

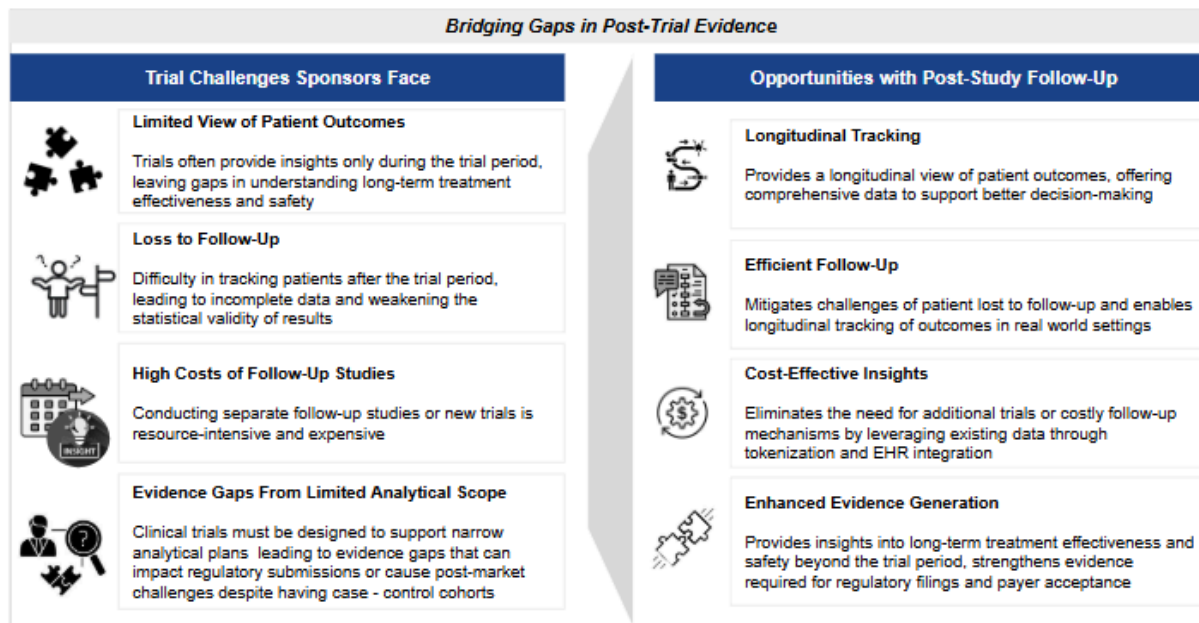
## Trial tokenization and EHR data interfaces for integrated post trial follow-up

As Graticule engages with life sciences sponsors on the challenges of long-term follow-up studies, a key theme has emerged: tracking patient data over long time scales is critical yet operationally complex. Trial tokenization is a powerful tool for this purpose, but using it to incorporate real-world data sources, particularly EHRs, remains a hurdle. Our discussions have highlighted the potential of honest broker systems, such as CLEHR, to bridge this gap. By combining tokenization with direct EHR-FHIR interfaces, these systems can streamline post-study follow-up, enabling sponsors to generate robust longitudinal insights while maintaining compliance and data security.

### **Enhancing Post-Study Follow-Up for Long-Term Insights**

Post-study follow-up enables sponsors to track patient outcomes beyond the trial period, providing a more comprehensive view of treatment effectiveness in real-world settings. By integrating multiple data interfaces—including tokenization and direct EHR connections—sponsors can extend the value of their clinical trials, generating richer long-term evidence while maintaining data security and compliance. This approach not only strengthens the evidence base for regulatory and payer decisions but also accelerates the adoption of best practices, empowering clinicians with real-world insights to improve patient care.

