BRS BioRes Scientia
Value your Research

Review Article **Open ⊕ Access**

Tokenization in Clinical Development-Current Trends, Challenges, and Opportunities

Aman Thukral[1]*, Kelsey Linsmeier[2], Heather Zelko[3], Sanjay Bhardwaj[4]

[1]*Director of Clinical Technology, AbbVie Inc., Clinical Data Strategy and Operations Department North Waukegan Road, North Chicago, United States.* [2]*Clinical System Design, AbbVie Inc., Clinical Data Strategy and Operations Department North Waukegan Road, North Chicago, United States.* [3]*Associate Data Scientist, AbbVie Inc., Clinical Data Strategy and Operations Department North Waukegan Road, North Chicago, United States.* [4]*Senior Director of Clinical Technology, AbbVie Inc., Clinical Data Strategy and Operations Department North Waukegan Road, North Chicago, United States.*
Corresponding author: Aman Thukral.

Abstract

Data collected for clinical research is acquired within the span of the clinical trial via eCRFs, eCOAs, and external vendor sources such as central labs. Meanwhile, Real-World Data (RWD) is continuously collected for the purposes of routine medical care. RWD is acquired throughout the entirety of a patient's life from a multitude of sources such as patient claims, pharmacy records, and primary care providers. While we request clinical sites to enter the patient's medical history data at the start of the clinical study, sites can't always capture all prior medical events. As such, there is untapped potential in leveraging RWD to supplement clinical datasets. Tokenization, the process of creating a unique and anonymized token using Personally Identifiable Information (PII), is one method of linking RWD with clinical data. Token linkage allows for greater analysis, while also protecting the contents of the original data. Tokenization in clinical trials is entering the initial stages of adoption across the pharmaceutical and medical device industry. During the pandemic, Janssen made global news for using tokenization in COVID-19 trials.

Keywords: data; tokenization; drug

Introduction

Real World Data and Tokenization

Data collected for clinical research is acquired within the span of the clinical trial via eCRFs, eCOAs, and external vendor sources such as central labs. Meanwhile, Real-World Data (RWD) is continuously collected for the purposes of routine medical care. RWD is acquired throughout the entirety of a patient's life from a multitude of sources such as patient claims, pharmacy records, and primary care providers. While we request clinical sites to enter the patient's medical history data at the start of the clinical study, sites can't always capture all prior medical events. As such, there is untapped potential in leveraging RWD to supplement clinical datasets. Tokenization, the process of creating a unique and anonymized token using Personally Identifiable Information (PII), is one method of linking RWD with clinical data [1]. Token linkage allows for greater analysis, while also protecting the contents of the original data [1]. Tokenization in clinical trials is entering the initial stages of adoption across the pharmaceutical and medical device industry. During the pandemic, Janssen made global news for using

tokenization in COVID-19 trials [2]. However, tokenization is not the first use of blockchain-based digital assets. Digital assets have been widely used in other industries in applications such as cryptocurrency. Tokenization is gaining momentum in clinical trials due to the applications in drug and medical device development. The use of real-world patient health data allows for a holistic view of a patient's health before, during, and after they participate in a trial. Having access to this data enables pharmaceutical companies to continuously monitor the effectiveness and potential adverse effects that may not be reported to the investigator. The use of this data and its applications will be further explored in the following sections.

Potential Applications in Drug Development

Token usage in clinical trials presents many applications in drug development, specifically in recruitment, long- term follow-up, and data integrity. When applied, tokenization has potential to decrease the financial burden on pharmaceutical companies and allow for greater trust between regulatory agencies, patients, and pharmaceutical sponsors.

