KYTHERA
DECIPHERING HEALTHCARE

Addressing Token Bridging & Data Optimization for Life Science Organizations

Introduction

Kythera Labs is a data technology company applying manufacturing engineering principles to the challenges of healthcare data sourcing, integration, standardization, and semantic master data management. Our approach systematically reduces inefficiencies, optimizes data utility, and ensures that data investments yield higher value—similar to how Lean Manufacturing principles maximize production efficiency.

Many life science companies experience a 25-30% patient drop-off due to tokenization mismatches, leading to fragmented patient journeys, underutilized data, and inefficiencies in patient identification. Rather than just solving isolated data issues, Kythera applies a structured methodology to diagnose, measure, and improve inefficiencies, ensuring data conforms to fit-for-purpose standards and maximizes its usability.

By applying a factory-like process to data management, we reduce waste across the data lifecycle:

Defective Data	Overproduction	Wasted Motion	Waiting	Overprocessing
Resolving quality issues from sources and suppliers	Preventing unnecessary data purchases	Eliminating inefficiencies in finding and structuring data	Reducing delays caused by slow data transformations	Removing redundant processing steps

Kythera's approach to patient mastering and token bridging directly addresses these inefficiencies. Our methodology ensures that patient records across multiple datasets can be accurately matched and retained, preventing unnecessary patient loss and allowing life science companies to fully utilize their data assets.

KYTHERA

⟩ Challenges in Token Bridging & Patient Mastering

Life science organizations face material challenges in fully leveraging their investments in data and technology for complex disease research and commercialization. The complexity of working with multiple data providers and the tokenization technology required to harmonize disparate datasets has created obstacles to achieving a unified and complete patient view.

Key Issues

- **Token Bridging Mismatches -** Many organizations experience patient drop-off due to tokenization mismatches Even when all datasets are aligned to a single token provider, discrepancies arise when non-native token sources introduce crosswalk complexities.

- **Data Overlap & Waste** – Estimates indicate that approximately 75% of purchased data overlaps across vendors. Tokenization inconsistencies lead to patient loss and limit the utility of available data. While consultants can de-duplicate records and construct patient journeys, they do not address the core issues of token bridging and resolution, leaving gaps in the data integration process.

- **Encounter-Level Duplication** – Beyond patient overlap, duplication at the encounter level leads to inefficiencies and unnecessary costs. Variations in reported patient interactions inflate data volumes and create inconsistencies, complicating analysis and obscuring patient histories.

⟩ Why These Issues Occur

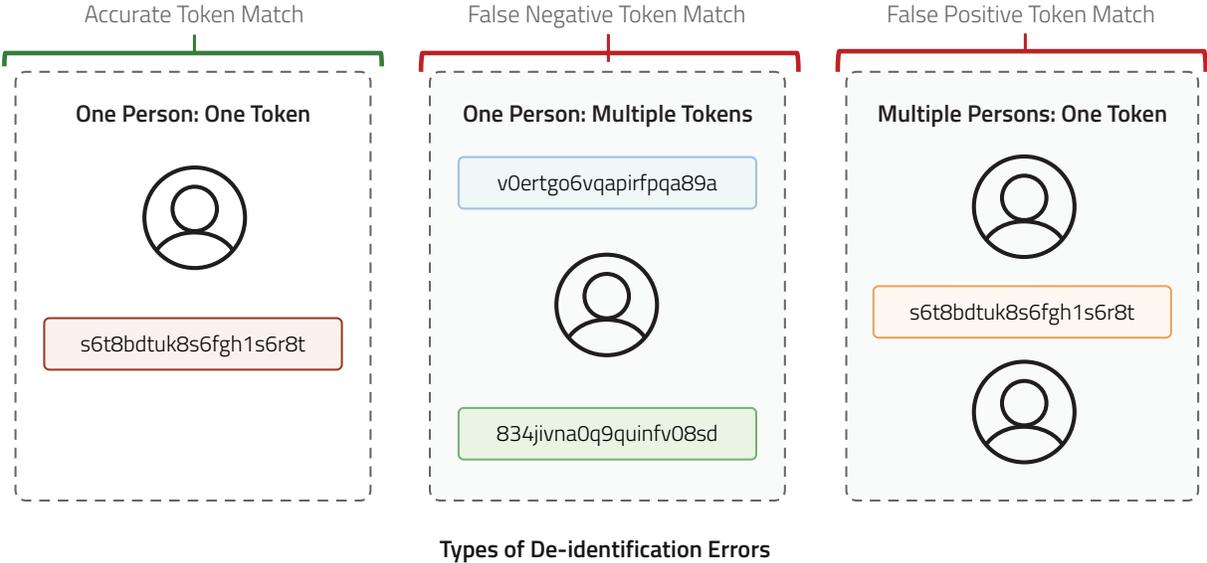Several factors contribute to challenges in token bridging and resolution:

- **Lack of Standardization** – Proprietary tokenization methods vary across data providers, leading to misalignment with common token standards. Data entry errors and legacy patient identifiers in source systems are a root cause of tokenization failures.

- **Incomplete Crosswalks** – Non-native token vendors require complex and often imprecise mapping to a common token standard.

- **Data Redundancy & Misalignment** – Overlapping data across multiple vendors creates inefficiencies and unnecessary costs, exacerbated by a lack of clear patient matching.

- **Third-Party Vendor Limitations** – Existing de-duplication vendors fail to resolve patient histories effectively, leading to patient drop-off.

- **Technology Solution Constraints** – Many solutions fail to accommodate tokenization inconsistencies, incomplete crosswalks, and redundant data, limiting seamless integration.
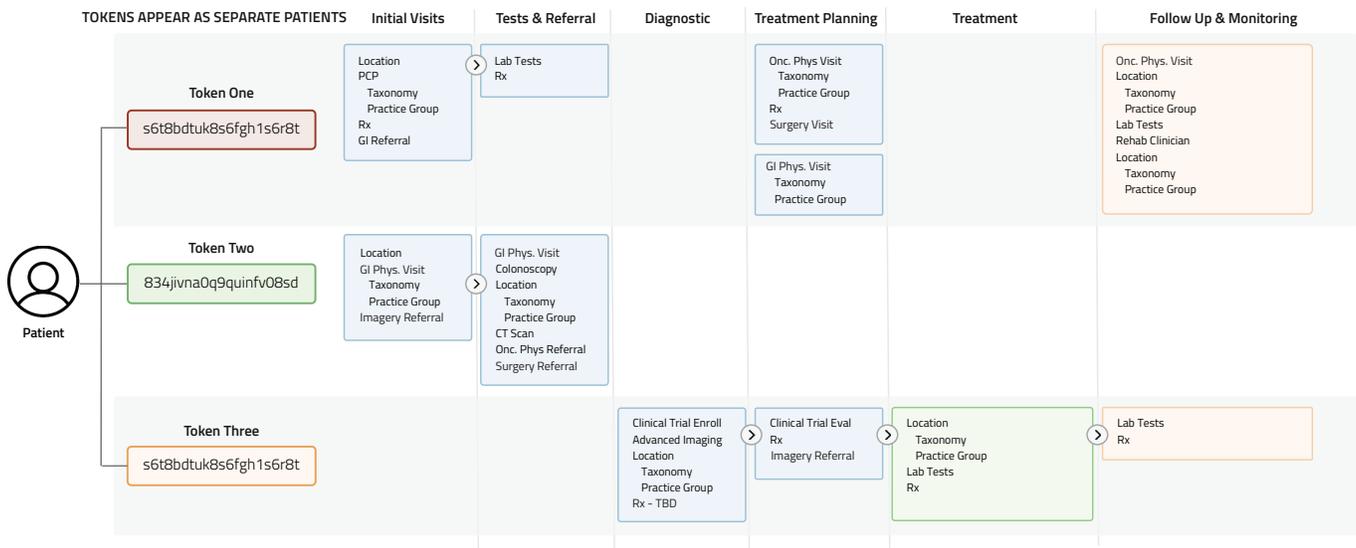
KYTHERA

