# Ensuring Governance and Traceability of Real-World Data (RWD) from Source to Regulatory Submission

## Use Case



# Introduction

- Objective: Transform RWD into reliable RWE for regulatory submissions
- Challenge: Ensuring governance, traceability, and integrity without accessing third-party source data
- Solution: Use cryptographic certifications to provide end-to-end verifiable proof of data reliability
  - ✓ Capture
  - ✓ Data utilization per licensing agreements
  - ✓ Transformations
  - ✓ Contemporaneous documentation for every action

# Importance of RWD Governance

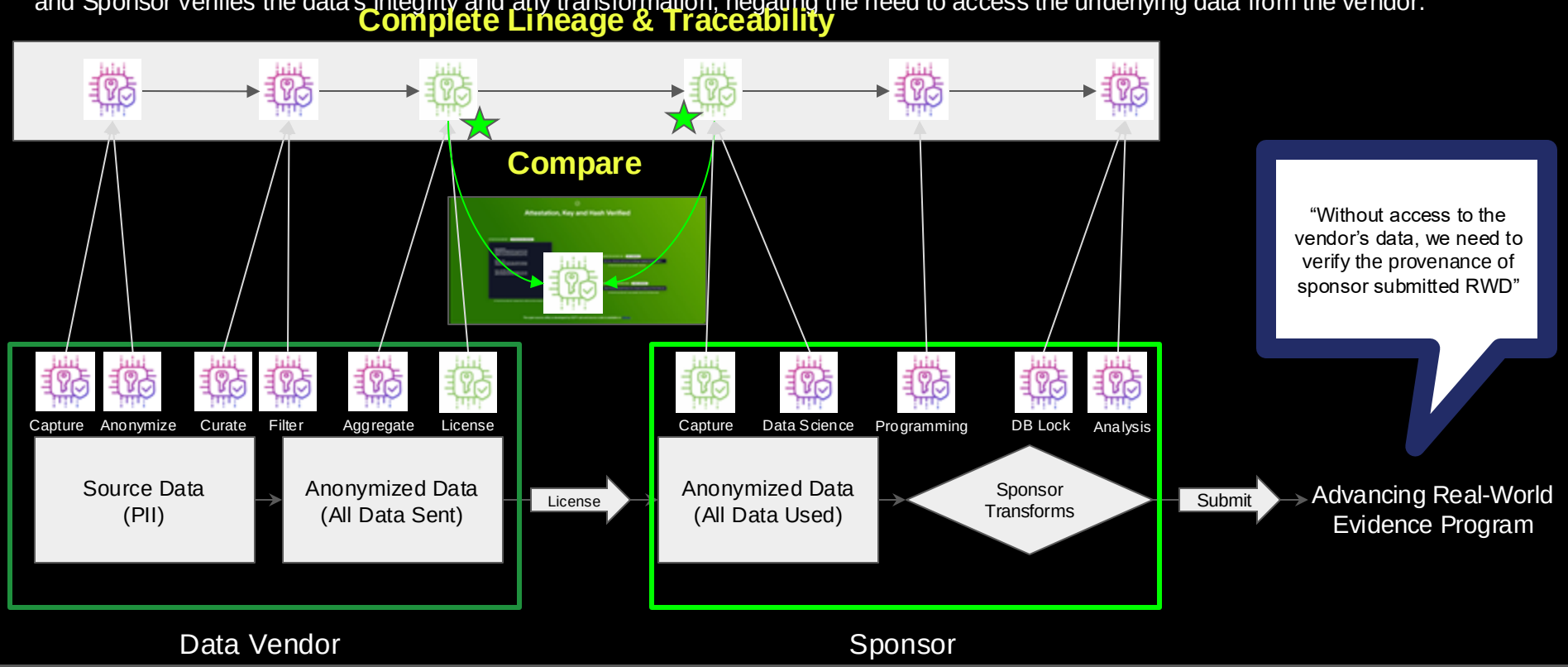
- Compliance:
  - Prove all parties meet HIPAA, FDA and EMA guidelines
- Data and process integrity:
  - Prove data quality throughout the lifecycle
- Transparency:
  - Provide auditors with complete audit trail for verifiable evidence
- Privacy Protection:
  - Variable level controls for appropriate data utilization proving anonymization, risk of reidentification or deletion

