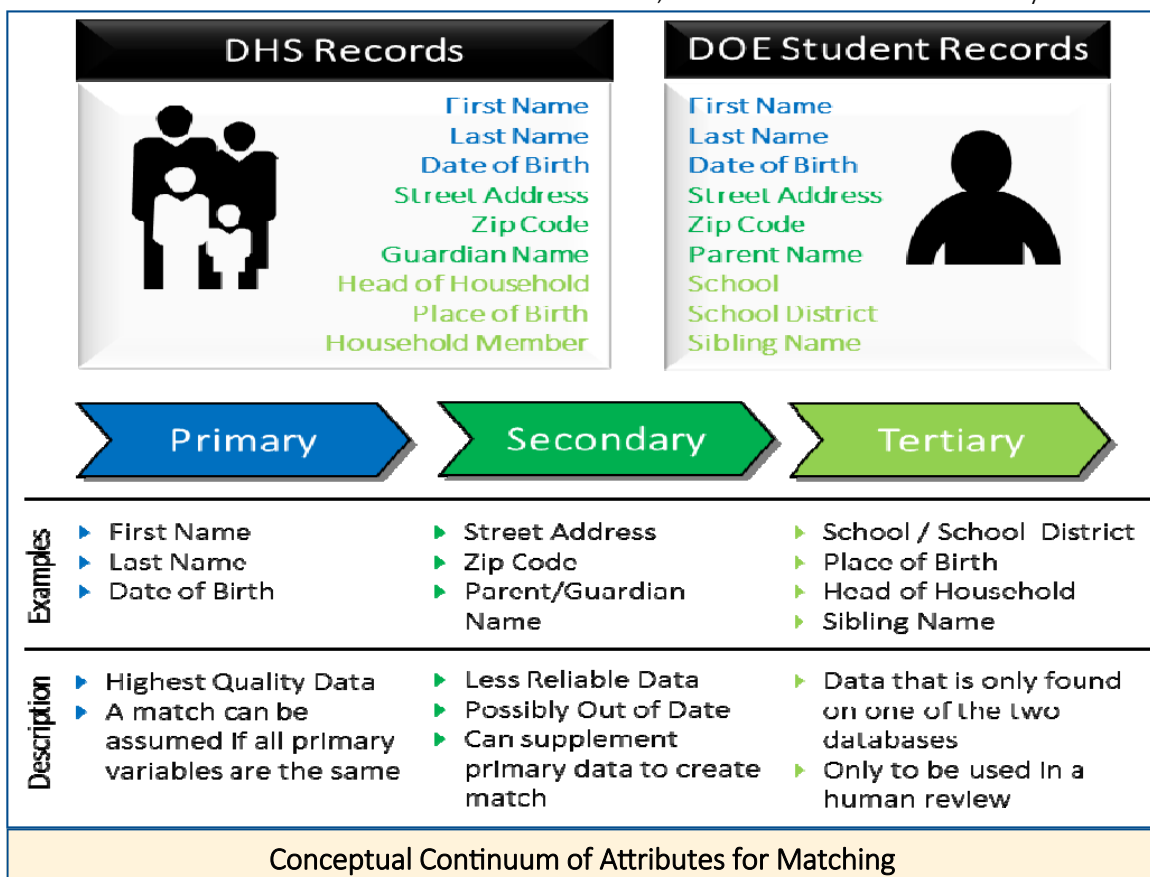
### 2 Secondary Match Attributes

Secondary match attributes are less stable for a child (e.g., address), may not receive as much attention to ensure accuracy when entered (e.g., correct legal name of guardian), or may not be very valuable in differentiating records (e.g., gender).

Consequently, these attributes should be considered only to augment confidence that two records are describing the same real-world child, particularly if a perfect match of the primary attributes for that child is not possible because of spelling errors, date format errors, etc. Examples of secondary attributes are address and parent name. These attributes are secondary for several reasons. Addresses change (sometimes frequently), and a parent's name may also change as a result of marriage, divorce, or custody issues. These changes may not be recorded in the database(s) in a timely manner.

### 3 Tertiary Match Attributes

Tertiary match attributes for a child are typically those attributes collected by one database, but not included in others. These may help confirm a match during a human review when primary and secondary attributes are insufficient to make a solid match. These attributes may include the school the child currently attends, previous address, or other members of the family.