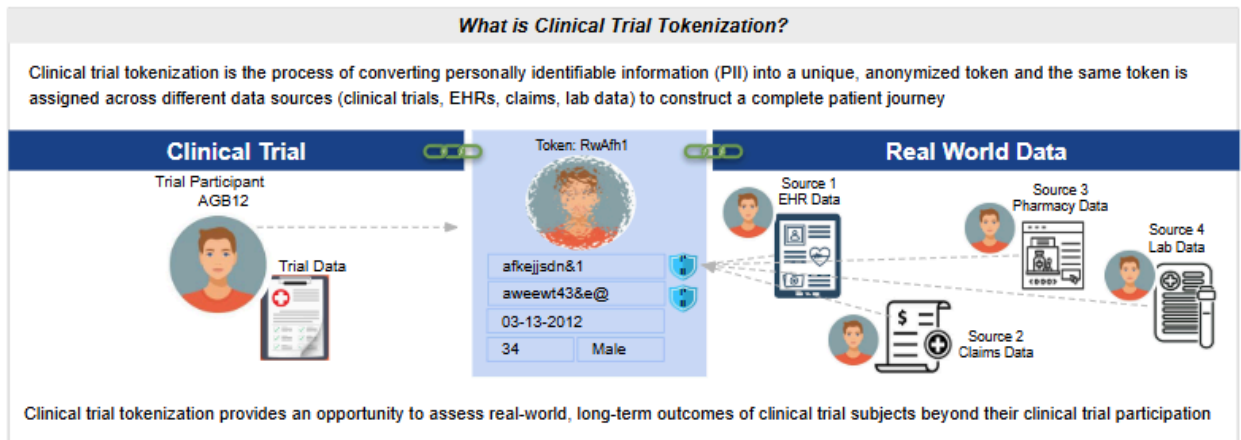
However, while clinical trials provide a controlled view of treatment outcomes, they often fail (or are not designed) to capture long-term effects once patients transition to standard care. As trials grow more complex and costly, sponsors face increasing pressure to limit study durations to what is strictly necessary for regulatory approvals. This may result in gaps in long-term safety and efficacy data, requiring additional post-approval investments to address uncertainties.



The challenge is particularly acute for emerging therapies, such as CAR-T cell therapies and treatments for rare diseases, where patient monitoring must extend years beyond the initial study period. Many sponsors miss the opportunity to obtain patient consent for long-term follow-up during initial enrollment, and current manual chart review and abstraction data collection methods remain fragmented and inefficient. As a result, valuable insights on long-term patient outcomes, comparative adverse event rates, and post-study treatment trajectories are either lost or are too expensive and complex to capture. Addressing these gaps is critical—not only to support regulatory and payer decisions but also to fully realize the potential of innovative therapies.

## Understanding Trial Tokenization

Trial tokenization has emerged as a powerful method for extending evidence generation beyond the confines of traditional clinical trial data collection. By using privacy-preserving record linkage, tokenization allows sponsors to securely link trial participants to external data sources such as administrative claims, aggregated EHR data, and registries like the Social Security Death Index. This approach enhances trial datasets by enabling a broader and more longitudinal view of patient outcomes, capturing information not collected through standard EDC tools.



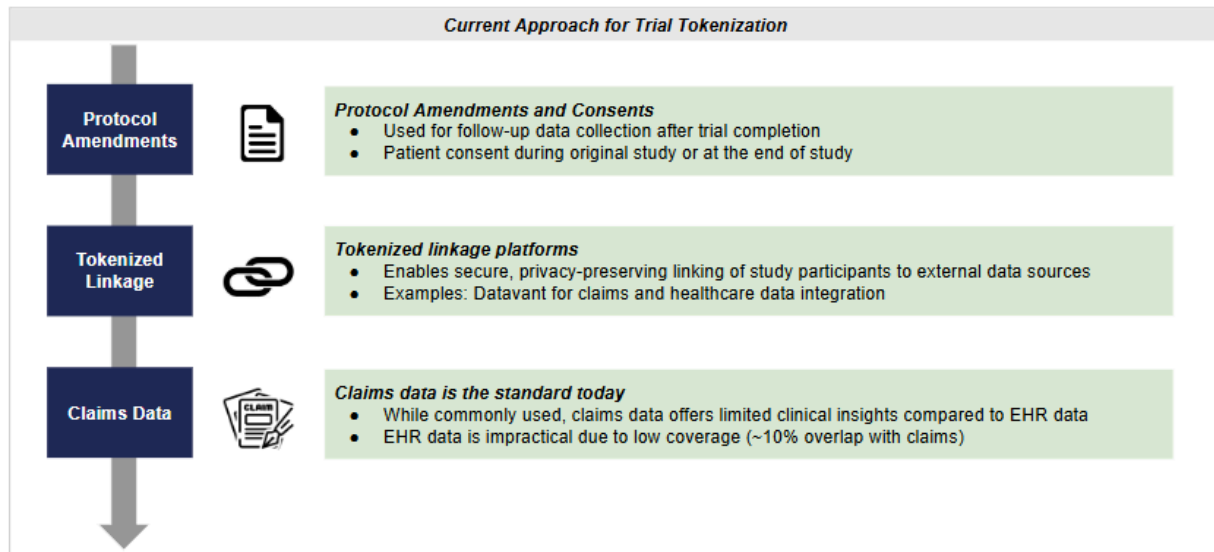
## The Current State of Trial Tokenization: Progress and Challenges