The following case study highlights how Kythera Labs partnered with a customer to address these challenges and optimize their data refinement approach:

Optimizing Patient Data in Life Sciences to Reduce Data Loss and Costs

› Challenge

A leading Life Sciences company sought to improve patient mastering and token bridging across multiple de-identified datasets. Patients often received multiple identifiers, causing a 25-30% drop-off rate and resulting in fragmented records. Additionally, 75% of the purchased data overlapped across vendors, leading to unnecessary costs and inefficiencies. Multiple data sources reported the same patient interactions, inflating data volume and complicating the reconstruction of patient journeys. Finally, inconsistencies among tokenization providers (e.g., Datavant, HealthVerity, IQVIA) made patient matching difficult, as different systems failed to align under a common identifier.

Accurate Token Match	False Negative Token Match	False Positive Token Match
One Person: One Token	**One Person: Multiple Tokens**	**Multiple Persons: One Token**
	v0ertgo6vqapirfpqa89a	
s6t8bdtuk8s6fgh1s6r8t	834jivna0q9quinfv08sd	s6t8bdtuk8s6fgh1s6r8t

Types of De-identification Errors

KYTHERA

Multiple, Fragmented Patient Journeys (False Negative Match Error)

TOKENS APPEAR AS SEPARATE PATIENTS

Patient

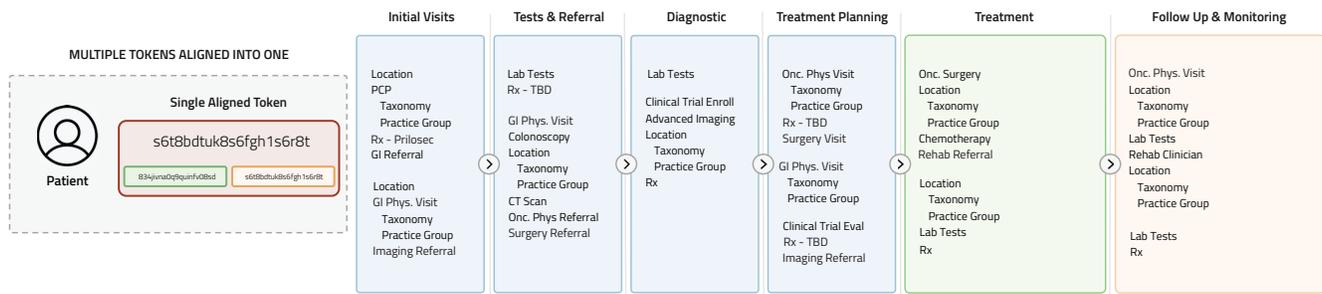|	Initial Visits	Tests & Referral	Diagnostic	Treatment Planning	Treatment	Follow Up & Monitoring
Token One `s6t8bdtuk8s6fgh1s6r8t`	Location PCP Taxonomy Practice Group Rx GI Referral	Lab Tests Rx		Onc. Phys Visit Taxonomy Practice Group Rx Surgery Visit GI Phys. Visit Taxonomy Practice Group		Onc. Phys. Visit Location Taxonomy Practice Group Lab Tests Rehab Clinician Location Taxonomy Practice Group
Token Two `834jivna0q9quinfv08sd`	Location GI Phys. Visit Taxonomy Practice Group Imagery Referral	GI Phys. Visit Colonoscopy Location Taxonomy Practice Group CT Scan Onc. Phys Referral Surgery Referral				
Token Three `s6t8bdtuk8s6fgh1s6r8t`			Clinical Trial Enroll Advanced Imaging Location Taxonomy Practice Group Rx – TBD	Clinical Trial Eval Rx Imagery Referral	Location Taxonomy Practice Group Lab Tests Rx	Lab Tests Rx

Multiple, Fragmented Patient Journeys as a Result of False Negative Match Error

> ## Solution

Kythera Labs deployed its advanced machine learning and token resolution methodologies to address these issues, implementing the following solutions:

- **Data Source Integration** – Standardized and normalized data across multiple vendors, ensuring consistent formatting and improved interoperability.