Data integrity is a pillar of the pharmaceutical industry. The use of tokens in clinical trials allows for linking trial data to RWD by creating a unique identifier that is different from the number assigned to each patient within the study [7]. The unique token identifier prevents patient data duplication, ensuring patients cannot enroll in the same clinical trial at different sites or enroll in a trial before completing a mandatory wash-out period. Additionally, each token linkage is marked as a unique transaction that can be audited to ensure data duplication does not occur [7]. Furthermore, integrating tokens into clinical trials allows for a more comprehensive patient health map. With tokens linked to RWD, more detailed information on adverse events (AEs), relapse rates, and survival outcomes can be tracked and analyzed [8]. Additionally, factors related to health economics research can be included in trials. When analyzing adverse events, additional information such as healthcare visits, pharmacy claims, and overall health status can be evaluated to provide additional insights. For example, a patient may visit a primary care provider for an adverse event but may not report this information to the trial sponsor. By having access to the primary care provider data, these events can be included with trial data for a complete representation of AEs. More detailed patient history can also be linked to trial data. This may allow the sponsor to better determine whether certain pre-existing conditions are contributing to adverse events seen during a trial. Finally, the use of RWD may allow the sponsor to follow patients' health after the conclusion of the trial, informing regulatory agencies of the long-term effects of a drug without the need for patients to return for on-site visits or for sponsors to run financially burdensome Phase 4 studies.

One of the most challenging aspects of a clinical trial can be recruiting patients that meet eligibility criteria. When evaluating a drug in long-term chronic disease, currently, only 10-20% of the population is willing to participate in a clinical trial [9]. With low participation, it is important to target regions and sites with large patient populations. Pharmaceutical companies can strategically target these areas and reduce the number of sites by using token data to identify regions or institutions with higher instances of disease. Since the overhead to open a site is quite large, this may reduce the cost of running a clinical trial by limiting low enrolling sites. Better yet, RWD insights can be used to strategically target areas with diverse patient populations, potentially allowing sponsors to increase patient representation.

Learnings from Pilot Studies

AbbVie's tokenization pilot consists of three stages – Stage 1: token generation pilot, Stage 2: portfolio token generation (all studies), and Stage 3: token application. The scope of Stage 1 included tokenizing eight clinical studies within the Immunology, Oncology, and Neuroscience therapeutic areas. Studies were eligible to participate if they were in study start-up, had sites located in the US, and had the potential to generate more evidence from Real-World Data. The extent of tokenization remained solely within the United States. Many learnings were gathered in the process of preparing for and implementing tokenization in Stage 1.

Overall Tokenization Process

Tokenization starts when the site enters a patient's Personally Identifiable Information (PII), such as name, birthdate, and zip code, into an external token portal along with the sponsor trial ID and patient ID. This information is used to assign a unique encrypted identifier, otherwise known as a token. For most patients, this token may already be assigned to the individual outside the context of the clinical trial. This data includes sources such as medical and pharmacy bills, insurance claims, and electronic medical records (EMRs). As Real-World Data is collected, it is encrypted by external parties and transferred to a large data lake. Once the token is linked to the sponsor trial ID and patient number, it can be used to locate corresponding RWD records. At any point, the sponsor may request transfer of the encrypted RWD, which can be used to supplement clinical data sources for enhanced analysis. With most encryption processes used, the token has an extremely low risk of being reversed [5].
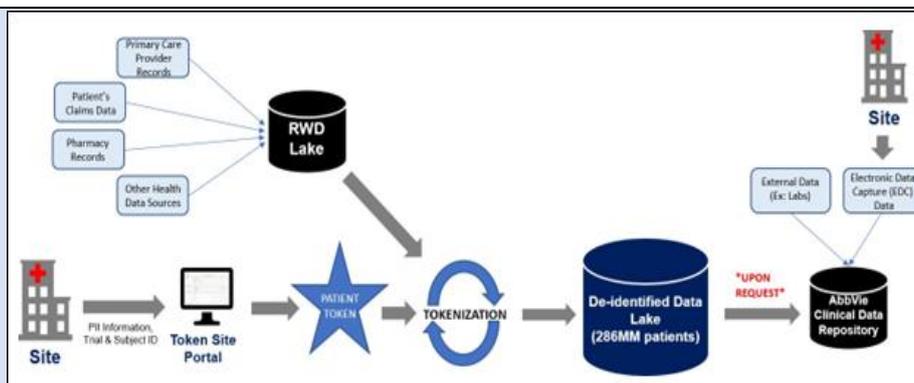
Figure 1: High-level data flow of token generation and linkage to Real-World Data

PII Collection – Internal vs. External

Sponsors should consider their PII data collection policies when designing a tokenization process. Some sponsors may be comfortable acquiring and storing PII information in-house, especially within commercial and medical affairs organizations. However, other sponsors may rely on third-party vendors instead. The AbbVie tokenization project team first explored using electronic Case Report Forms (eCRFs) within the Electronic Data Capture (EDC) system to collect token inputs. Ultimately, they chose to leverage a third-party vendor as it better aligned with internal legal policies and decreased risk and liability. The project team will continue to evaluate new ways to generate tokens. Nonetheless, PII collection and storage will remain outsourced for the near future.