# Cryptographic Certifications

- Definition:
  - Digital fingerprints providing mathematical proof of data integrity
    - Real-time automated documentation: Who, what, where, when, when and how
    - Pair with human attestations for context: Why
- Benefits:
  - Immutable records
  - Reliability without accessing sensitive or prohibited data or code
  - Supports end-to-end data lineage, traceability and context for human-decisions
- Outcome:
  - Builds trust in data reliability for regulators

# Reliability and Traceability

**Cryptographic certificates** provide tamper-proof verification, creating a trusted chain of custody. Comparing hashes from the Data Vendor and Sponsor verifies the data's integrity and any transformation, negating the need to access the underlying data from the vendor.



# Use Case Overview

- Scenario: Create an Alzheimer's disease cohort for a mock regulatory submission
- Participants:
  - Data Providers
  - Data Vendors
  - Sponsors/CROs
- Process: Trace data from EHR source licensed to data vendor then through transformations to submission-ready datasets
- Streamlined Review: Build LLM prompt based review tool to expedite review with proof of lineage and governance over AI and workflow

# RWE Workflow

Step 1: Patient selection and licensed from EHR source (FHIR format)

Step 2: Transfer selected data to Data Provider

Step 3: Map FHIR data to OMOP-like structure

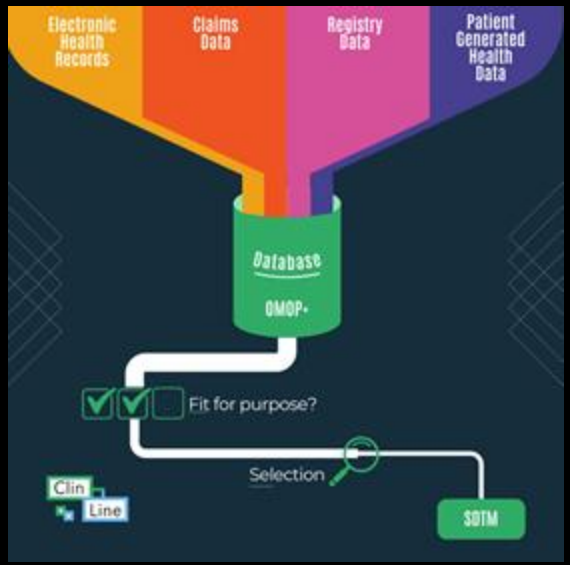
Step 4: Anonymize data

Step 5: Select data subset for Sponsor

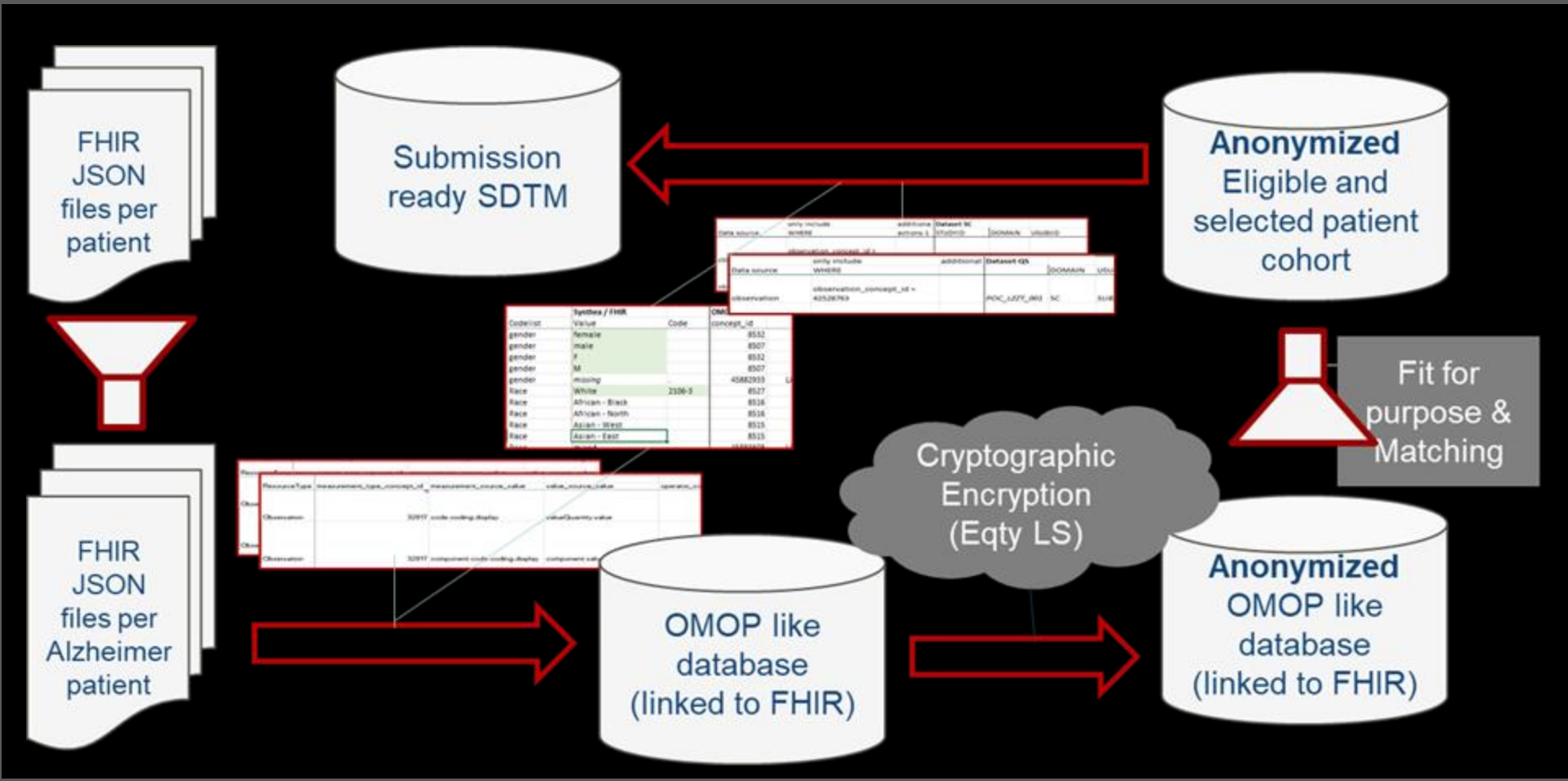
Step 6: Transfer data to Sponsor

Step 7: Sponsor processes data (filtering, imputations)

Step 8: Transform data to SDTM for submission



# Data Flow Diagram

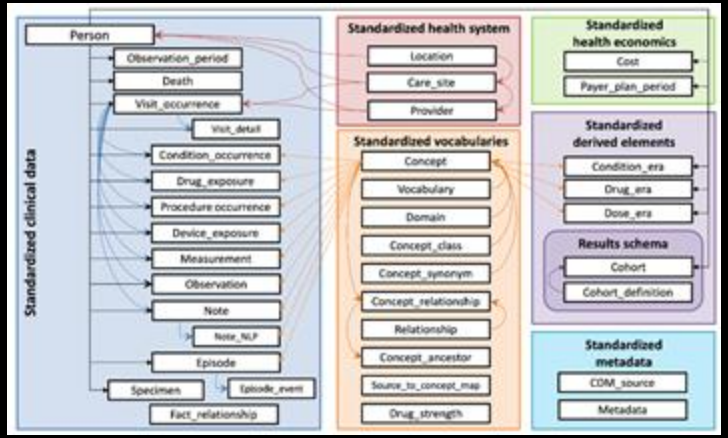


# FHIR Data to OMOP CDM Mapping

- Proportional mapping - Not all FHIR data is relevant for research
- Mapping challenges - FHIR
  - Standard mapping tools do not include all the traceability information
  - Same information nested at different levels or resources in FHIR
  - Complexity in some resources like AllergyIntolerance
- Mapping challenges - OMOP
  - Distinction between OMOP measurement or observation
  - Variation in source concepts
  - Some relevant source information not accounted for in OMOP standards