Protocols have been successfully adapted to incorporate tokenized data sources, enabling long-term evidence generation beyond traditional clinical trial endpoints. Several studies have demonstrated the feasibility of this approach, with claims datasets emerging as the primary linked source.<sup>1</sup> Given their extensive coverage—spanning over 200 million U.S. patients—claims data offer substantial overlap with trial populations, at close to 50%. One example of this was published in an article “Probabilistic Linkage of Randomized Controlled Trial Data to Administrative Claims: A Case Study of Patients from Baricitinib Clinical Trials”<sup>2</sup>. The article discussed trial patient match rates with IQVIA medical and RX claims. The results showed that 38 out of 78 (48.7%) consented RCT patients were successfully matched 1:1 to claims database patients. Additionally, the study noted that only 37% of trial participants opted to participate in post-trial follow-up when given the option to consent at the end of the study.

Efforts have also explored linking to death registries to capture mortality events that may be absent from medical records. While tokenization is technically viable, challenges remain—particularly with EHR integration. Inconsistent record matching and limited patient overlap have hindered the widespread use of EHR data for post-study follow-up, restricting its role in longitudinal evidence generation.

<sup>1</sup>Walters C, Langlais CS, Oakkar EE, Hoogendoorn WE, Coutcher JB and Van Zandt M (2025) Implementing tokenization in clinical research to expand realworld insights. *Front. Drug Saf.* 5:1519307. doi: 10.3389/fdsfr.2025.1519307

