Best Practices in *Data Quality*

Moderator: **Linda Green**, *Vice President, Programs*
Freedman Healthcare

Provider Identification and Linkage:

Kevan Edwards, *Health Services Research Director*, Minnesota

Validation Checks:

Mary Fields, *Business Systems Analyst*, New Hampshire

Historical Data Files:

Kyle Russell, *APCD Data Analyst*, Virginia

APCD COUNCIL All-Payer Claims Database SM

8th Annual APCD Workshop

Provider Identification and Linkage in All Payer Claims Data

Kevan Edwards, PhD
Health Services Research Director

Health Economics Program
Minnesota Department of Health

health reform MINNESOTA
A Better State of Health

Agenda

1. Shaping the question – Data Quality vs Functionality
2. Provider Identification challenges in APCD data
3. Provider identification and consolidation process in MN
    - Step 1: Vendor provided identification and consolidation
    - Step 2: MDH directed effort
4. Enhanced APCD quality through more robust provider identification/linkage

Data Quality vs Functionality

- Data Quality – focus on questions of validity and reliability in the data
  - Is submitted data accurate
  - Is submitted data complete
  - Is submitted data representative

- Provider Identity – a question of a different flavor
  - Data maybe perfect in its submission validity and reliability from each of your submitters **BUT**
  - Is it functionally enabled ?
  - Do you need to enable it ?
  - What do you have to do to improve the data and make it “more” usable

Provider Identification Issues Con't

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

- How you are using the data can guide your decision to commit resources to improving the quality of your database

Research requiring minimal provider identification effort

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

Research 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

Data Issues Related to Provider ID

- Data from multiple sources is often misaligned
- Submitted data can be of varying depth or completeness, 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

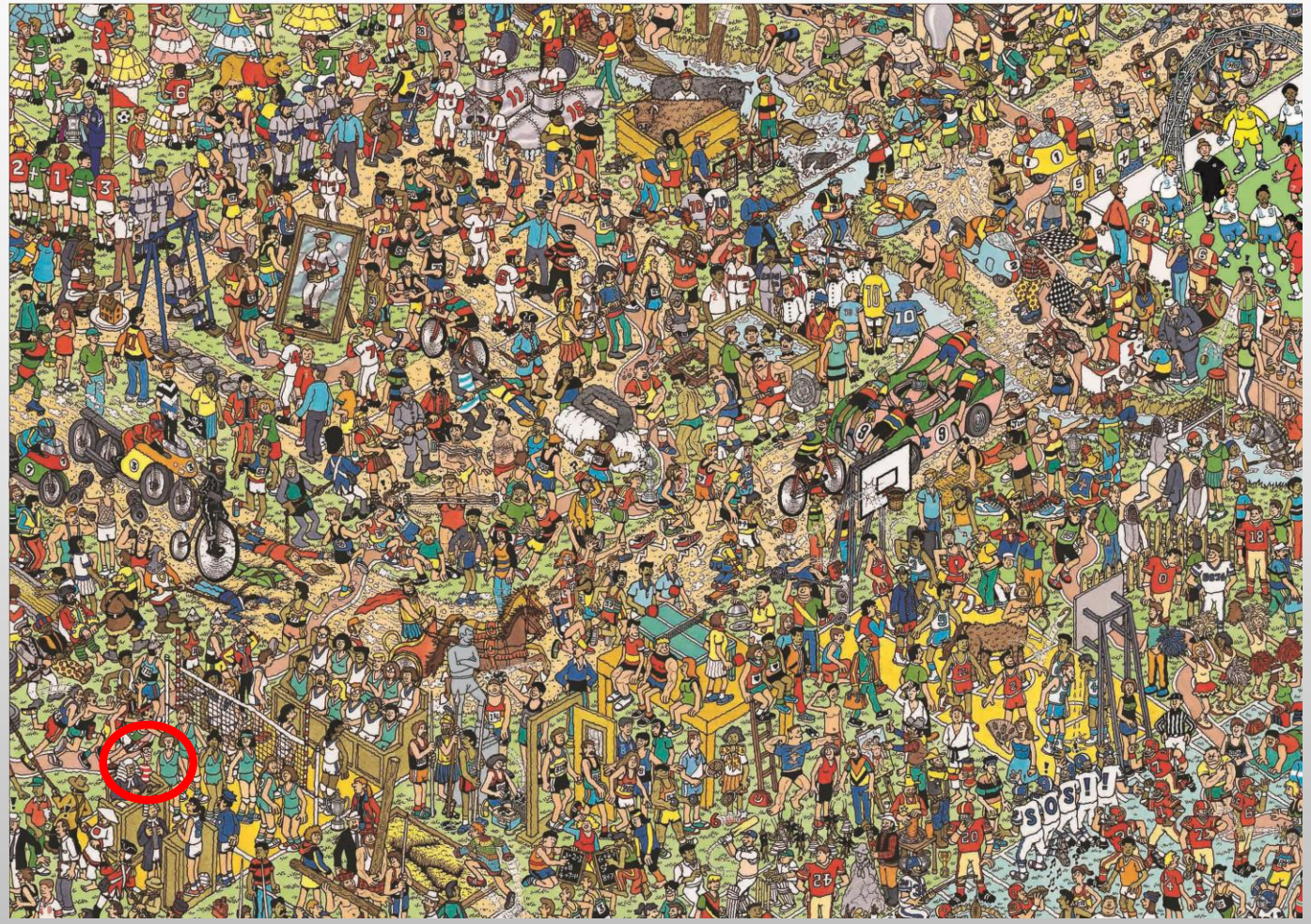
Enhancing Data: Provider Identification & Linkage

- Minnesota uses a two-step approach to provider identification and enhanced linkage
    1. Initial 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 data enhancement 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

