

Minnesota Health Care Claims Reporting System: Provider Identification and Linkage

Kevan Edwards, PhD
Health Services Research Director

Chelsea Georgesen, MPH
Project Coordinator

Minnesota Department of Health

Webinar Agenda

MDH will review a number of concepts related to our APCD and provider ID and consolidation including:

1. Background and future direction for Minnesota APCD for context of the discussion
2. How intended APCD use impacts both the need for and degree of provider identification necessary for project success
3. Provider identification and consolidation process in MN
 - Two steps
 - Vendor provided identification and consolidation
 - MDH directed effort
4. Impact of enhanced identification/linkage on our APCD

Overview of the MN APCD

- Authorizing legislation in 2008 under state health reform
- Up to now, authorizing language limits use to one purpose
 - Cost and Quality Reporting effort referred to as the Provider Peer Grouping Project (PPG)
 - New legislation signed into law in April 2014 expands the authorized uses of the APCD and suspends PPG
- Collection began in 2009
 - Data includes claims from 2008 - present
- MDH contracts with Onpoint Health Data for data collections and aggregation services
- Currently covers ~ 90% of MN covered lives approx 4.2 million individuals and 78,000,000 medical claims

Planned Use of APCD & Impact on Provider Identification

As in all decisions regarding your APCD, ask yourself what are we planning to do with the data?

- Legislative restrictions or CMS reporting restrictions
- Political climate and what is “acceptable” use in your environment
- Does my planned / authorized use warrant investment in provider ID effort or will it become necessary in the future ?

Research usually requiring minimal provider identification effort

- Epidemiological studies
- Small area variation studies (Dartmouth Atlas type reporting)
- Disease burden

Research usually requiring maximum provider identification effort

- Care system cost reporting (either total cost of care or condition specific)
- Facility identified cost reporting (TCOC or Condition Specific)
- Individual provider cost / quality profiling efforts

Provider Identification and Linkage

Step 1: Vendor Provider Identification and Consolidation

MN Provider Identification and Linkage

- MDH uses a two-step approach to provider identification and linkage
 1. Work performed by the vendor to uniquely identify providers then combine providers across data submitters when provider data matches on certain key id data fields
 2. The value-added work that MDH completes to enhance the data and provider linkage effort to combine the same provider who may falsely appear as different unique providers
- It is critical to work with a vendor partner capable of providing the initial provider ID and linkage service that enables additional value-added work later

Data Issues

- Data from multiple sources is often misaligned
- Submitted data is of varying quality and accuracy, especially across different payers
- Conflicting data exists due to entry errors, delivery system complexity, name changes and multiple practice locations in-state and out-of-state
- Necessary provider detail such as registration information (ex. NPI values) is not as functional and clean as initially hoped / intended

Necessary Data Elements

- Legacy provider number
- Tax ID (collected for non-individuals)
- NPI
- Provider Entity
- First name, middle initial, last name, suffix
- City, state and zip code

Note: There are varying rates of completeness and quality across these elements, including payers who submit an individual's entire name in the last name field. It is also difficult to determine whether location is reported as the billing address rather than the service or facility location

Creation of Unique Identifiers

- Assignment of a “*unique ID*”
 - Allows records to be assigned to one individual, regardless of where the data originated
- Continuous mapping of new and existing data to established IDs
 - Process must identify recognized data versus unrecognized data, and determine if a link can be established
- Auto- and manual-clustering of records creates a more authoritative and smaller file of provider IDs
 - Extensive manual review is required to refine the linkages

Next Steps: Further Refining “Unique IDs”

- At this point, an individual may still be assigned multiple “unique IDs”
 - Dr. John Smith and Dr. John M. Smith may be the same person, but are assigned two different unique IDs based on a small difference in the data
- Additional steps are needed to further refine these links and definitely identify John Smith (unique ID#1234) and John M. Smith (unique ID#2345) as the same person
- MDH’s value-added steps examine these links at a more nuanced and granular level

Step 2: MDH's Value-Added Work

Step 2 – Consolidation & Linkage to External Data

- Purpose of this step to use a probabilistic matching methodology to combine different “unique provider IDs” that very likely point to the same individual
- Like distinct puzzle pieces the goal is to fit these separate pieces into whole entities again