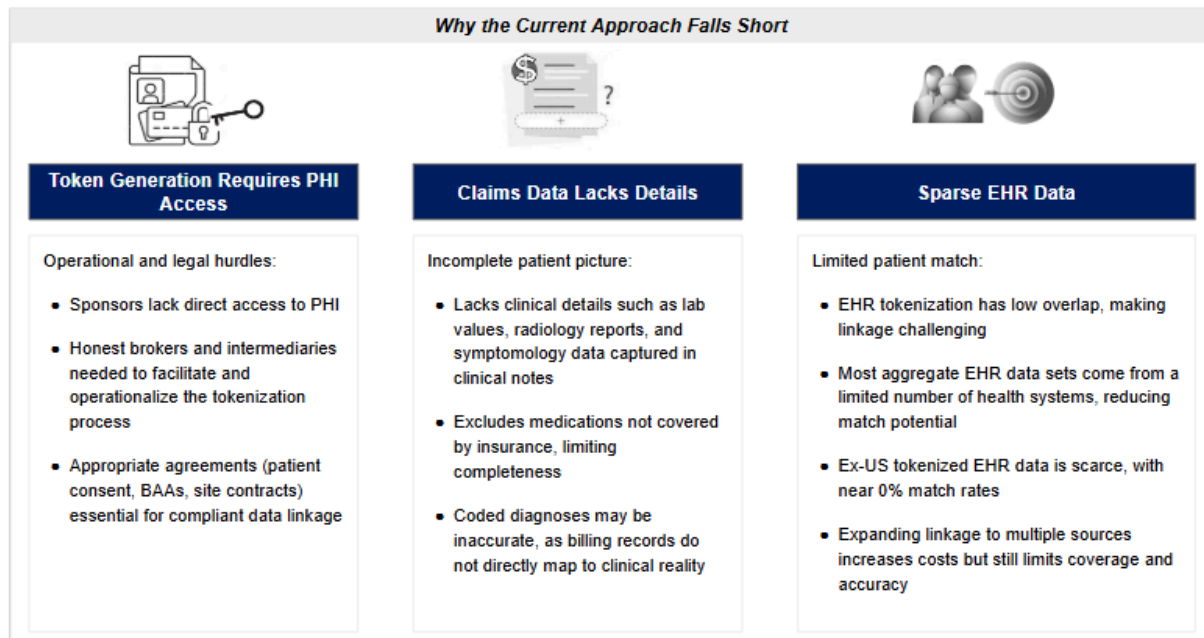
<sup>2</sup> McGuinness CB, Boytsov NN, Zhang X, Wang X, Kannowski CL, Wade RL. Probabilistic Linkage of Randomized Controlled Trial Data to Administrative Claims: A Case Study of Patients from Baricitinib Clinical Trials. *Rheumatol Ther.* 2021 Jun;8(2):793-802. doi: 10.1007/s40744-021-00302-2. Epub 2021 Apr 2. PMID: 33811317; PMCID: PMC8217382.



Currently, trial tokenization follows a structured approach involving protocol amendments and consents, tokenized linkage platforms, and claims data integration. Protocol amendments allow for follow-up data collection, with patient consent obtained during or after the trial. Tokenized linkage platforms, such as Datavant, enable secure connections between study participants and external data sources. While claims data remains the standard for post-trial follow-up, its lack of clinical detail presents limitations. EHR data, despite offering richer insights, is impractical due to low coverage, typically overlapping with only ~10% of claims data. As sponsors seek to maximize the utility of tokenization, addressing these limitations will be essential for improving long-term patient outcome tracking.

## Challenges in Trial Tokenization: Key Barriers and Considerations

Discussions with sponsors have surfaced three major challenges: generating linking tokens, the limitations of claims data for outcomes assessment, and the sparse coverage of EHR data in aggregated datasets.



## Generating Tokens Requires PHI Access

Tokenization relies on protected health information (PHI) to generate unique patient tokens. However, life sciences companies typically do not have direct access to PHI, creating logistical and regulatory hurdles. To facilitate tokenization, sites and intermediaries must be contracted to either send PHI to tokenization tools or implement on-site token generation. This requires comprehensive agreements, including patient consent protocols, business associate agreements (BAAs) with health systems or third parties, statistical determinations, and compliant data management contracts. Given that many life sciences companies and health systems lack mature infrastructure for this process, establishing the necessary distributed infrastructure for tokenization remains a significant challenge.

## Claims Data Lacks Clinical Detail

While claims data provides broad coverage, it is limited in clinical granularity, making it less effective for capturing patient outcomes. Claims datasets do not include lab values, self-paid medications, detailed imaging reports, or symptom descriptions such as 'shortness of breath.' Additionally, diagnosis codes in claims data often reflect billing requirements rather than true clinical outcomes—for instance, ICD-10 codes may