### PROBABILISTIC MATCHING EXPLAINED

Identifying potential matches is a key component of a highly efficient direct certification process, but with all the different ways of determining what a 'potential match' is, it can be helpful to think of it along a continuum.

The following continuum provides a way to think about and categorize 'how probable' it is that two records describe the same child. Automation allows the creation of rules by which a computer can use to classify match results, comparing and combining multiple data attributes for a child:

**1 Exact match:** attributes are recorded exactly the same

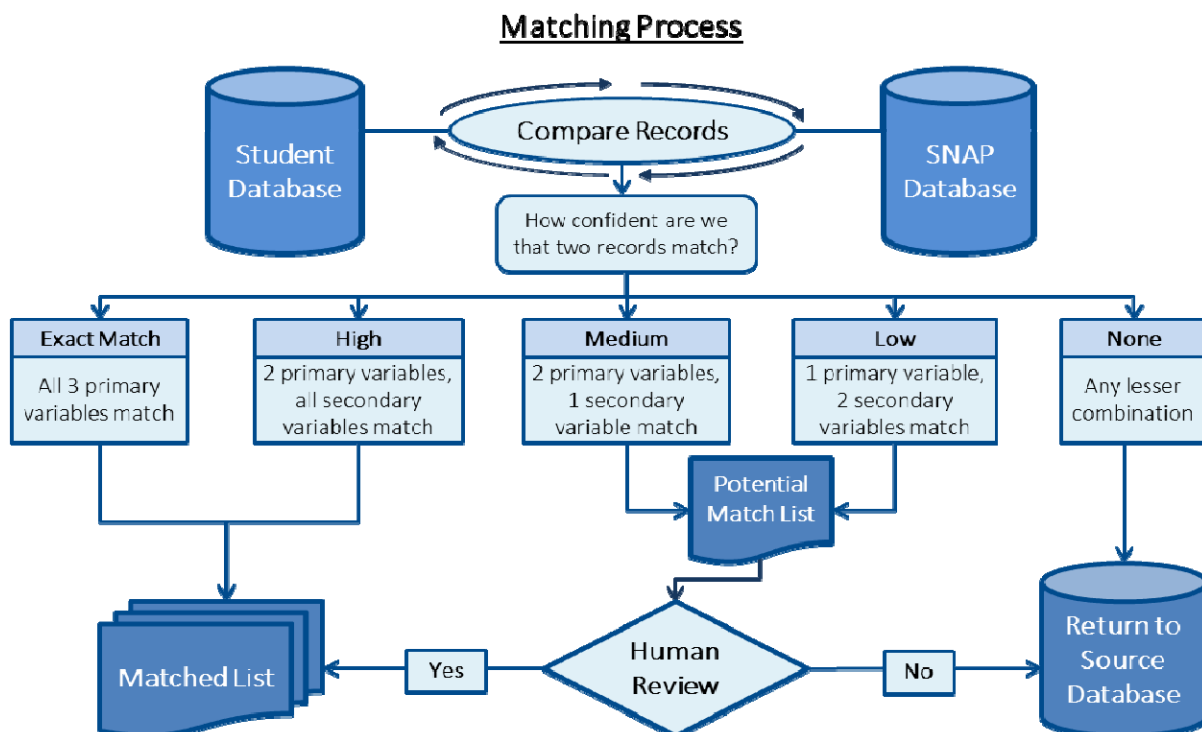
**2 High confidence:** attributes match so closely that review to confirm that they are describing the same child is not necessary

**3 Medium confidence:** attributes do not match well enough to avoid review, but confidence is high enough to justify human review to try to confirm the match.

**4 Low confidence:** attributes do not match well enough to avoid review, but the misalignment is such that review should only be attempted if and when resources permit

**5 No match:** comparison records for the child are not found

While not every direct certification system uses all of these categories, the following graphic represents what they might look like when used in a matching process:



### CONTACT INFORMATION

Operational Support Branch  
 Child Nutrition Programs  
 USDA Food & Nutrition Service  
[CNStateSystems@fns.usda.gov](mailto: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](mailto:CNStateSystems@fns.usda.gov).