Step 1

Initial Provider Identification & Consolidation

In a Nut Shell : Where is Dr. Waldo??



health reform MINNESOTA
A Better State of Health

Usually Available 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

Vendor Creation of Unique Identifiers

- Assignment of a "*unique ID*"
  - Allows identical records to be assigned to one individual, regardless of where the data originated
- Continuous mapping of new and existing data to established IDs
  - Identify recognized data versus unrecognized data
  - Determine if a link can be established
- Auto- and manual-clustering of records creates a more authoritative 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
- Depending on your state and vendor expect from 10X to 100X as many "unique" providers as you might expect.
- Additional steps needed to further refine links and unify the same provider uniquely identified more than once
- You can think of this as Where's Waldo 2.0

Next Task: Combining Dr. Waldo and all His Alter Egos

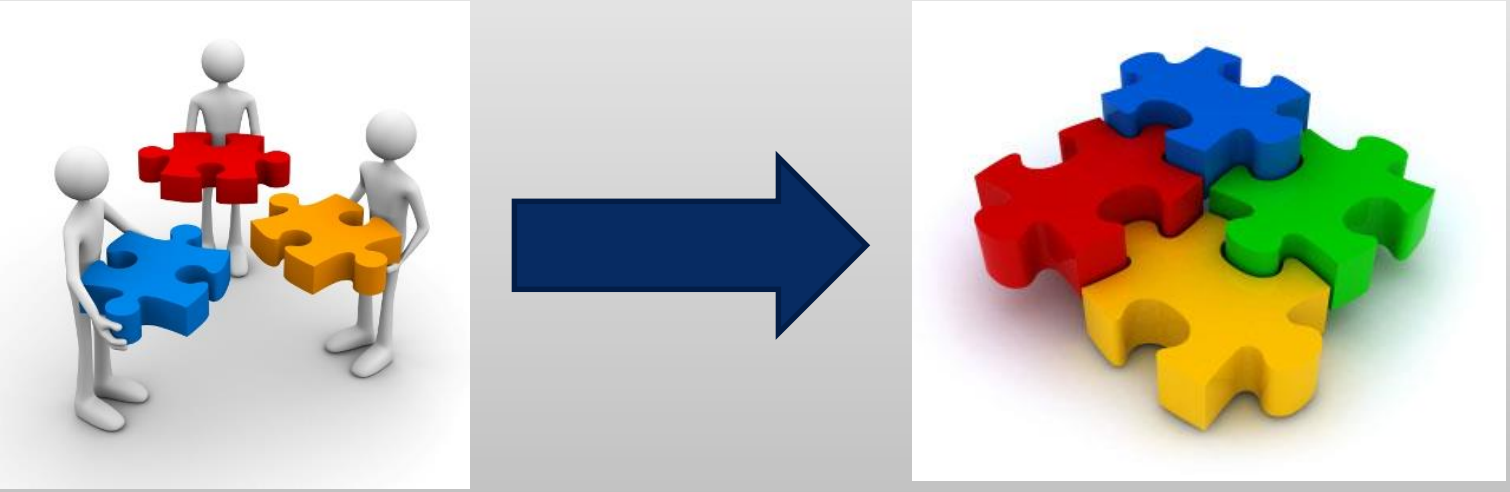
health reform MINNESOTA
A Better State of Health

Step 2: Bringing The Pieces Together

health reform MINNESOTA
A Better State of Health

Step 2 – Consolidation & Linkage to External Data

- Purpose of this step to use probabilistic matching and clustering techniques to combine different “unique provider IDs” that with high certainty point to one individual
- Like distinct puzzle pieces the goal is to fit these separate pieces into whole entities again



