

DE-IDENTIFICATION ASSESSMENT OF MA CHIA DATA COLLECTION & SHARING PRACTICES



Kathy Hines

Senior Director of Partner Operations and Data Compliance,
Massachusetts Center for Health Information and Analysis
(MA CHIA)

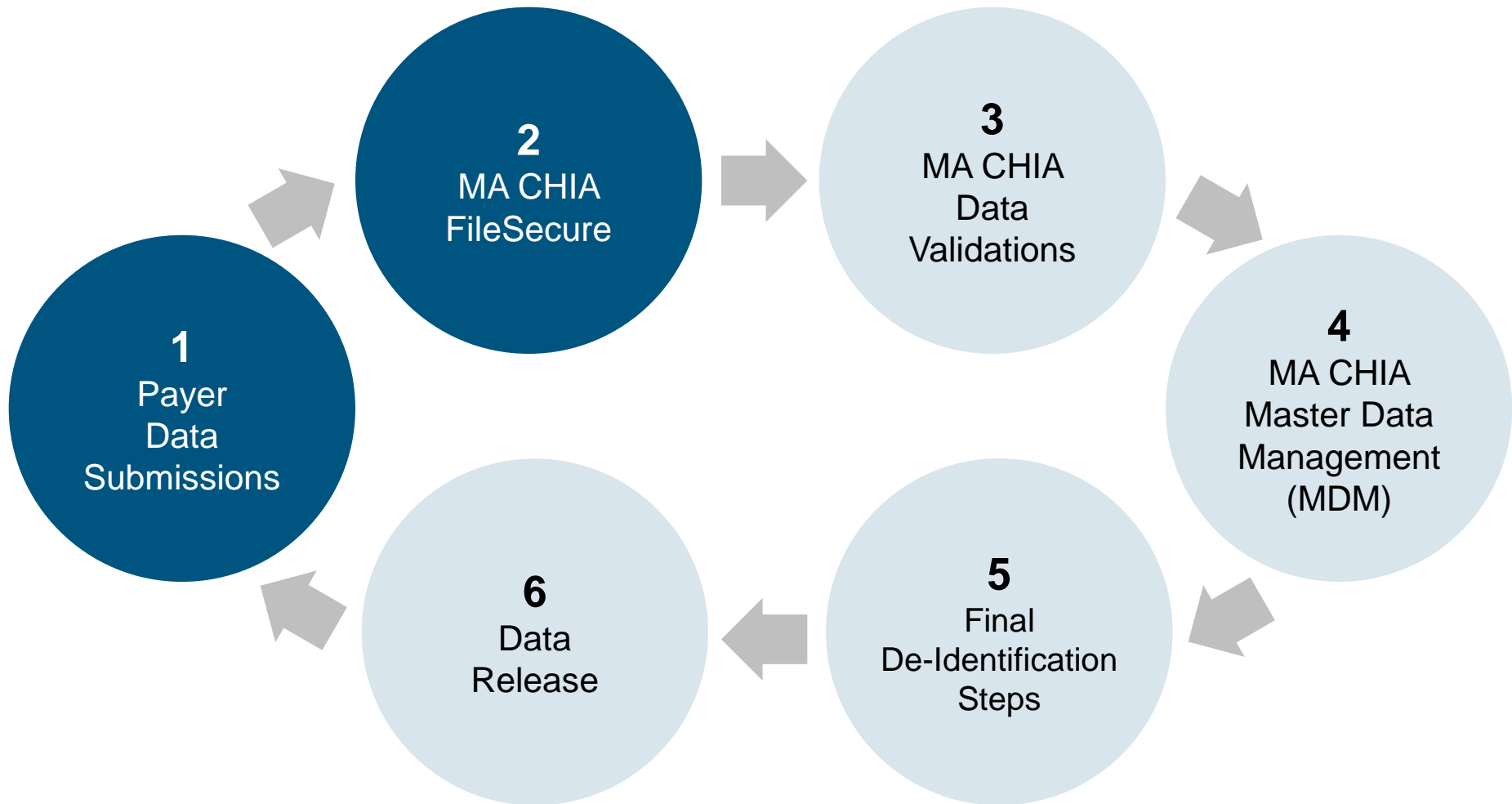


Samuel Chick

Data & Analytics Process Manager,
Onpoint Health Data

November 6, 2019

De-Identification Cycle

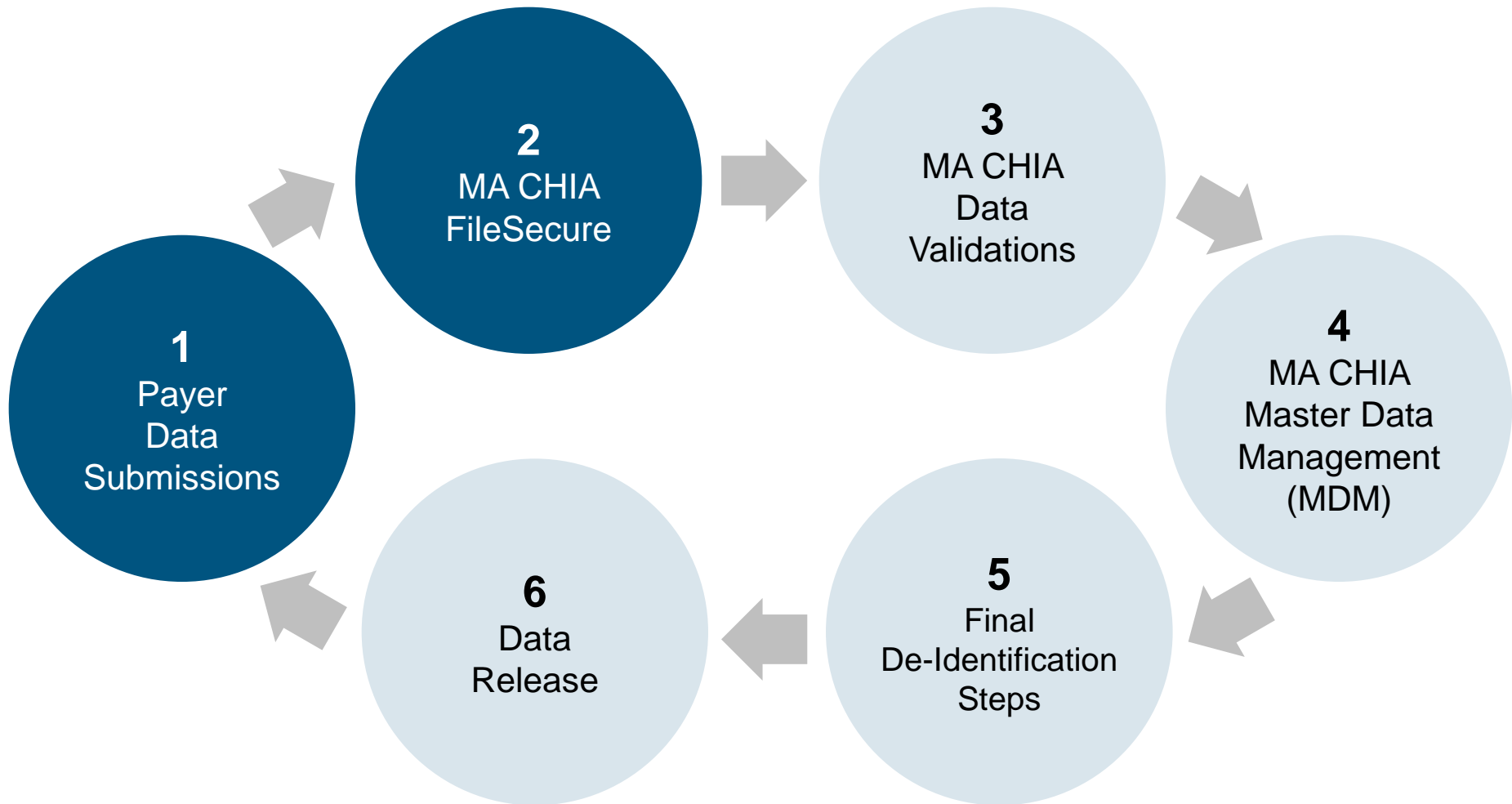


First Steps: 2016

Submission Guide Changes – Data Removal

- Claims
 - First/Last names
 - Social Security numbers (SSNs)
 - Street/City address information
 - ZIP code limited to first 5 digits
- Eligibility
 - Street/City address information
 - ZIP code limited to 5 digits

De-Identification Cycle



First Steps: 2016

FileSecure

- Created by CHIA to be installed at data submitters' sites
- Successfully masks identifiable information prior to submission
- Performs validation/standardization before masking
- Utilizes a NIST-certified SHA-3 hashing algorithm