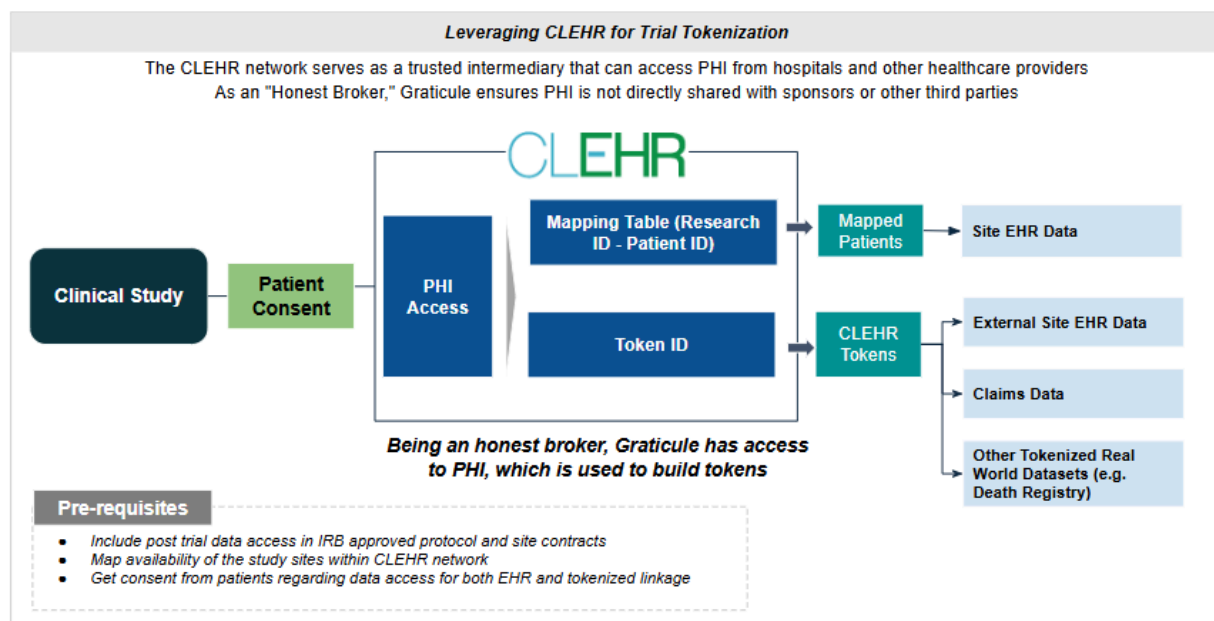
indicate a diagnostic process was performed, but the actual results remain unavailable. As a result, claims-based tokenization alone is insufficient for generating a complete and accurate picture of post-trial patient outcomes and long-term safety signals.

## EHR Data Is Too Sparse for Meaningful Trial Coverage

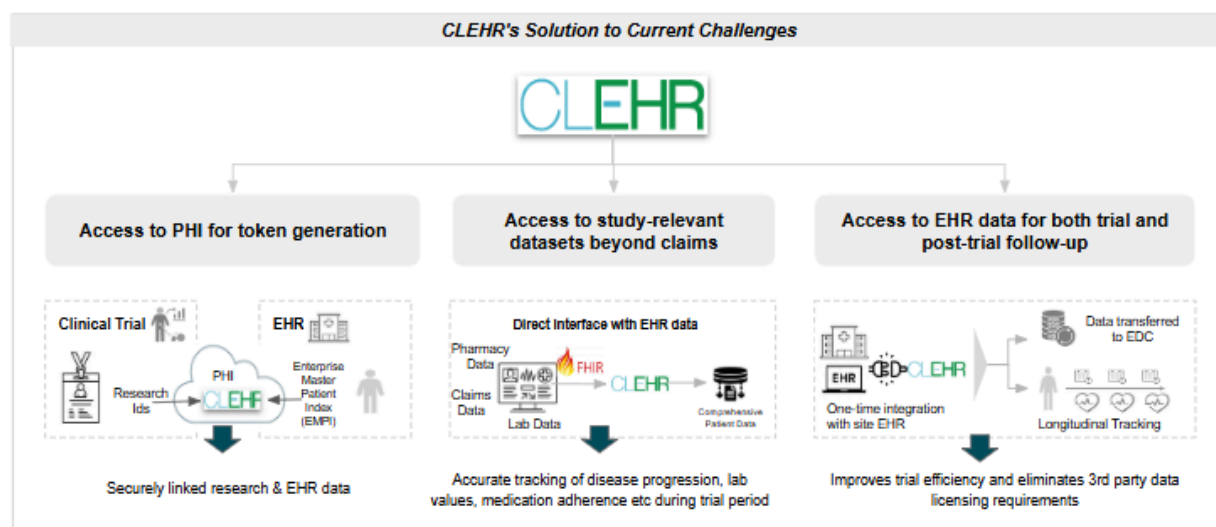
While EHR data offers the necessary clinical detail, achieving adequate patient overlap with trial participants is a persistent challenge. Most aggregated EHR datasets cover only 5-15% of any given trial cohort due to their limited geographic and institutional reach. Moreover, since EHR datasets are typically sourced from a small subset of U.S. health systems, only trial sites that overlap with the EHR licensing partner have a reasonable probability of generating a successful token match.