# Mapping to OMOP+ Datasets

- Why OMOP?
  - Observational research standard facilitates patient selection and fit-for-purpose assessment
  - Basic structure - comparable to structures used at data vendor sites
  - Contains source information variables
  - Univocal: no repeated instances of same information
  - Includes mapping capabilities
  - Includes child/ancestor relationships
- OMOP+
  - Added variables to optimize traceability
  - Excluded OMOP derived datasets
- Challenges
  - Learning curve
    - Relationships
  - Mapping and coverage



# Anonymization

Enhanced Privacy via Cryptographic Encryption

OMOP like database (linked to FHIR)


***Anonymized*** OMOP like database (linked to FHIR)

Scramble ID variables
Shift date and time variables
Remove potential sensitive information (pregnancy, mental health, family conditions

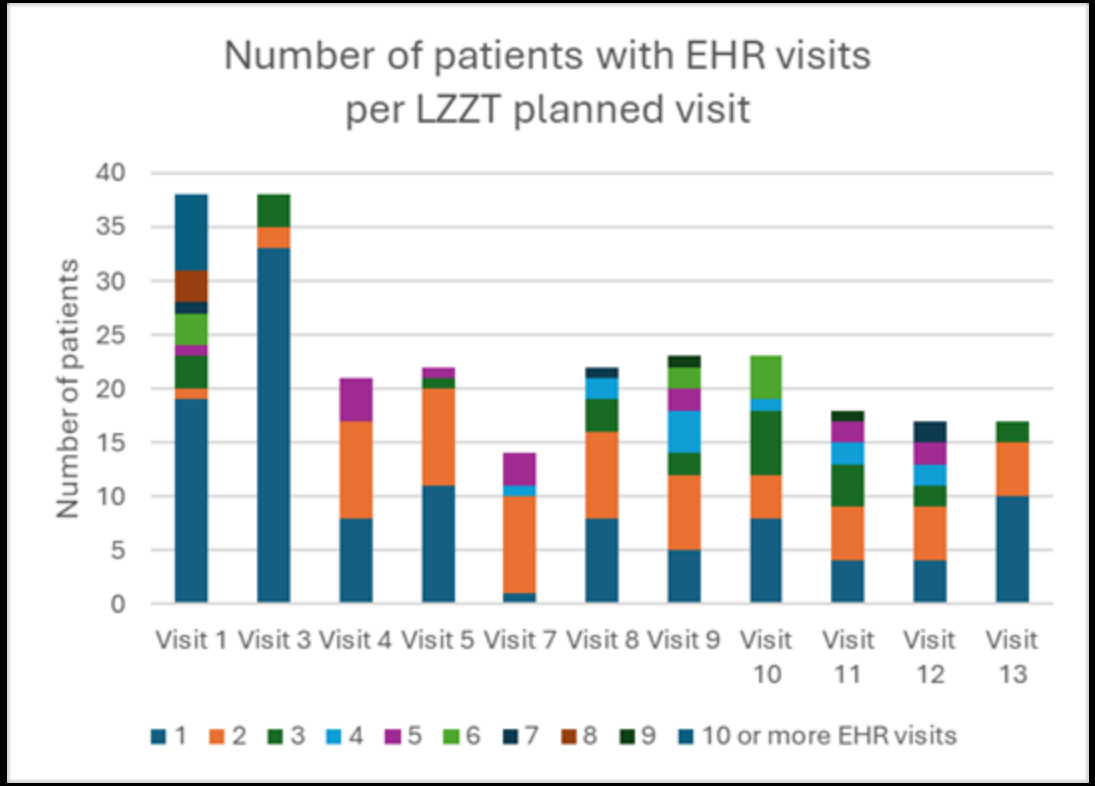
Audit and verification of privacy and integrity