Informed Consent

A large part of successfully introducing tokenization includes ensuring that individuals are educated and informed on the process and its benefits. Many people may not be familiar with the term tokenization, especially in the context of health. To ensure that patients comprehend the tokenization process, sponsors should consider developing standard informed consent language which outlines the token lifecycle from start to finish. Standardization can create consistency and ensure high quality results, while also minimizing site-specific change requests and reducing IRB questions. Additionally, teams should determine whether consent language is baked into the main informed consent or whether it will stand alone as a separate, optional consent. In situations where a patient population may be more likely to deny participation due to tokenization, sponsors may want to pursue optional consent. On the other hand, there may be benefits of including tokenization in the main consent if enrollment is not a concern. On initial AbbVie pilots, tokenization consent was made optional out of caution for participation impacts. However, this approach led to lack of consent and lack of site buy-in. Some sites opted out of the tokenization process on a site-level due to wariness of increased workload. However, other sites that opted in still experienced low consent. The latter scenario led to increased risk for protocol and compliance deviations, as some sites preemptively set up the patient for tokenization despite the denial of consent. To combat these issues, tokenization consent was made mandatory for all US patients moving forward and included within the main consent. With the change in consent strategy, an increase in participation was observed without impacting enrollment. However, we advise each sponsor to carefully consider which strategy is best suited for their trial.

Protocol Language

To ensure regulatory compliance and transparency, sponsors should ensure tokenization information is included in the study protocol. The language should outline the types of PII and de-identified data that will be collected and how the data types will be used (as seen in Table 1). Sites can use this information as a reference tool to help answer patient's questions during the consent process. The protocol should clarify whether the sponsor and/or a third- party vendor will be able to see the encrypted patient token and which party will link the token to Real-World Data. The protocol language at AbbVie clarifies that the sponsor will not have access to patient identifiers and will only be able to view aggregate data at a study level. The distinction between AbbVie and the third-party vendor allows for the protection of patient privacy. Sponsors should also state in the protocol language whether participation in tokenization is mandatory or optional and which countries will participate.

Table 1: Description of PII and Other De-Identified Data Sources

Personally Identifiable Information (PII)	Other Sources of De-Identified Data
Required:	Medical Claims
First Name	Pharmacy Claims
Last Name	Hospital Charge Master Records
Gender	Behavioral and Demographic Information
Data of Birth	Electronic Medical Records
Zip Code	Long-term Care Pharmacy
Optional:	Laboratory Data
Insurance ID	
State Code	

Site & Internal Training

Ensuring proper external and internal training is another key factor to implementing tokenization successfully. When sponsors select a vendor, they should evaluate the vendor's training resources for the sponsor and site. Sponsors should also consider how sites will be trained on this information; for example, will this training be conducted by a third-party vendor, at an investigator meeting, or by the sponsor's Clinical Research Associates (CRAs)? In AbbVie's case, the CRAs were trained by the vendor and then given resources to subsequently perform site trainings independently. The vendor also partnered with the pilot project team to create training resources for the sponsor's internal use. Many study teams were not familiar with tokenization, so there was a mass effort to raise awareness on the benefits of tokenization and the overall project status. As more study teams became familiar with tokenization, patient and site skepticism seemed to decrease as well. This proved that without adequate training, trials were more likely to run into issues with site and patient buy-in.

Data Reconciliation

Sponsors will need to consider how to reconcile the token information collected. Whether using a third-party vendor or internal systems for PII collection, teams will still need to cross-verify information between data sources. If a patient is tokenized and does not have token consent record or vice versa, this should be flagged as a potential discrepancy. Without reconciliation, there is risk that patients could be tokenized without providing consent. Initially, the process of reconciliation between the token portal and consent records was a manual process at AbbVie. However, AbbVie has since programmed edit checks to automatically flag discrepancies between sources. The edit checks are included within the study-specific Data Review Plan and actioned on a regular cadence.