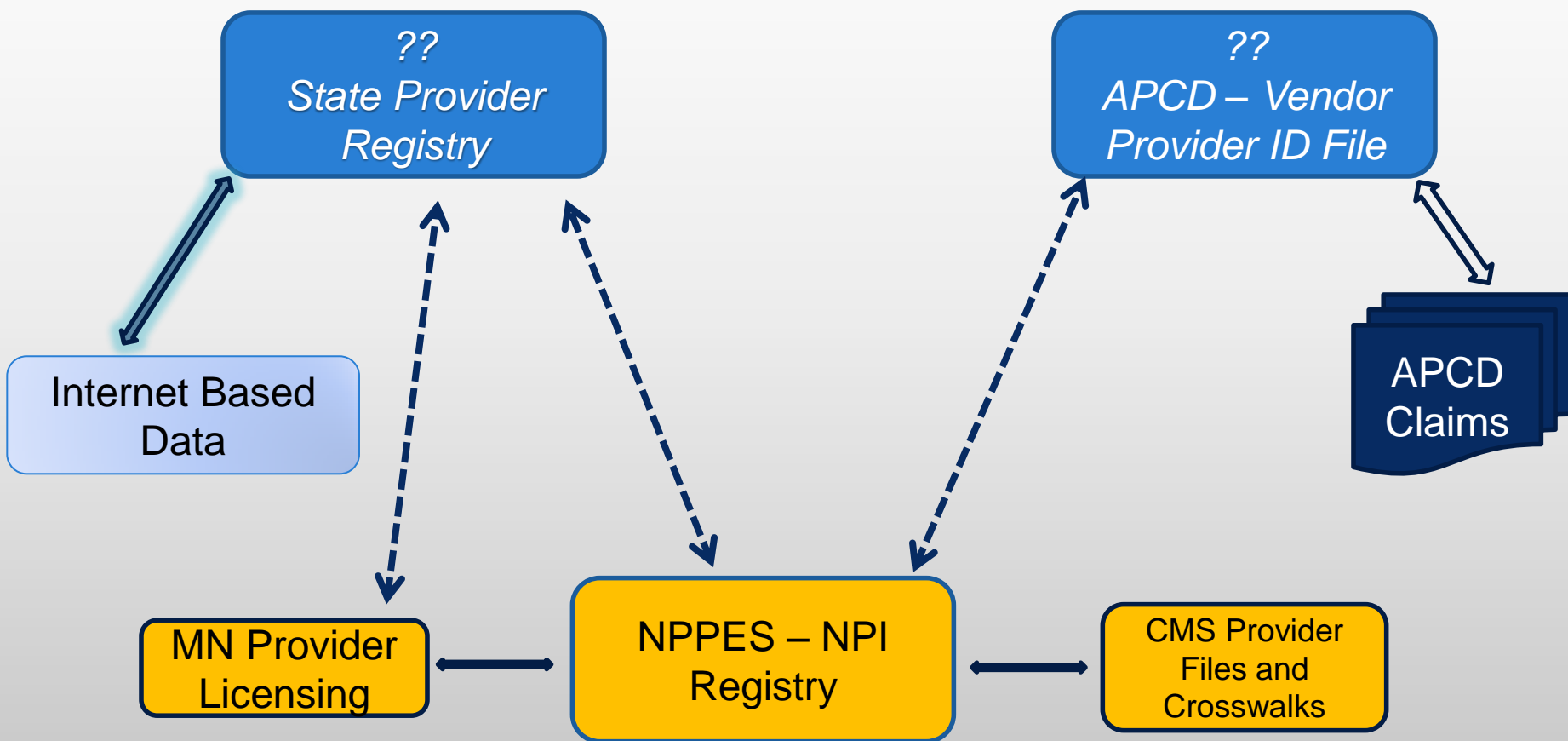
Necessary Data Sources

Critically important data sources

1. **MN Provider Registry Database – collected by state contract pursuant to our quality reporting rules and statute**
2. NPES Registry – downloaded from CMS via http://npes.viva-it.com/NPI_Files.html
3. A summary file provided by the Vendor – listing all distinct combinations of identifying provider information from the APCD

Additional “Bonus” supplemental data sources

1. State licensure files,
2. Other ID files from outside sources (eg: CMS MPIER, CMS UPIN Directory, HCCIS, Discharge Data) etc.
3. Current Internet Provider Web Sites
4. Prior Internet Provider Web Sites (the way back machine).