# Cohort Selection and fit for purpose assessment

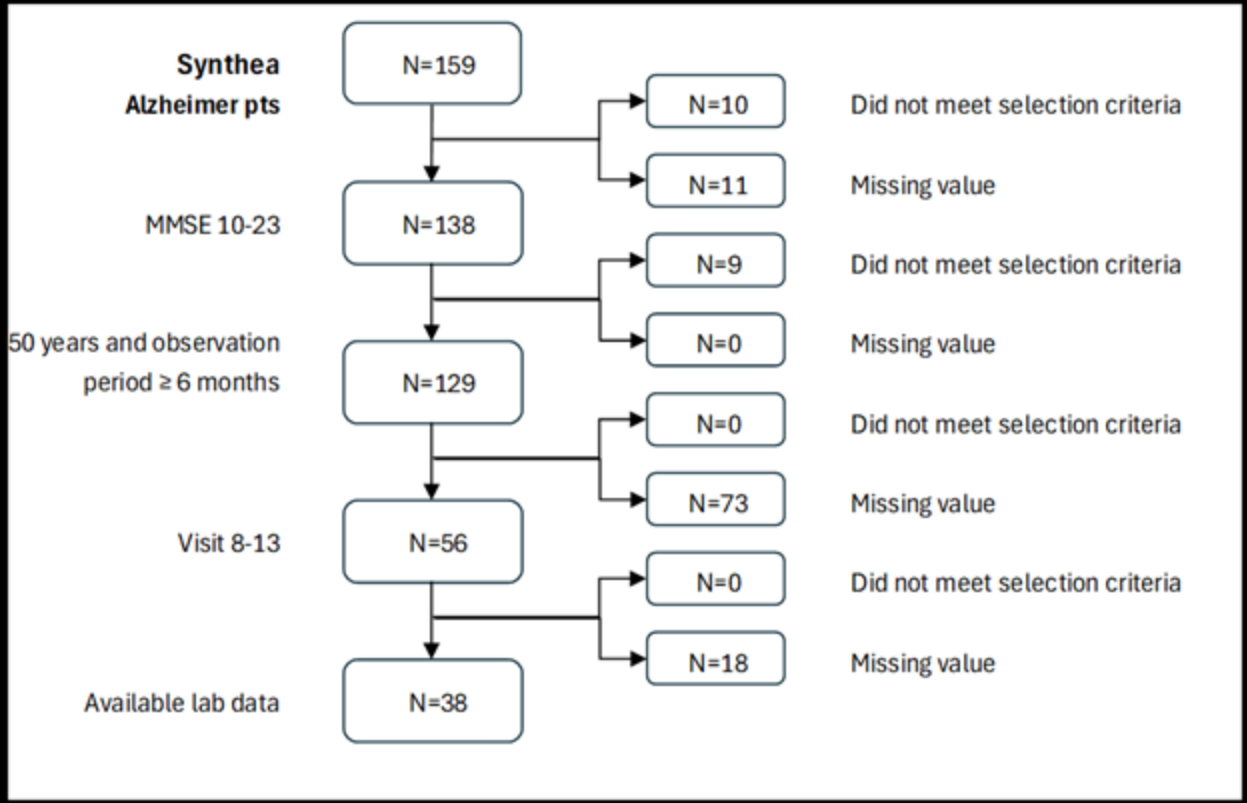
- Based on
  - Eligibility criteria
  - Data availability and validation
  - Windowing
  - Matching
- Automation
  - Mapping files
  - Defensive programming
- Index date
  - When eligibility criteria are first met

| Interventional trial arm | | | Real-world control arm | |
|---|---|---|---|---|
| **Visit** | **Week** | **Planned visit Day** | **WindowLower Day** | **WindowUpper Day** |
| 1 | -2 | -14 | -180 | -1 |
| 2 | -0.3 | -2 | **DO NOT** | **INCLUDE** |
| 3 | 0 | 0 | **INDEX DATE** | 7 |
| 4 | 2 | 14 | 7 | 20 |
| 5 | 4 | 28 | 21 | 34 |
| 7 | 6 | 42 | 35 | 48 |
| 8 | 8 | 56 | 49 | 69 |
| 9 | 12 | 84 | 70 | 97 |
| 10 | 16 | 112 | 98 | 125 |
| 11 | 20 | 140 | 126 | 153 |
| 12 | 24 | 168 | 154 | 174 |
| 13 | 26 | 182 | 175 | 188 |

