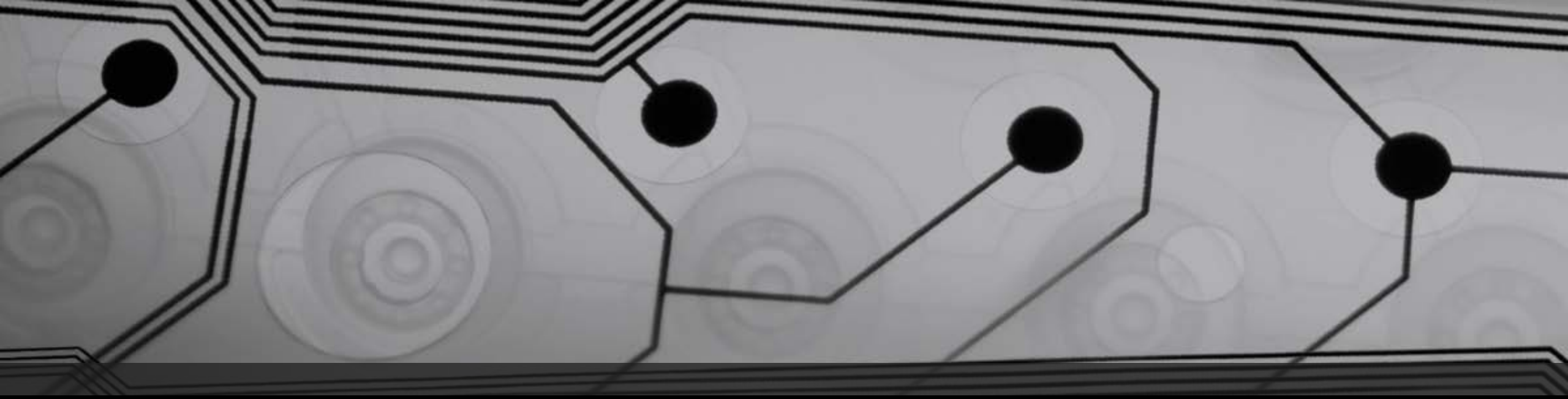

Data Migration Lessons Learned

Data Quality Opportunities for Data at Rest

Data Migration Goals/Challenges

- The California Immunization Registry (CAIR) is consolidating 7 regional systems into a single state-wide immunization system.
- CAIR has more than 20 million clients with close to 200 million immunizations
- Complete extensive analysis on the existing data at rest for any potential data quality improvements
- Develop a strategy to minimize data integrity risks that might occur as a result of the migration
- Fully Utilize the new IIS functionality with the legacy data

Begin the Process

- Weekly meetings early on with key stakeholders between the California Department of Public Health (CDPH) and the Hewlett Packard Enterprise (HPE) teams
- CDPH provided an ERD and Test Database with test data to HPE for analysis against the new CAIR Database
- Determine how the tables would be grouped and categorized to complete the analysis
  - Review the legacy acceptable data values as compared to the new IIS Database
  - Discuss the mapping and transformation of data elements
  - Determine if there are any data fields that could be updated in the legacy CAIR before the migration process
  - Repeat analysis process for each grouping

Table and Data Groupings

- Provider Organization Data
- User Data
- Vaccines
- Manufacturers
- Tradenames
- Inventory
- Patients and Immunizations
  - Patient Flat File
  - Immunization Flat File
  - Comment Flat File
  - TB Test Flat File
  - Responsible Persons Update

Provider Data

- Multiple tables for both the legacy CAIR and new CAIR database
  - Determine what data fields from the tables were a 1:1 match
  - Determine what, if any fields could be transformed
    - Use of composite keys to fill data fields
    - Crosswalk Legacy codes to new codes
  - Determine if default values can be used for missing data fields
- Possible Duplicate Providers across the 7 regions
- Provider Surveys for updates required for the migration

User Data

- Multiple tables for both the legacy CAIR and new CAIR database
  - Determine what data fields from the tables were a 1:1 match
  - Determine what, if any fields could be transformed
    - Use of composite keys to fill data fields
    - Crosswalk Legacy codes to new codes
  - Determine if default values can be used for missing data fields
- User Surveys for update to user accounts needed for the migration

Vaccine/Trade Name/Manufacturer

- Utilized Vaccines in the new database
  - Determine what data fields from the tables were a 1:1 match
  - Determine what, if any fields could be transformed
    - Use of composite keys to fill data fields
    - Crosswalk Legacy codes to new codes
- Utilized Tradenames in the new database
  - Legacy database stores very few Tradenames
  - Use of composite keys and data mapping to put in correct Trade Names
- Addition of missing manufacturers into the new database
  - Allows for any legacy data using old manufacturers to be added correctly

Inventory

- Multiple tables for both the legacy CAIR and new CAIR database
  - Determine what data fields from the tables were a 1:1 match
  - Determine what, if any fields could be transformed
    - Use of composite keys to fill data fields
    - Crosswalk Legacy codes to new codes
  - Determine if default values can be used for missing data fields

Patient Data

- Multiple tables for both the legacy CAIR and new CAIR database
  - Determine what data fields from the tables were a 1:1 match
  - Determine what, if any fields could be transformed
    - Use of composite keys to fill data fields
    - Crosswalk Legacy codes to new codes
- Client Deduplication
  - Utilizing Flat File for batch Data Exchange in new CAIR application
  - Sending and storing CAIR Legacy ID for data quality purposes
  - Sending and storing CAIR region ID for data quality purposes

Immunization Data

- Multiple tables for both the legacy CAIR and new CAIR database
  - Determine what data fields from the tables were a 1:1 match
  - Determine what, if any fields could be transformed
    - Use of composite keys to fill data fields
    - Crosswalk Legacy codes to new codes
- Client Deduplication
  - Utilizing Flat File for batch Data Exchange in new CAIR application
  - Sending and storing CAIR Legacy ID for data quality purposes
  - Sending and storing CAIR region ID for data quality purposes

Patient Comments/TB Test Data

- Map Legacy comments to existing comments
  - Add legacy comments to new database that could not be mapped
- TB Test Data was able to be a 1:1 match
  - Created TB Flat File as part of the batch Data Exchange process

Summary

- More complete User Information
- Better Quality and unique Provider Organizations
- Client Deduplication
- Quality Immunization Data that utilizes the functionality of a new IIS