Task 1 - Clean The Provider Registry File

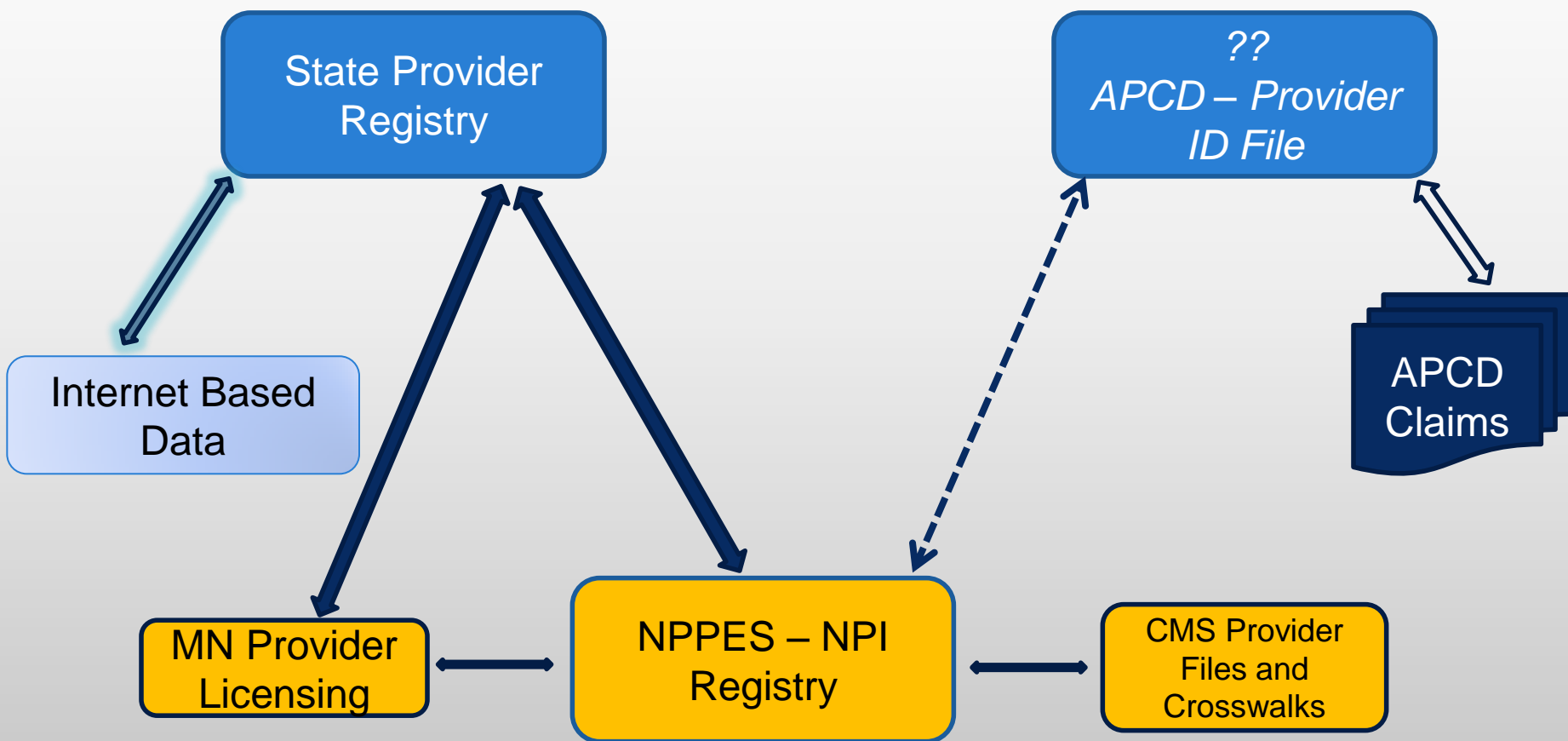
1. Since the provider registry data does not come to us completely clean we have to validate Provider information (NPIs) to use this file

NOTE This step is crucial, as the NPI is the key to linking providers to the claims data.