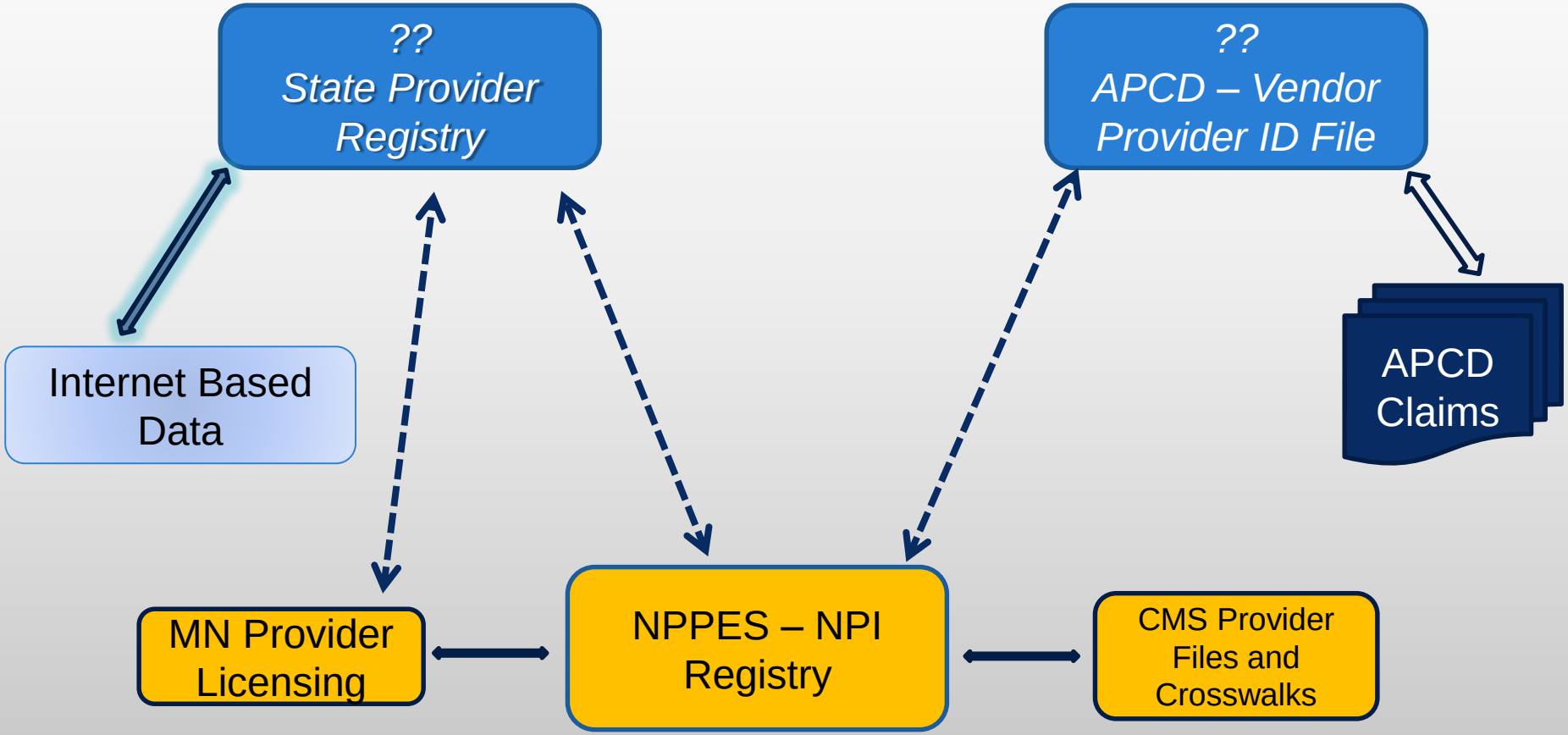
Bring in Outside Data Sources

Critically important data sources

1. **Provider Registry Database – collected by state contract pursuant to our quality reporting rules and statute**
2. NPPES Registry – downloaded from CMS via http://nppes.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 helpful 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).


?? State Provider Registry
?? APCD – Vendor Provider ID File
Internet Based Data
APCD Claims
MN Provider Licensing
NPPES – NPI Registry
CMS Provider Files and Crosswalks

Task 1 - Clean The Provider Registry File

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

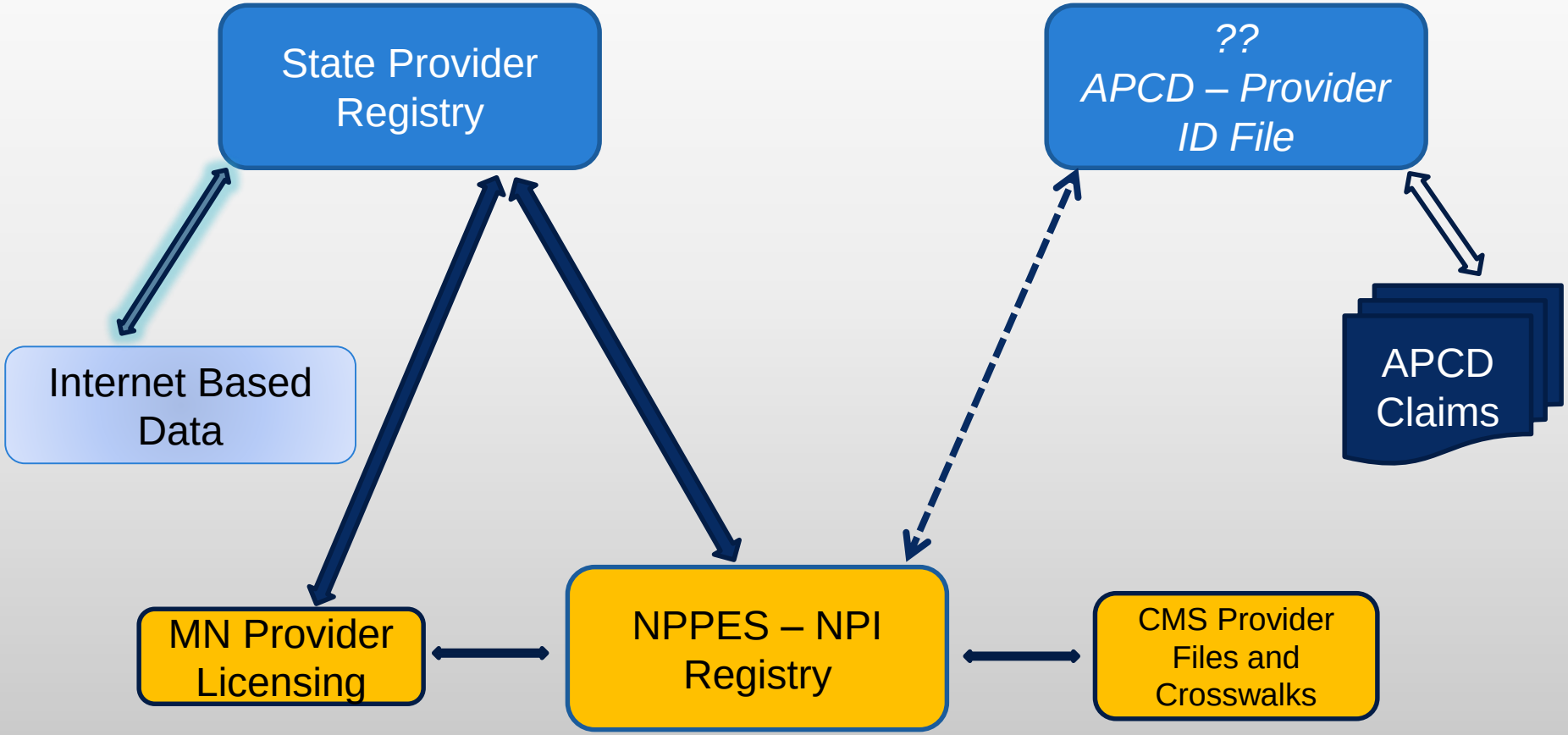
Validate and correct by comparing the registry data to NPPES data, licensure data, & web based resources linking on combinations of:

- NPI + Last Name + First Name
- NPI + Last Name + First Initial + etc.
- NPI + Last Name + etc.