Data Certification

The ownership of data certification is the responsibility of the vendor providing tokenization services. To comply with HIPAA in the United States, tokens must be de-identified through the Safe Harbor de-identification process or the Statistical de-identification process. When using the Safe Harbor method to de-identify data sets, valuable information is stripped from the dataset, ensuring that the data cannot be re-identified to the patients [6]. However, this method may strip a significant amount of data needed to allow for meaningful analysis [6]. Alternatively, the Statistical de-identification process retains valuable information, allowing for more complete statistical analyses to be conducted. Once statistical analysis has been completed, identifying health elements are removed from the data set, but information such as service provider dates are still present [6]. The data certification process then requires that a statistician or HIPAA certified professional has reviewed the dataset and determined that enough identifying information has been removed to ensure the risk of re-identification is minimal [6]. Although the risk of potential re-identification is higher than the Safe Harbor method, the risk is still very low when adequately certified [6]. When considering a vendor for tokenization services, sponsors should evaluate which method of Data Certification the vendor provides and ensure that the method being used is HIPAA compliant.

Preliminary Conclusion

AbbVie has completed the pilot phase of the project (Stage 1), with overall participation from eight studies and 199 US patients. Initial pilot successes included reducing re-identification risk through the use of a third-party vendor and creating robust training and protocol language that increased awareness and understanding. Areas for improvement included enhancing the data reconciliation process and re-evaluating the consent strategy to boost participation.

Finally, we learned the criticality of ensuring that tokens are HIPAA certified and were able to select a vendor that used the certification process that better fit AbbVie's needs. While these learnings are specific to AbbVie studies, other sponsors may use these examples when considering what is right for their studies. At this point, AbbVie has not realized the benefits of tokenization. However, the company is confident that the framework developed will successfully support the collection and analysis of tokenized data later on.

Further Research

AbbVie plans to continue into Stage 2 of the project, in which company-wide implementation will begin. Prior to beginning Stage 2, AbbVie will evaluate the ability to manage tokenization on a large scale within the United States. For this to occur effectively, all risks and benefits from the pilot must be reviewed. Upon completion of stage 2 and commencement of Stage 3, AbbVie will begin requesting the tokenized datasets from the vendor to allow for anonymized cohort analyses. By linking clinical data to Real-World Data using tokenization, AbbVie is hopeful that additional insights can be gained on patient populations in areas such as recruitment, long term follow-up, and fit for purpose evidence. Once the full project has completed, AbbVie will also determine the cost-effectiveness of using a third-party vendor and if this process can be efficiently done in-house while maintaining regulatory requirements.

References

1. Boldyreva, A. S., & Grubbs, P. (2018, April). Tokenization vs encryption. *McAfee.*
2. Moline, H. L., Whitaker, M., Deng, L., et al. (2021). Effectiveness of COVID-19 vaccines in preventing hospitalization among adults aged ≥65 years — COVID-NET, 13 states, February–April 2021. MMWR. Morbidity and Mortality Weekly Report, 70(32):1088–1093.
3. Langley, P. C., & Martin, R. E. (2018). Blockchains, property rights and health technology assessment in the pharmaceutical and device industries. *Innovations in Pharmacy*, 9(4).
4. Silva, J. S. (2002). HIPAA administrative simplification standards: Provisions of the final privacy rule related to clinical trials. In Cancer informatics: Essential technologies for clinical trials. *Springer*, 206–211.
5. Rockhold, F., Bromley, C., Wagner, E. K., & Buyse, M. (2019). Open science: The open clinical trials data journey. Clinical Trials, 16(5):539–546.
6. (2018). Overview of Datavant's de-identification and linking technology for structured data. *Datavant.*
7. Maslove, D. M., Klein, J., Brohman, K., & Martin, P. (2018). Using blockchain technology to manage clinical trials data: A proof-of-concept study. *JMIR Medical Informatics*, 6(4):e11949.
8. Yaeger, K., Martini, M., Rasouli, J., & Costa, A. (2019). Emerging blockchain technology solutions for modern healthcare infrastructure. *Journal of Scientific Innovation in Medicine*, 2(1).
9. Armitage, J., Souhami, R., Friedman, L., et al. (2008). The impact of privacy and confidentiality laws on the conduct of clinical trials. Clinical Trials, 5(1)70–74.