Updated Data Transfer

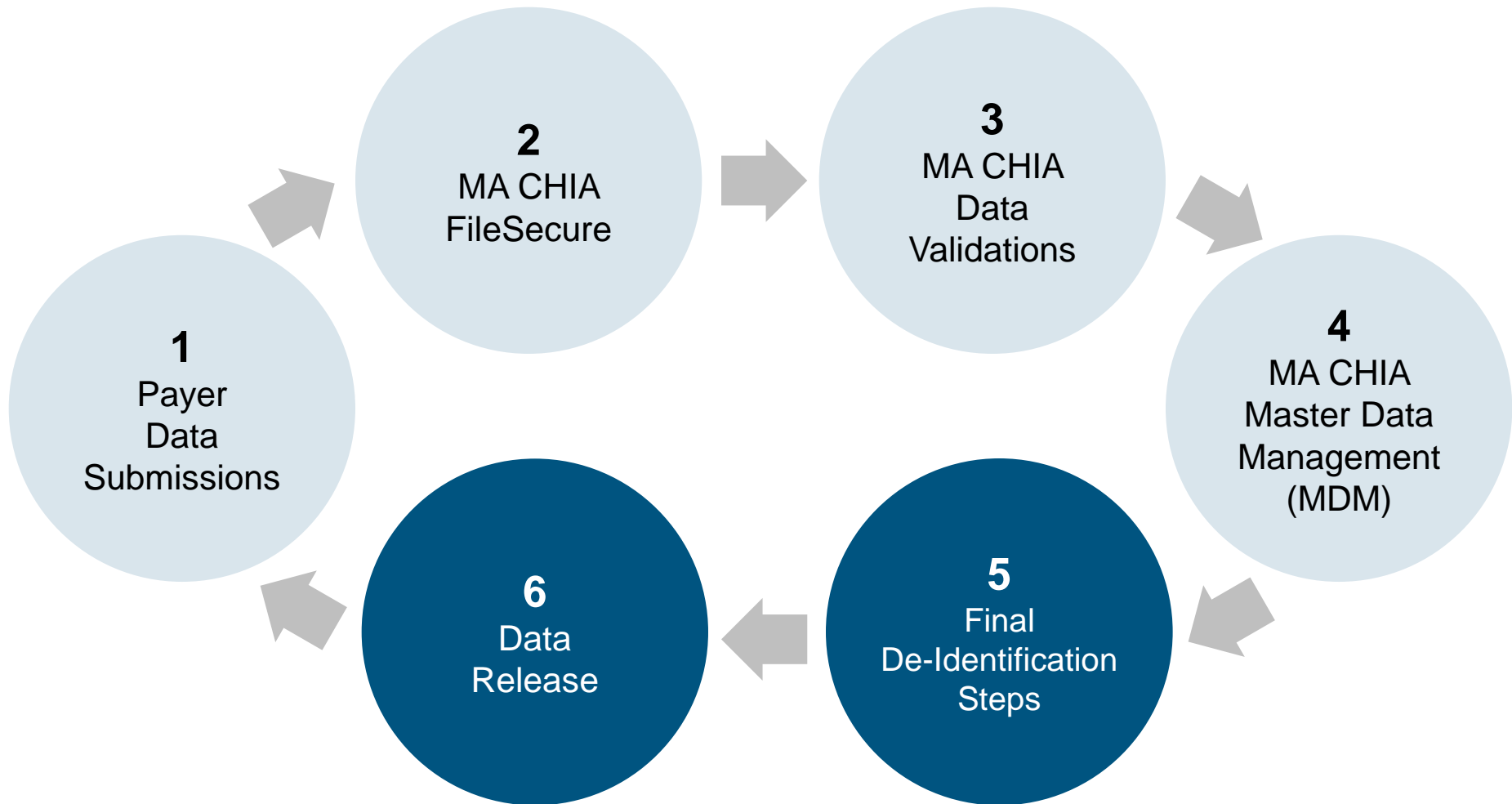
- Standard SFTP client with AES-256 encryption

First Steps: 2016

FileSecure Process on Member Eligibility

- Performs validation/standardization before masking
 - First Name – Standardized
 - Last Name – NYSIIS phonetic standardization
 - SSNs – Incorrectly formatted SSNs removed
 - Date of Birth – Incorrectly formatted dates removed; creates YYYYMM field
- Successfully masks identifiable information, including names, SSNs, and full dates of birth

De-Identification Cycle



HIPAA Safe Harbor vs. De-identification

Safe Harbor

Pros

- Easy to implement and maintain

Cons

- 18 data elements redacted or removed entirely
- More restrictive than de-identification with respect to birth dates, service dates, and geographic data