Build Auto links as you go 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

health reform MINNESOTA
A Better State of Health


State Provider Registry
?? APCD – Provider ID File
Internet Based Data
APCD Claims
MN Provider Licensing
NPPES – NPI Registry
CMS Provider Files and Crosswalks

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

Develop business rules that indicate when two or more unique IDs from vendor may actually be the same provider and establish confidence ratings to those 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.	

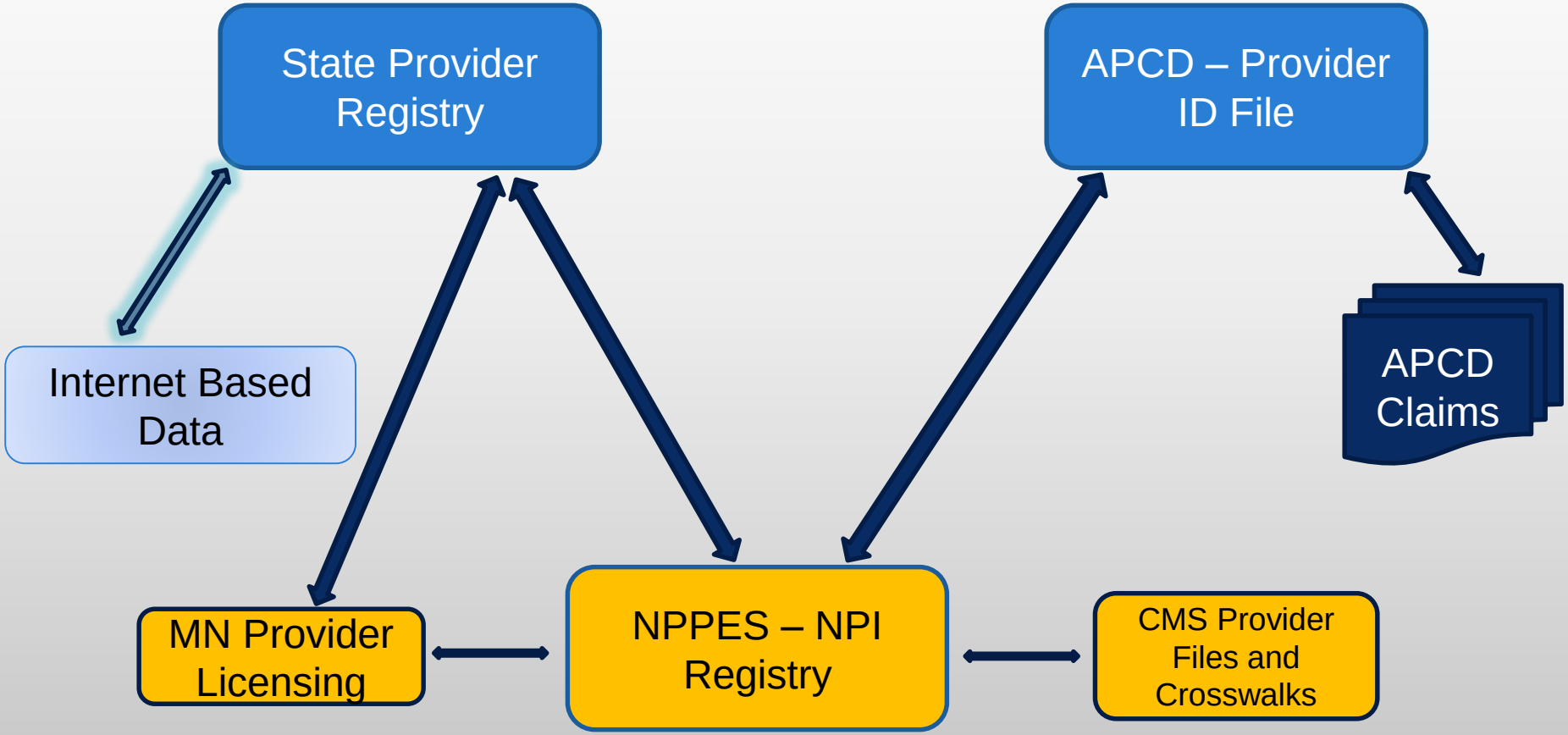
Run business rules against the provider data to create "Matched Clusters" of provider IDs

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

Merge resulting clusters using Modal data categories (NPI, Vendor ID etc.)