The challenge is even greater in ex-U.S. settings, where tokenized EHR datasets are not yet widely available. Without broad EHR coverage, tokenization is impractical for global trials unless sponsors invest heavily in linking multiple, disparate EHR sources. However, this approach dramatically increases costs while still yielding limited coverage, inconsistent longitudinal data, and challenges in standardization.

## How does CLEHR as a platform expand on tokenization?



The goal of post-trial data collection extends beyond tokenization alone. The broader objective is to maximize the investment in each clinical trial by integrating follow-up data that provides deeper insights into long-term patient outcomes. Additionally, accessing pre-trial data can offer valuable context for evaluating long-term safety and efficacy. CLEHR acts as an honest broker, directly connecting trial sites with health systems to facilitate this data capture. By leveraging a centralized approach to PHI and integrating EHR data from participating sites, CLEHR addresses key limitations of traditional tokenization methods.



## Addressing PHI Challenges in Token Generation

CLEHR streamlines the tokenization process by incorporating verification mechanisms through Clinical Research Coordinators (CRCs) or during randomization using identified data that is later deidentified. Unlike conventional methods, CLEHR maintains patient-to-study mappings within the Enterprise Master Patient Index (EMPI) of the health system via FHIR-based integration. This ensures that token generation occurs securely within CLEHR's cloud environment while maintaining compliance with privacy requirements. Additionally, CLEHR's integration with trial consent frameworks ensures that all post-study tracking adheres to regulatory and ethical standards.

Given how CLEHR streamlines tokenization, its built-in infrastructure can generate vendor-compatible tokens for tokenization vendors such as Datavant or HealthVerity,



eliminating the need for life sciences companies to contract sites or intermediaries to send PHI to tokenization tools or implement on-site token generation. Rather than competing with Datavant or HealthVerity, CLEHR complements their services by acting as an in-house intermediary within the health system, enabling tokenization for linking EHR data to external datasets through these vendors. This can be done centrally in the CLEHR cloud, where PHI needed for tokenization is managed, enabling secure, centralized tokenization that can be applied across multiple studies instead of a study-by-study bespoke approach. As a result, life sciences companies do not have to handle the complexities of facilitating tokenization. CLEHR simplifies the process, ensuring seamless data linkage while maintaining compliance and security.