De-identification

Pros

- Methodology tailored to data set in question
- Lower overall risk of re-identification

Cons

- No single method for implementation
- Requires routine reassessment
- More restrictive than Safe Harbor with respect to some individual claim lines

Stage 1. Worked with MA CHIA to Define Approach

- Established the variables to be considered for a formal re-identification risk analysis
 - Catalogued all **direct identifiers** and **quasi-identifiers**
- Determined acceptable risk levels
 - Minimum cell size, maximum risk, average risk
 - Assumed an adversarial environment where the recipients of the data have knowledge of quasi-identifying values for the individual

Stage 2. Developed Initial Data Profile

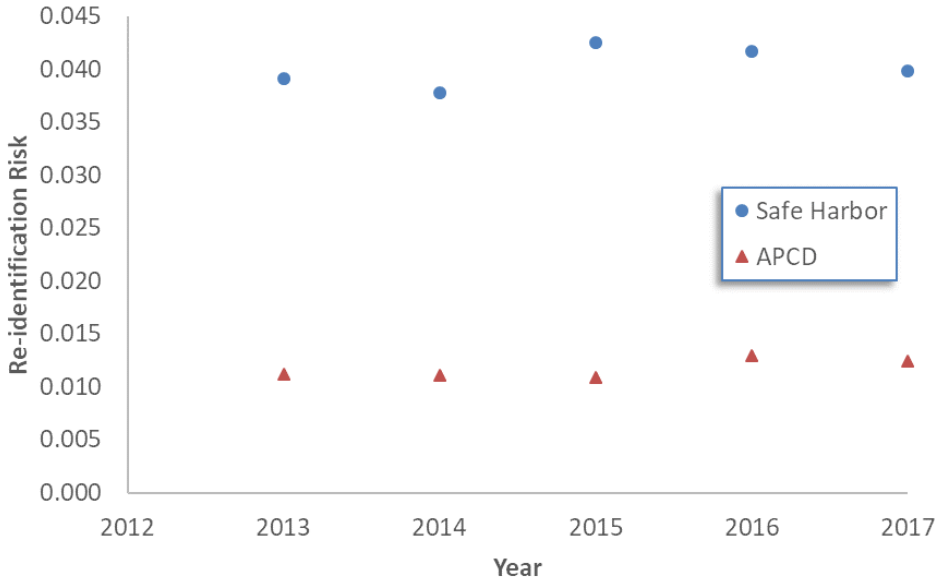
- Profiled quasi-identifying variables independently and jointly
 - Generated a distribution of risk for the data
- Developed acceptable transformation strategy and performed **risk mitigation** where risk was deemed to be above acceptable ranges

Note: The following risk calculations are examples only; they do not reflect actual CHIA reporting results.

Number of Fields	Fields in Policy	Risk
1	ZIP*	0.000578
2	ZIP*, GENDER	0.000484
2	ZIP*, BIRTHYEARMONTH	0.006937
2	ZIP*, STATE	0.000624
3	ZIP*, STATE, GENDER	0.000649
3	ZIP*, STATE, BIRTHYEARMONTH	0.006917
4	ZIP*, STATE, GENDER, BIRTHYEARMONTH	0.013033
4	ZIP*, STATE, GENDER, ETHNICITY	0.001232
5	ZIP*, STATE, GENDER, BIRTHYEARMONTH, ETHNICITY	0.026261