Both Manual and Automated verification to ensure against both false Neg and Pos

health reform MINNESOTA
A Better State of Health

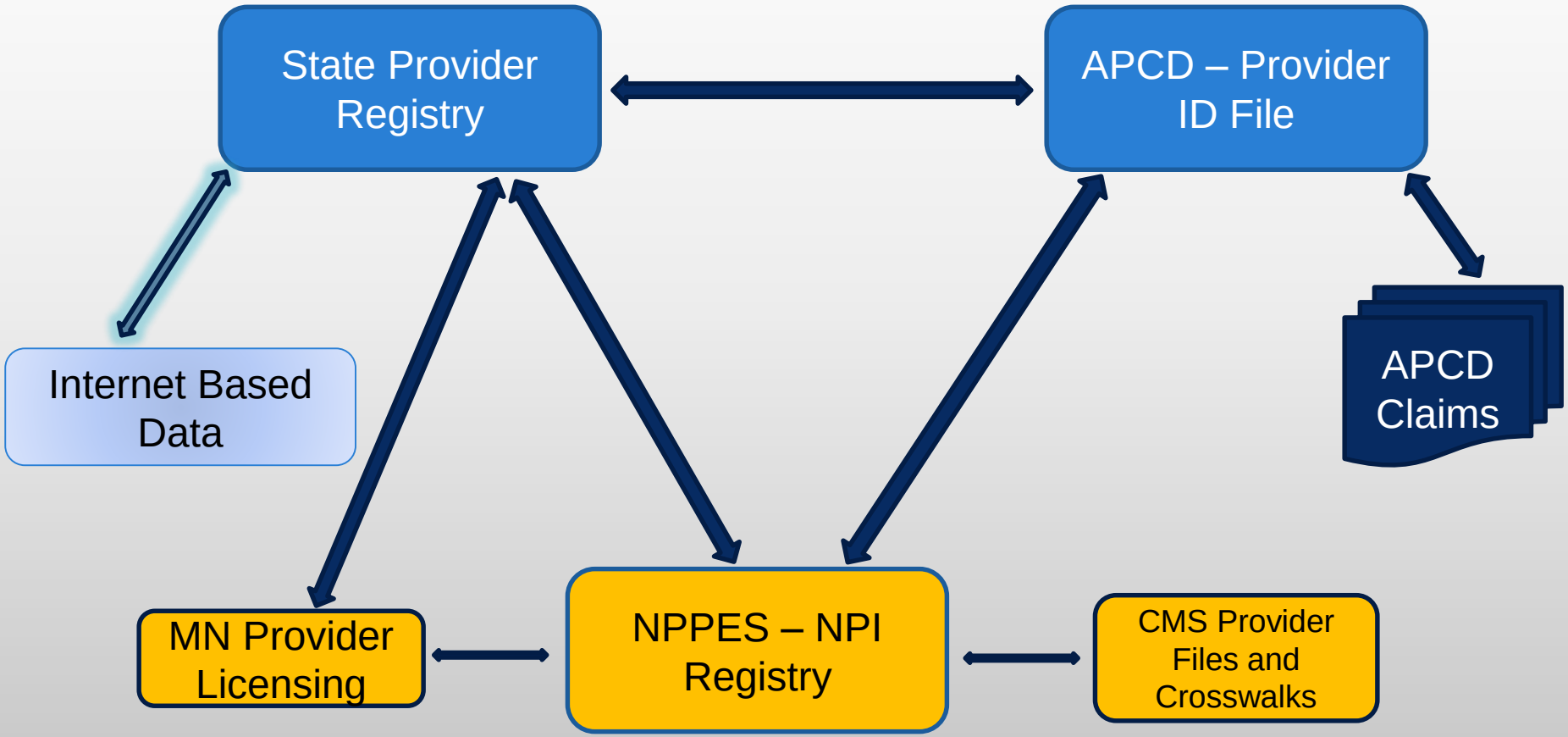

State Provider Registry
APCD – Provider ID File
Internet Based Data
APCD Claims
MN Provider Licensing
NPPES – NPI Registry
CMS Provider Files and Crosswalks

Task 3 – Merge APCD Provider / Registry Provider Tables

1. Prepare data for merge by merging of "Modal" data values for the clusters in your Provider data from Task 2

    - Find the value that occurs most often within the created clusters
    - Require a minimum number / percentage of occurrences for the modal value
    - Assign the "modal" cluster value to represent the cluster on:
        - 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