- Procedurally validate by comparing the registry data to the NPPES data registry, linking on various combinations of:
 - NPI + Last Name + First Name
 - NPI + Last Name + First Initial + ?
 - NPI + Last Name + ?
- Auto validate and when necessary manually correct NPIs using various tools
 - NPI Lookup tools, Provider Websites,
 - <http://archive.org/web/>
 - These manual “look ups” make great summer intern projects

Task 1 Continued - Evaluate, Confirm, Finalize

- Analyze the auto-corrected NPIs in the registry
- Manually correct to ensure false positives are NOT introduced in later matching and identification steps
 - example - the same NPIs is assigned to providers with differing first and last names (i.e., sharing an NPI in our registry)
- Compare enhanced provider registry to prior or more recent versions
 - look for provider gaps – i.e., providers who failed to register at points in time.
 - example - John Smith appears at Clinic A in 2009 and 2011 but does not appear in the registry at all in 2010 (and also has claims in the APCD in 2010)
- Develop business decision rules to finish the data processing
 - In the John Smith example we decide to connect two book ended identical entries and manually add him to the registry as staff under Clinic A
- Append supplemental data IDs (license numbers, etc.) to registry



Task 2: Vendor IDs and Data Elements - Enhanced Clustering Analysis

1. Identify rules that would indicate that two or more unique IDs provided by your vendor may actually be the same provider.
