From Clinical Notes to Research-Ready Data



Four stages, each with a checkpoint for trust

Marissa Seligman, PharmD
VP, Clinical Data Optimization
Real World Data Solutions
Optum Life Sciences

Alyce Kuklinski, RN, NP
VP, Client Solutions
Real World Data Solutions
Optum Life Sciences

AI in Healthcare & Pharma Summit • Boston, MA • November 2026

Trust at Scale

Fireside chat: What patients, professionals and regulators need to see
Discussion themes:

- Evidence of appropriate use, fairness and accountability
- Clear communication for patients and frontline professionals
- Transparent responses to model errors and unintended outcomes
- Measurable trust supported by operational controls

Optum

© 2026 Optum, Inc. All rights reserved.

? How does your team currently use real-world data?

1. Evidence generation for health outcomes studies
2. Post-marketing epidemiologic studies (safety, etc)
3. Clinical trial solutions (patient/provider/investigator identification
4. Clinical care

Optum

© 2026 Optum, Inc. All rights reserved.

? What is the most important criterion that you consider when evaluating the utility of any real-world data set/solution?

1. Latency
2. Completeness
3. Longitudinality
4. Validity

Optum

© 2026 Optum, Inc. All rights reserved.

? What utility do synthetic data have in your real-world evidence generation?

1. Synthetic data are critical to our work
2. We use synthetic data, but they are not an absolute necessity
3. We do not use or wish to use synthetic data

Optum

© 2026 Optum, Inc. All rights reserved.

? My organization has a clear, well-defined and transparent data governance framework in place.

1. Agree strongly
2. Agree
3. Neither agree nor disagree
4. Disagree
5. Disagree strongly

Optum

© 2026 Optum, Inc. All rights reserved.

Agenda

WHAT WE'LL COVER

The problem

Clinical notes add rich context and introduce privacy risks

The pipeline

Four stages with a checkpoint for trust at each stage

Collaboration

The teams needed alongside NLP specialists

Validation

Independent checks and meaningful measures

Governance

Controls that make decisions auditable

What trust looks like

Practical takeaways and lessons learned

Clinical notes: rich context, difficult to use

THE PROBLEM

Clinical notes add context that structured fields can miss, including symptoms, clinical reasoning and social factors.

- Abbreviations, boilerplate and copied text complicate extraction
- Protected health information (PHI) appears throughout notes
- De-identification must preserve useful context while reducing re-identification risk
- Research uses include cohort discovery, real-world evidence and trial feasibility

Confirm volume

Monthly note volume needs confirmation

18+

PHI and confidential data categories tracked

≈80%

Research-relevant data in unstructured notes
Industry estimate (Truveta)

What can you do with clinical notes data?	How can you apply learnings from clinical notes data?
Determine triggers in medication switching, factors in patient adherence and nonadherence	Understand the patient story underpinning treatment changes
Monitor physician prescribing patterns around clinical events	Inform physician medical education strategies and assess the quality of care being provided to patients
Discover patient cohorts for conditions not identified well or underrepresented in structured codes	Run studies focused on rare and underdiagnosed conditions
Classify lifestyle status (e.g., physical activity, diet, etc.) of patients based on clinical outcomes	Develop detailed phenotypes to support the development and commercialization of products

Optum

© 2026 Optum, Inc. All rights reserved.

Trust depends on the process around the model



Clear team responsibilities, independent checks and a defined response to errors make the pipeline trustworthy.

The real-world data pipeline converts clinical notes into de-identified research data. Collaboration, validation and governance support every stage.

NLP: natural language processing

AI: artificial intelligence *ML: machine learning*

© 2026 Optum, Inc. All rights reserved.