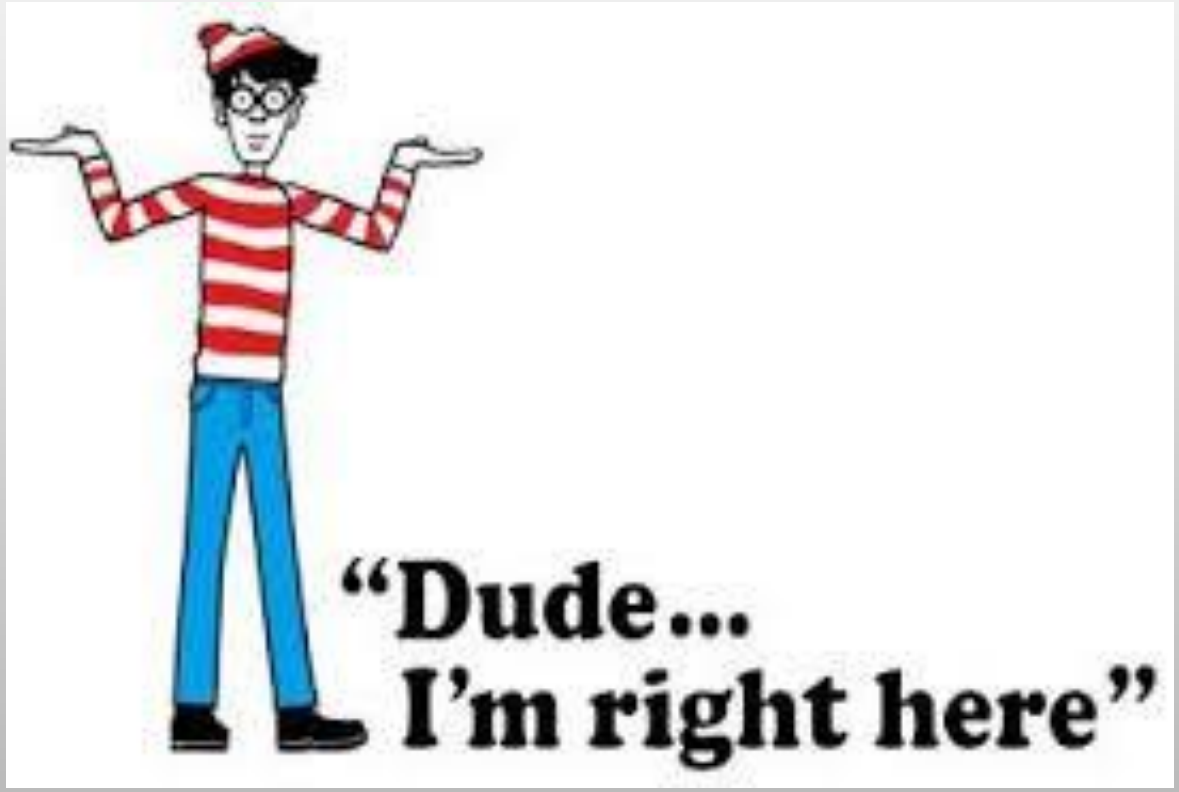

State Provider Registry
APCD – Provider ID File
Internet Based Data
APCD Claims
MN Provider Licensing
NPPES – NPI Registry
CMS Provider Files and Crosswalks

Finally Verify and Analyze the Results

- Compare modal categories of the resulting merges.
  - Do they correspond to expected values in other sources of data (CMS, Licensure)
- Evaluate resulting data enhancement impact
  - analyze % of visits and dollars in your APCD that can be 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

Lessons Learned?

- Provider identification is a known data quality challenge
- As many as 100X as many providers in your APCD than in your state
- **An external provider data you can trust is critical**
- **Staff resources and/or vendor expertise also critical**
- Time and perseverance pays off - the identification patterns in the data are there you just need to find them



Best Practices in Data Quality

Validation Checks

New Hampshire CHIS comprehensive health care information system

Mary Fields
October 7, 2014

Overview

- Set Expectations
- Screen for Format
- Data Validation Process

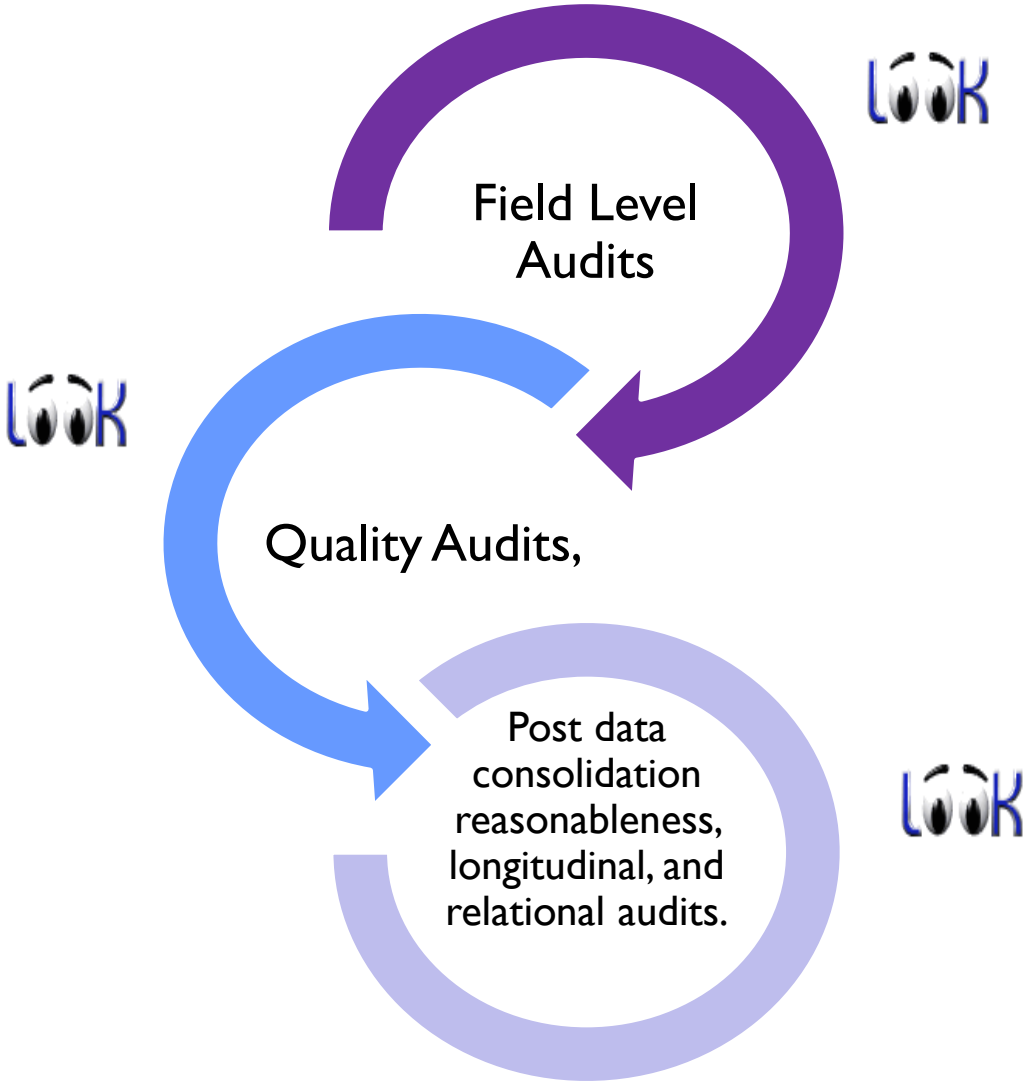
Clear Expectations of Carriers

- Develop Standards
  - Data dictionary
  - Rules
  - Submission Guide
  - Provide File Status

Screen for Format & Data standards

- Before the file is transmitted, files are screened through a preprocessor to ensure the format is correct and meets the basic requirements.
  - Data in header and trailer records match dates and record counts; all required fields are included.
- This step also encrypts patient identifiers.