### **Enhancing Outcomes Data Beyond Claims**

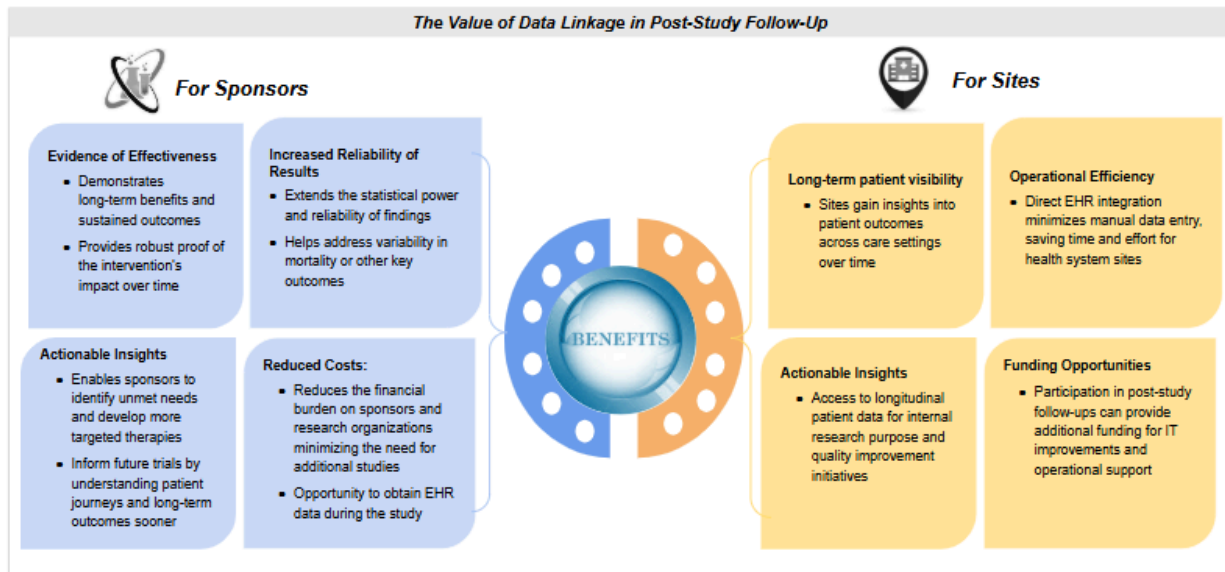
While claims data serves as the backbone of tokenized linkages, it lacks clinical granularity. CLEHR addresses this gap by directly interfacing with site EHR systems, enabling access to real-time, study-relevant clinical data such as disease progression, lab values, medication adherence, and safety signals. This richer dataset enhances Post-Approval Safety (PAS) monitoring and real-world evidence generation. Furthermore, CLEHR allows sponsors and CROs to streamline Source Data Verification (SDV) by providing direct access to EHR source data, reducing reliance on manual data reconciliation within Electronic Data Capture (EDC) tools.

### **Overcoming EHR Data Sparsity**

CLEHR mitigates one of the biggest challenges in trial tokenization—the low overlap of EHR data. By directly interfacing with the same sites where trials are conducted, CLEHR ensures that longitudinal patient tracking remains within the study's ecosystem. This eliminates the need for external EHR licensing agreements or third-party RWE vendors, simplifying both data access and trial economics. Since long-term follow-up can be integrated into trial contracts from the outset, sites can budget for extended tracking without additional infrastructure costs.

Thus, by expanding beyond traditional tokenization, CLEHR offers a scalable, integrated approach to post-trial follow-up—enabling sponsors to capture more complete, accurate, and cost-effective real-world evidence.

## Looking Beyond Tokenization: A More Integrated Future



The future of post-study follow-up with CLEHR extends beyond tokenization—it is about redefining how we track and understand patient outcomes beyond the trial period. By embedding connectivity directly at trial sites, CLEHR ensures that tokenization is not just a standalone process but part of a broader infrastructure for longitudinal patient tracking. This approach enables continuous EHR data capture, maximizing the long-term impact of clinical trial site interoperability investments.

While tokenization has gained traction in the U.S., clinical trials are inherently global, and data integration challenges remain, particularly in Europe, where tokenization adoption is still limited. CLEHR offers an immediate alternative by leveraging site-based EHR connectivity, enabling real-world data collection in European trials today. Given ongoing gaps in standardization and research network coverage across the region, Graticule's Real-World Data partnerships with health systems and data hubs in countries such as Spain, France, Italy, Germany, and Finland provide a scalable solution for enriching trial data with long-term follow-up insights.

Looking ahead, new approaches are likely to emerge, including patient-directed access to health records, expanded integration with health information exchanges (HIEs), and the incorporation of novel data sources such as digital health applications and wearable



devices. At Graticule, we are committed to helping sponsors maximize the value of their clinical trial investments. Whether piloting these capabilities or deploying enterprise-wide solutions, we are ready to collaborate with your team to drive the next evolution of post-study follow-up and real-world evidence generation.

Interested in learning more?

**Reach out to us at [info@graticule.life](mailto:info@graticule.life)**