Stage 3. Applied the Data Strategy

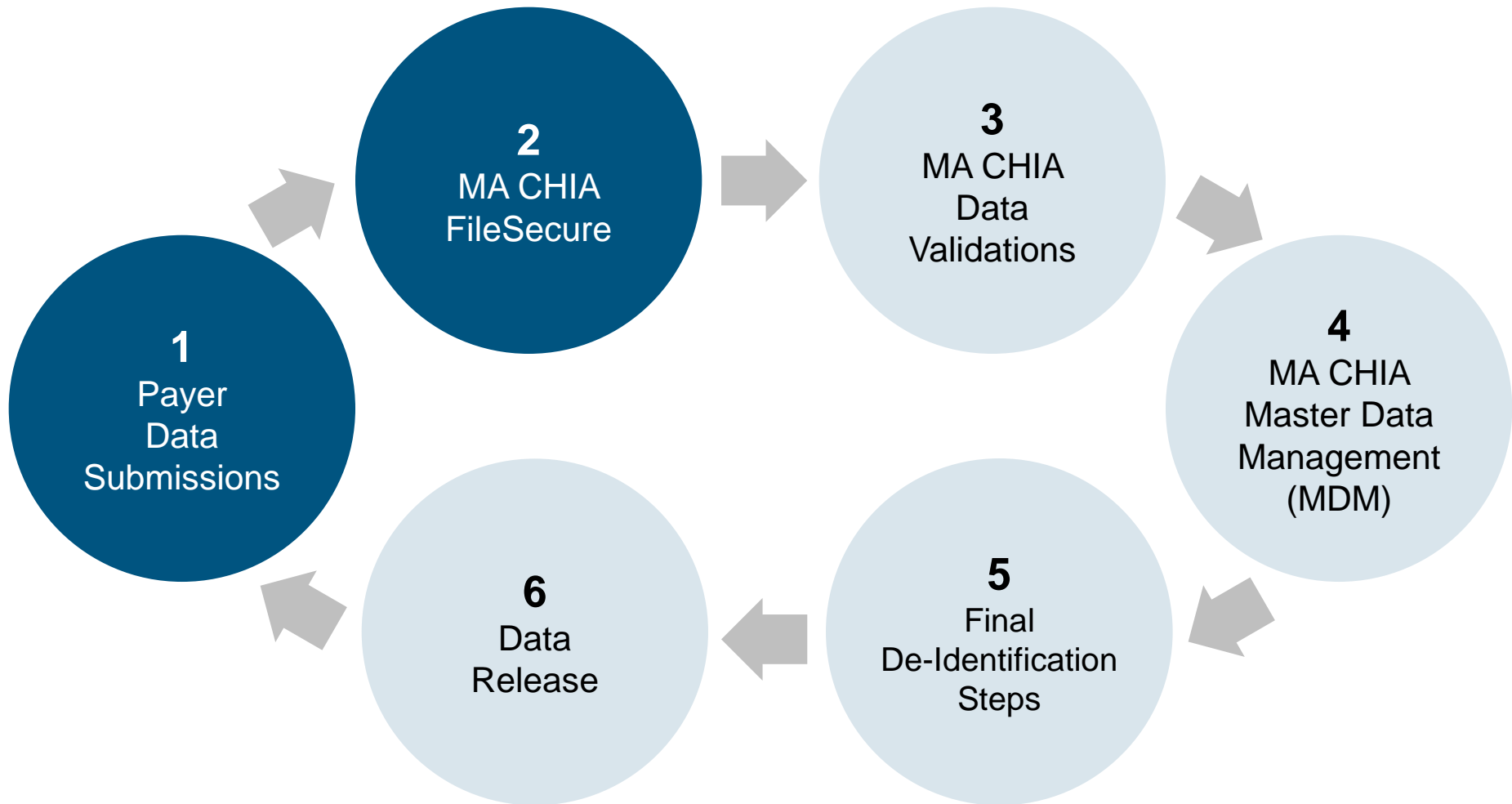
- The risk mitigation model was applied to multiple years of data (MA APCD data set years 2012–2017) to assess the risk stability over time and project a solution for the following year



Stage 4. Assessed Data Strategy Results

- The data shared with CHIA was determined to meet the de-identification requirements set forth in the HIPAA Privacy Rule and related relevant law and regulation
- Determination based upon the transformation of the fields and applies to CHIA as well as anticipated downstream recipients, provided they:
 1. Enter into a data-use agreement with CHIA that prohibits data linkage that would lead to the re-identification of patients
 2. Correspond to CHIA-approved data recipients that do not have access to identified or identifiable patient data

De-Identification Cycle



Where are We Now?

Submission Guide Changes – Data Removal

- Claims
 - Member state
 - Member ZIP code
- Eligibility
 - Race/Ethnicity indicators
 - Disability/Marital/Student/Family size indicators
 - Language (list abbreviated)

Where are We Now?

FileSecure Process Updates on Member Eligibility

- ZIP code processing
 - Flag if invalid ZIP code
 - Retain MA ZIP codes only
 - Map MA ZIP codes to mask small areas in MA APCD
- State code processing
 - Flag if invalid state
 - Retain only New England and New York state codes

Where are We Now?

In Progress

- One ZIP code per person per year
- Exclusion of select diagnosis codes

Contact Us

- Kathy Hines, MA CHIA
kathy.hines@state.ma.us
- Samuel Chick, Onpoint Health Data
schick@onpointhealthdata.org
- Brad Malin, Privasense, LLC
brad.malin@protonmail.com