The validation process is primarily composed of three groups of audits:



Field Level Checks

- All transmitted files are first checked to determine if they are in the correct format and have been created using the provided pre-processor.
- Field level audits are then employed to evaluate:
  - All required data elements are included.
    - Examples:
      - Payer ID
      - Line Counter
      - Insurance Type
      - Paid Amounts
  - Field length is within the required limits;
  - Any fields required to report specific codes only contain valid values, such as gender (e.g., Gender "M", "F" or "U");

Quality Audits

- Quality audits are employed to determine if the data submitted meet a pre-determined level of reasonableness
  - Some examples include:
    - % Records with member zip code not within primary state
    - % Total paid to charges
    - % Inpatient records missing admit type
    - % of institutional claims vs. % of professional claims
- Default thresholds have been established for approximately 200 quality audits.

Threshold Establishment and Alteration

- Default thresholds are applied to the field level audits for each element in file, and for each quality audit.
- The standard acceptable threshold for field length, field type, and data value audits is 100%. However, there are some fields where the acceptable thresholds are less.
- Individual field completeness thresholds are established for each data element and are specific to the file type.

Reasonableness, Longitudinal, and Relational

- Additional audits are run to identify any global issues that would not be evident during the field and quality level audit process.
- Examples of these audits are: frequency of individual field values; volume reconciliation; and cost/utilization reasonableness.

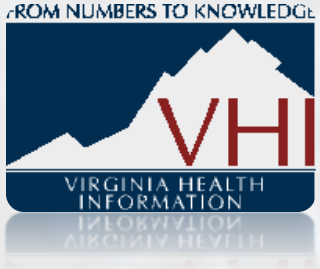
Exception Requests

- When a file fails the carrier is provided a report containing the errors.
- They are required to correct all files, if possible and resubmit.
- If they are not able to correct the file due to systematic issues they may request an exemption or adjustment for data variances through a standardized process.
  - If approved it will be in effect (at a maximum) for that calendar year.

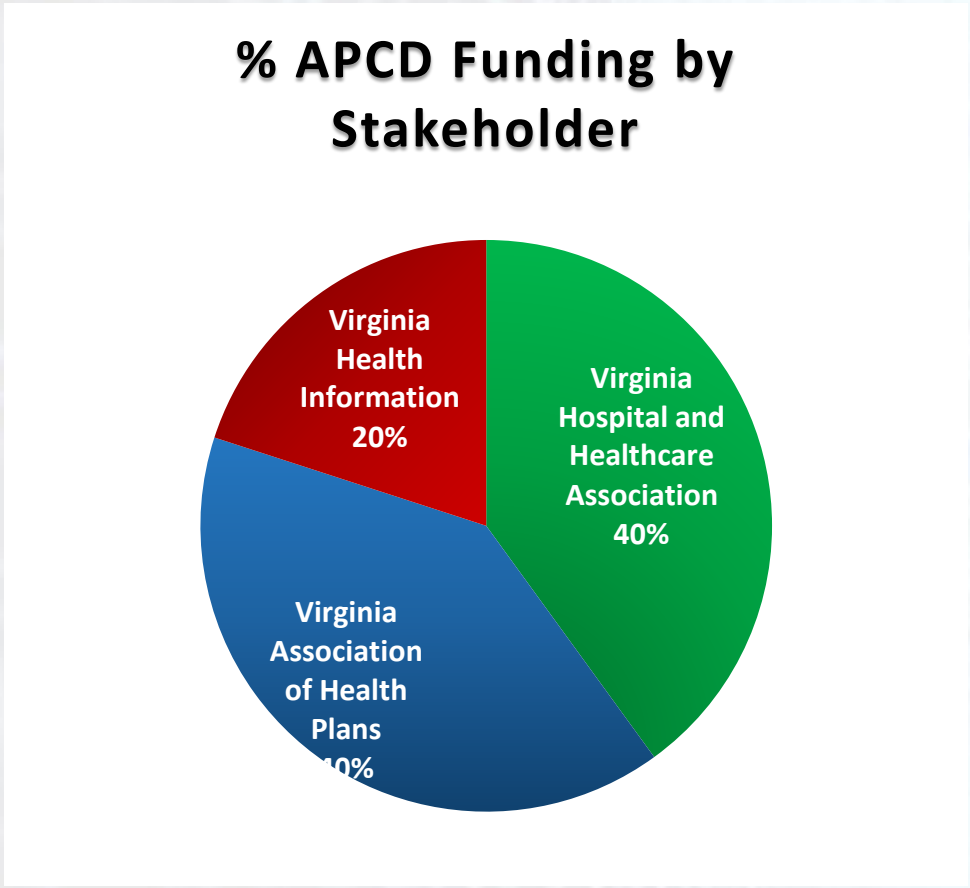
Historical Data Files

Kyle Russell

NAHDO Conference
October 7th, 2014

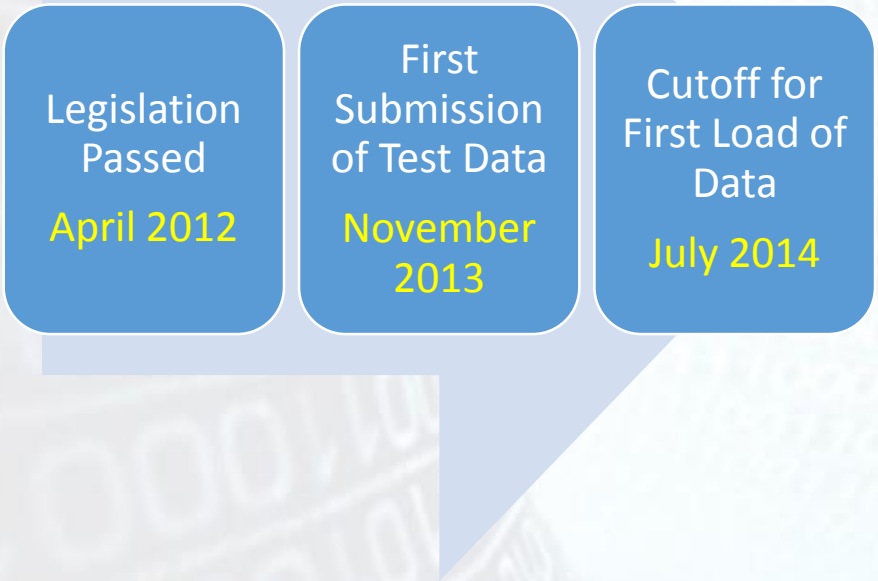


Virginia's APCD



- Voluntary Program
- Multi-Stakeholder Funding
- Milliman chosen as vendor for data collection, warehousing, and analytics