2. Develop business rule confidence ratings to that list of rules (note how ID from the Data Vendor is NOT included in the clustering rules)

Rule #	Match Criteria	Confidence (10 high/1 low)
1	LName, MName, FName, NPI, TaxID, Zip, Suffix where all are Not NULL	10
2	LName, FName, Mname, NPI, Zip where all are Not NULL	8
3	LName, FName, TaxID, Zip where all are Not Null	5
4	Etc.	

3. Run business rules against the provider data to create “Match Clusters”
4. Merge resulting clusters that share vendor provided IDs by determining business rules that will trigger merging Ids – balance between rigor and volume

Minimum Confidence for IDs linked clusters*	Minimum # of distinct IDs in common
10	1
9	3
8	5
7	7
<= 6	Do Not Merge

Task 2 Continued - Cluster Analysis

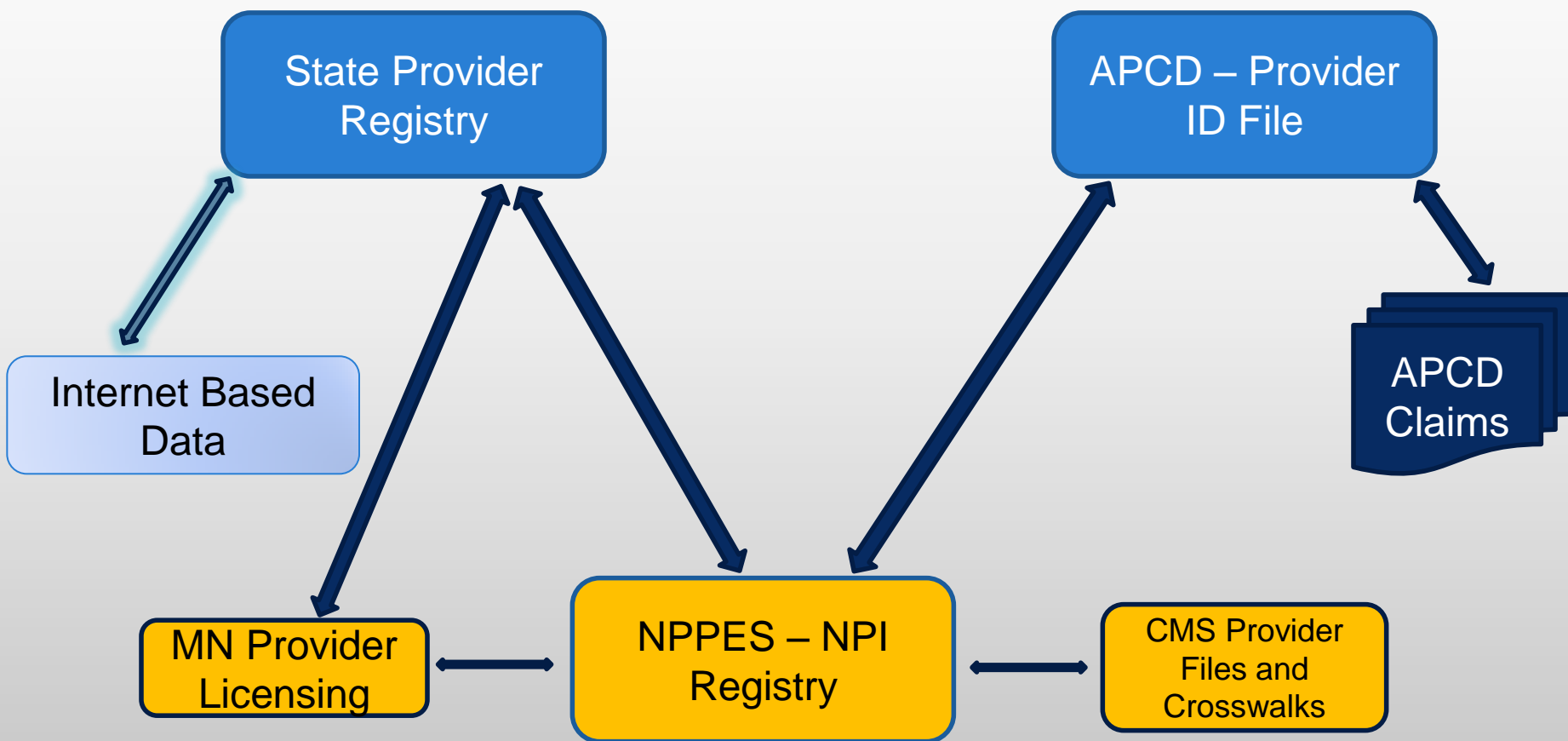
5. Perform manual analysis on cluster merges
 - ensure no incorrect merges are occurring by ensuring they do not merge
 - tweak confidence rankings or minimum common ID criteria
 - develop a “Do Not Merge” table
6. Continue merging clusters according to your rules until no more merges exist (iterative approach)
7. Post merge ensure IDs exist in only 1 merged cluster
 - If not, we keep the IDN in the cluster to which it matched with the highest confidence and delete from other clusters
8. After deletion of duplicate IDs identify IDs that are now part of a single ID cluster (i.e., itself).
 - Analyze that IDN’s lower confidence links that were deleted to see if any of them link to a cluster that still survives.
 - Reassign IDN to the cluster to which it is linked with the highest confidence.

