***Current CommIT Matching Rule:***

* The CommIT algorithm currently does an exact (case-insensitive) match of first name, last name, city and DOB (full or partial). They all need to be equal to declare a match.

***Penn State currently follows or is considering the following for their matching algorithm:***

* A match is made if a userid, psuid, or ssn is passed into CIDR...if none of this data is present a match is attempted on the data that is presented.
* The grid is used to determine a score if there is a match on the presented data....matches are only made above a certain score (330?)
* Grad School uses the international grid; everyone else uses the standard grid.
* Required fields for matching are different between international and US.
* non-Roman alphabet issue will be accommodated by the match code software (maybe)
* match codes currently produced by the vendor software are 15 characters....it is generated by the match code software after data normalization and is stored with the person record
* match code value changes if any data used to generate match codes changes
* algorithm is used to create grid ranking....application program uses ranking to make decision about match....it is easy for an application to decide about really good and really bad matches.....the "maybe OK" matches are harder for a program to decide on....this grey area is probably the source for matching errors.
* there are occasions when matches are made were the name and address are not even similar.....this may be because the PSUid is not forced to be unique or the match was made on an incorrectly supplied userid, PSUid, or SSN
* different application programs utilize different matching algorithms....centralizing the matching function in the CPR may help resolve current issues
* John A Smith and John B Smith get the same match code because match code sensitivity is set to 85....if the sensitivity setting was set to 100 this would not occur, however this would have ramifications on other data fields where less sensitivity is desirable

**CommIT Data Matching – Questions and Considerations**

1. Do we want to use PSU’s algorithm or develop our own?
   1. Standard Grid
   2. International Grid
   3. Hybrid
2. What will the tolerance level be for CommIT?
3. Do we want to integrate fuzzy matching into the Pilot?
4. Review and Reconciliation are necessary if:
   1. two or more accounts are suspected to be owned by the same person
   2. two or more people lay claim to the same account
5. What steps need to be followed for the review/reconciliation?
   1. Notification to affected user(s)
   2. Notification to Pilot Participants
      1. Use the CommIT logs to identify which participants have a relationship with affected user(s)
   3. Record Merging
   4. Record un-Merging or Splitting
   5. Is Account Linking (multiple CommIT ID’s associated with a single user) a possibility for the Pilot?
6. Who will perform these tasks?