Timeline



- 9 Largest Commercial Submitters
- 2011 data to current
- Testing soon

Test File Process

Courtesy Creative Commons

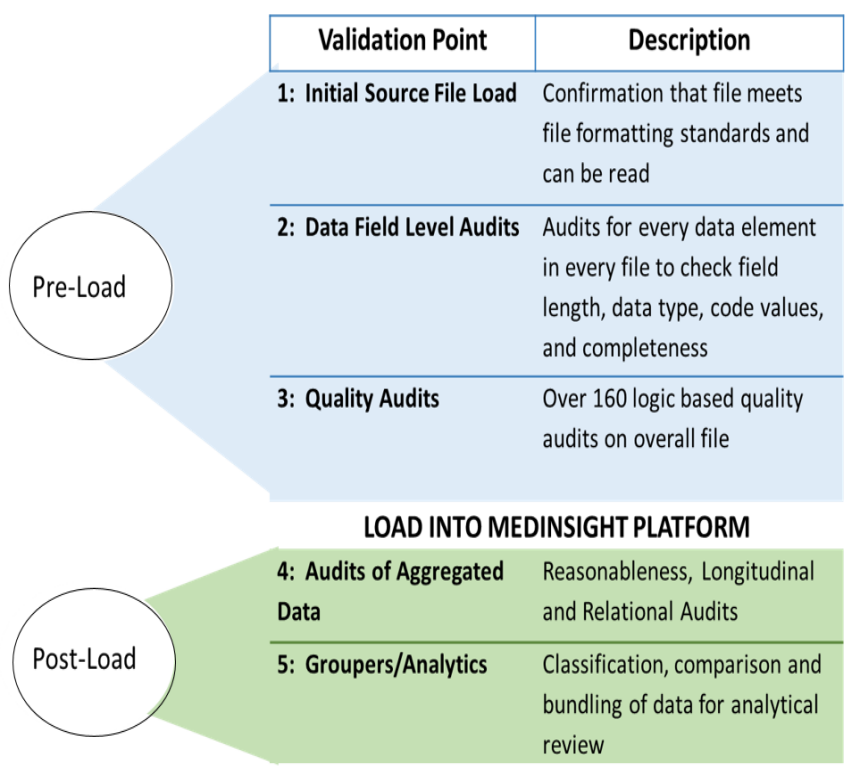
- 1 month of data
- Learn Payer Specific modifications
- Test data may not be reflective of earlier data

Receiving Historical Files

Courtesy Creative Commons

- Begin with earliest time period collected
- Quarterly or Monthly files
- Monitor claims and eligibility volume

Data Validations and Exemptions



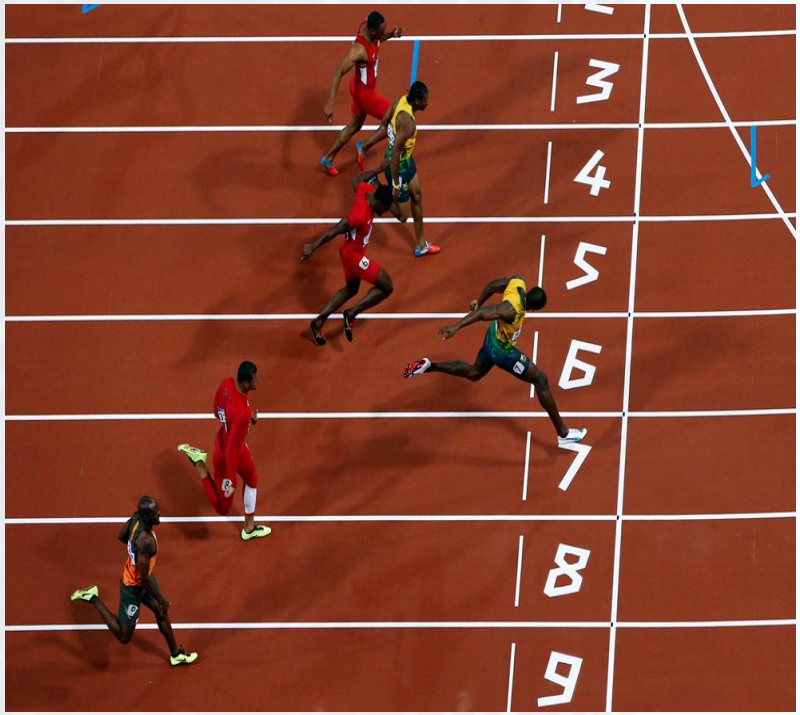
- How long?
- How wide?
- Have shared expiration dates as often as possible

Data Resubmissions

Courtesy keitherspring.info

- What warrants a resubmission?
- Timeline- how old is the data?
- Analyze new files beyond expected corrections

Load Cutoff



Courtesy images.76themes.com

- All data submitters may not have made the same progress
- Stay flexible with timeline

Takeaways

Have shared exemption expiration dates when possible

Have an idea of expected file volume

Resubmissions will happen

Load cutoffs can be a moving target