# Data Availability per Visit



# Fit-for-Purpose Assessments

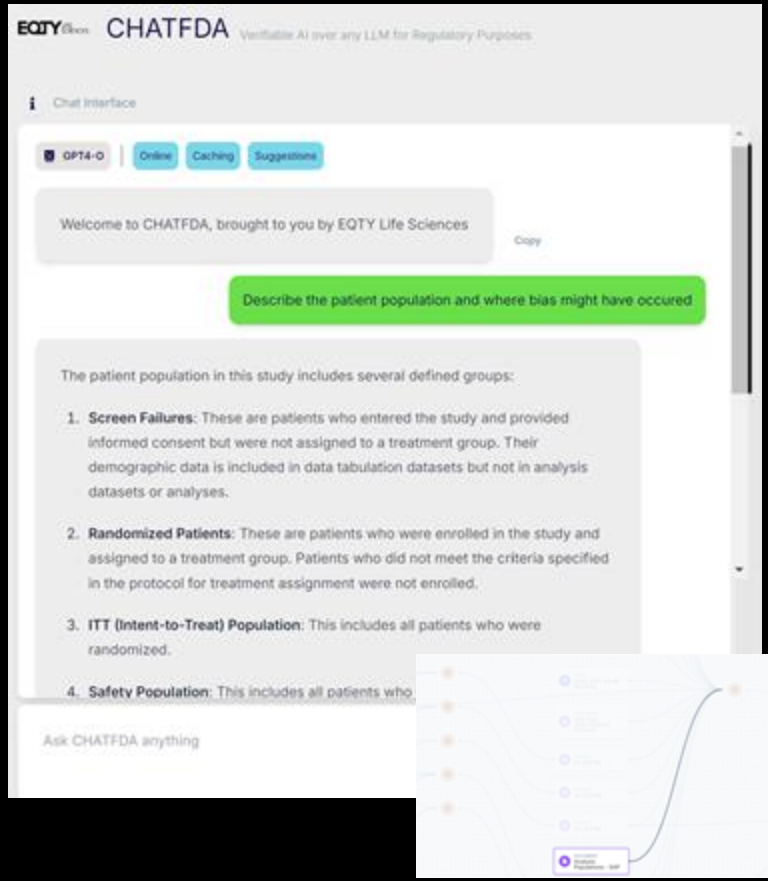


# Lineage and Traceability

- Lineage Graph:
  - Proves relationships between data sources, transformations, and outputs
  - Visualizes governance policy adherence
  - System interoperability with permissioned clickable navigation into dataset for audit
- Governance Sudio:
  - Utilizes cryptographic techniques like hashing
  - Enables traceability from source to submission
- Benefits:
  - Provides end-to-end audit trail
  - Demonstrates adherence to governance controls

# Backwards Mapping

- Capability:
  - Reconstruct original FHIR data from SDTM datasets
  - Power search over all data and analysis
    - Drill-down lineage
    - Proof of EU AI Act governance controls
- Limitations:
  - Excludes data omitted for confidentiality or proportionality



# Final Thoughts

- Key Takeaways:
  - Cryptographic certifications enhance data governance and traceability
  - Automated Documentation and Verifiability
  - Sponsors can prove data reliability without exposing sensitive data
  - Regulatory reviewers gain tools for efficient and confident evaluations
  - Extendable by adding USDM to drive protocol based data decisions

- Final Thought:
  - Adopting these methods strengthens trust in RWE for regulatory decisions. **Cryptography can inform auditors of the integrity of real-world data, ensuring suitability for RWE.**

# Next Steps

- For Sponsors and Data Providers:
  - Consider implementing cryptographic tools in data workflows
  - Collaborate on establishing industry standards
- For Regulators:
  - Engage with sponsors adopting cryptographic certifications to trust RWE and AI
- For Industry:
  - Advocate for cryptography as the uniform approach to append to datasets for RWE and AI integrity

Berber Snoeijer
b.snoeijer@clinline.eu
clinline.org

Ali Dootson
alistair.dootson@eqtylab.io
eqtylab.io