- **Advanced Token Resolution & Bridging** – Applied proprietary machine learning models to resolve token mismatches and improve patient retention.

- **Encounter-Level Refinement** – Eliminated duplicate encounters and reconciled fragmented patient journeys.

- **Confidence-Based Patient Matching** – Leveraged vector-based representations of patient histories to link patients across disparate datasets.

Token Alignment

MULTIPLE TOKENS ALIGNED INTO ONE

Single Aligned Token
`s6t8bdtuk8s6fgh1s6r8t`
`834jivna0q9quinfv08sd`   `s6t8bdtuk8s6fgh1s6r8t`

Patient

Initial Visits	Tests & Referral	Diagnostic	Treatment Planning	Treatment	Follow Up & Monitoring
Location PCP Taxonomy Practice Group Rx – Prilosec GI Referral Location GI Phys. Visit Taxonomy Practice Group Imaging Referral	Lab Tests Rx – TBD GI Phys. Visit Colonoscopy Location Taxonomy Practice Group CT Scan Onc. Phys Referral Surgery Referral	Lab Tests Clinical Trial Enroll Advanced Imaging Location Taxonomy Practice Group Rx	Onc. Phys Visit Taxonomy Practice Group Rx – TBD Surgery Visit GI Phys. Visit Taxonomy Practice Group Clinical Trial Eval Rx – TBD Imaging Referral	Onc. Surgery Location Taxonomy Practice Group Chemotherapy Rehab Referral Location Taxonomy Practice Group Lab Tests Rx	Onc. Phys. Visit Location Taxonomy Practice Group Lab Tests Rehab Clinician Location Taxonomy Practice Group Lab Tests Rx

Token Alignment Matches Patient with Multiple IDs

KYTHERA

> ## Results

Kythera Labs successfully implemented its methodology, delivering measurable improvements:

- Consolidated records from 525 million to 353 million mastered patient IDs.

- Increased patient retention, ensuring more complete and accurate cohort development.

- Reduced costs on redundant data purchases, enabling better resource allocation.

- Expanded the framework to additional therapeutic areas, driving broader application and continued improvements.
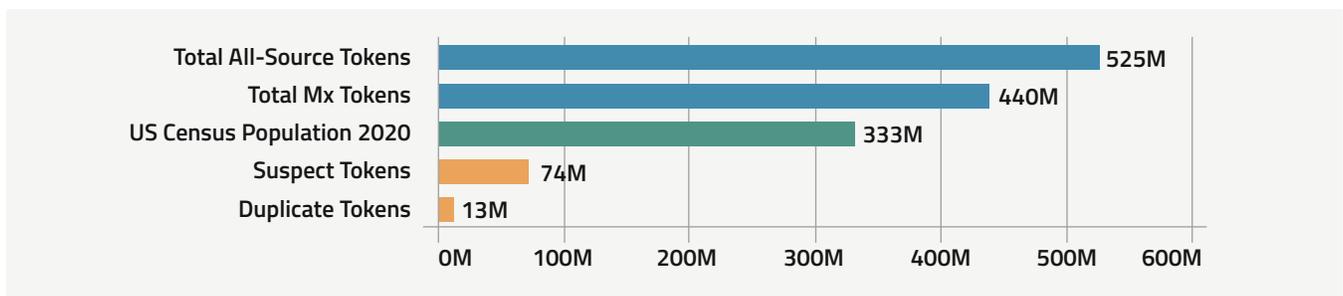
Summary of Analyzed Identifier Counts

By leveraging Kythera Labs' approach, the organization refined its dataset with greater precision, ensuring critical patient journeys remained intact and improving overall data accuracy, efficiency, and utility.

KYTHERA
DECIPHERING HEALTHCARE

346 Main Street
Franklin, TN 37064
www.kytheralabs.com