Task 2 Continued - Post-Cluster APCD Analysis Verification

- Perform additional analysis to ensure incorrect links are not occurring
- Example: Annual consistency ?
 - Look for Provider IDs that were assigned different **Modal NPIs** from one year to the next year
 - We have noted this usually occurs when the identifying information associated with that Provider ID is vague in nature.

Ex. J. J. Smith linked to John James Smith in 2009, but to James Jordan Smith in 2010

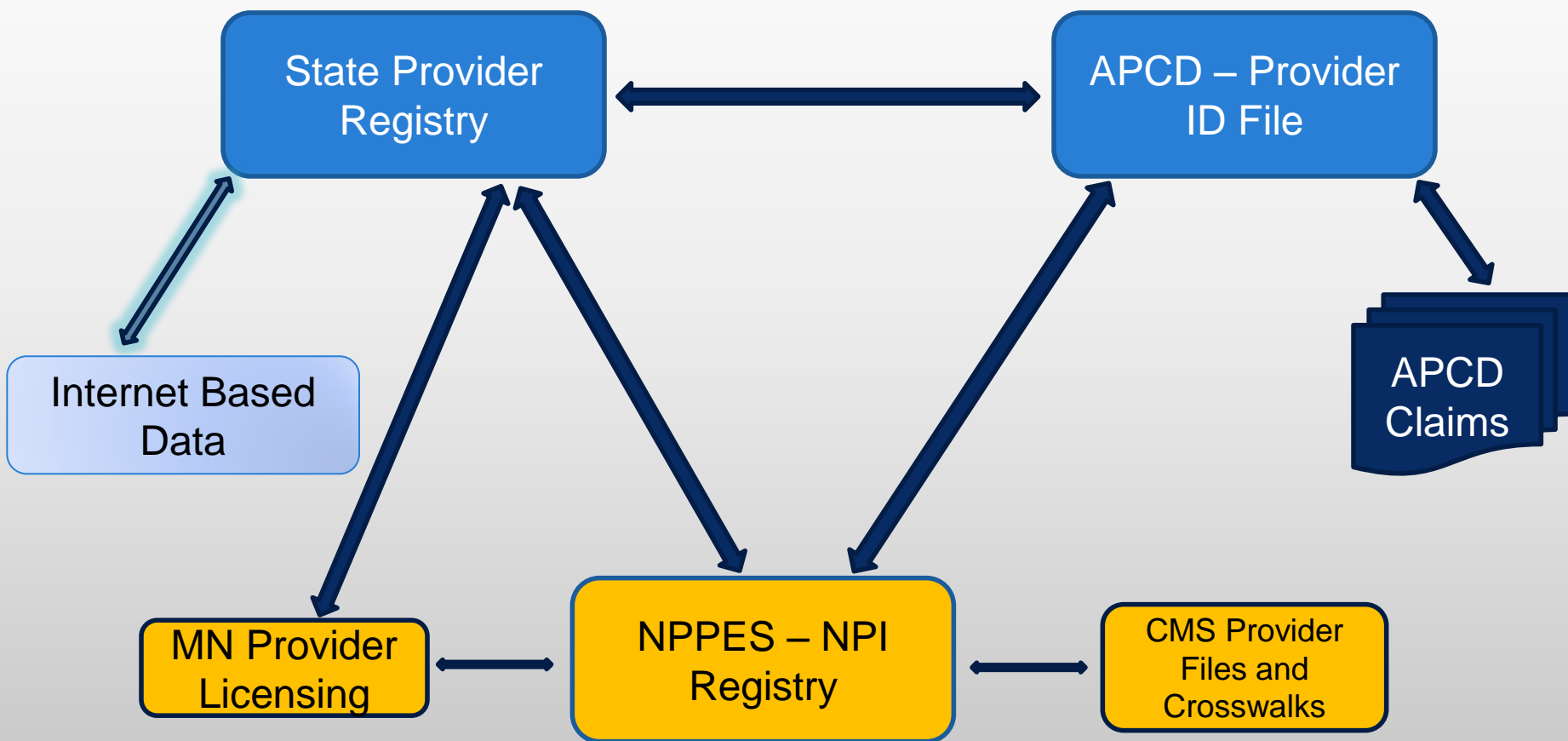
- Analysis has shown that this problem amounts to less than 0.1% of total claims and is therefore best left as is



Task 3 – Merge APCD Provider / Registry Provider Tables

1. Determine “Modal” value for the clusters in your Merged Provider ID file created Task 2
 - Find the value that occurs most often within the cluster
 - Require a minimum percentage of the occurrences for the modal value to be used confidently
 - Assign the “modal” value to represent the cluster on the following data fields
 - Provider First , Middle and Last Name
 - **NPI** & TaxID
 - City, State & Zip

2. Link Provider APCD Data Summary Modal NPI to the Provider Registry Verified NPI



Finally Verify and Analyze the Results

- Compare modal categories of the resulting merges.
 - Do modal names from the claims ID match the names in provider registry?

- Using the merged Ids
 - analyze % of visits in your APCD that can be assigned to confidently identified / consolidated providers
 - analyze % of dollars assigned to confidently identified / consolidated providers

- Examine changes in both volume of **visits** and **dollars** that can be assigned to providers annually for extreme variation

- If anomalies can't be reasonably explained options include:
 - Tweaking business rules re confidence intervals and / or merge logic
 - Manually merging or preventing merging of particular IDNs
 - Other ideas that may be appropriate for you data and needs

Effects of Enhanced Linkage

Statistics from 2011 APCD data (prior to enhanced linkage)

- Number of distinct unique provider identifiers = 656,885
- \$4,789,000,923 Paid in Evaluation & Management Paid Claims

After MDH Enhanced Linkage

- Number of MDH-generated Unique IDs = 16,124
- \$3,331,079,334 in Evaluation & Management Paid Claims
 - This represents nearly 70% of E&M Paid Claims linked to comprehensively identified Minnesota registered providers and roughly 30% linked to either 1) non MN providers, 2) providers NOT in the MN provider registry or 3) providers for whom enhanced linkage efforts failed to improve consolidation

Contact & Additional Information

- Kevan Edwards 651-201-3551
kevan.edwards@state.mn.us
- Chelsea Georgesen 651-201-5957
chelsea.georgesen@state.mn.us
- PPG Hospital Total Care Homepage
www.health.state.mn.us/healthreform/peer/hospitaltotalcare.html