|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | ***21132 - Clinical Data Repository (CDR) Audit – Data Protection Walkthrough***  **Date: 07/13/2021**  **Physical Location: WebEx** | | |
|  | | |  |  |  |
| **Attendees** | | | | | |
| **Customer** | | **Internal Audit Department** | | | |
| Kishore Puvvala – Lead Data Engineer  Lisa Larsen – Director, Enterprise Systems  Olivia Lawson – Intern Associate | | Sol Vazquez – Manager, Internal Audit  Jason Nazare – Advisor, Internal Audit  Seun Mafi – Senior Consultant, Internal Audit  Aerozona Obiadazie – Senior Consultant, Internal Audit  Tyrell Jarett – Audit Consultant  Moriah Striegel – Intern Associate. | | | |
|  | | | | | |
|  | | | | | |
| **Business Unit** | Data Protection Team | | | | |
| **Process** | Data Protection | | | | |
| **Process Owner(s)** | Kishore Puvvala – Lead Data Engineer | | | | |
| **Policies and Procedures** | N/A | | | | |

|  |
| --- |
| **Purpose of the process walkthroughs** |
| The purpose of this walkthrough is to get a complete understanding of the data governance process completed by the Clinical Data Repository (CDR) team as well as identify all systems used, reports used / generated. |

|  |  |
| --- | --- |
| **Roles & Responsibilities of the Personnel involved in the process** | |
| **Role** | **Responsibilities** |
| Kishore Puvvala | Ensure that all data within CDR is protected both in transit and at rest. |

|  |  |
| --- | --- |
| **Supporting Systems** | |
| **System Name** | **System Description** |
| JIRA | Ticketing system used to ingest data into CDR |
| Service Now | Cloud system used to store and access tickets |
| SQL | Domain specific language used in programming for managing data held in CDR |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Key Source Reports used within the processes** | | | | | |
| **Report Name** | **Report Frequency** | **Report Description** | **Origination System** | **Report Owner** | **Report Example (embedded)** |
| *N/A* |  |  |  |  |  |

|  |
| --- |
| **Process Walkthrough** |
| A high-level overview of Data protection involves the data protection team checking to see if there are any PII data, making sure it is encrypted at rest, data on disk is always encrypted, amongst other security measures. It was also noted that data in transit should always be secured. PII information through any tables is encrypted and secured. Kishore Puvvala, lead data engineer; noted that CDR is an operation use case, and as a result PII data is not encrypted per operation use case exception; however, all other data is encrypted.  For every data that is brought into CDR, a data injection ticket is used. The data team use the data dictionary to review this information. As part of the ingestion process, they tag any column as PII column. The whole process including the targeted items are part of the JIRA ticket. The Data Protection team owns the process automation aspect. They document all the different types of tickets. Kishore noted that in the cloud, they team uses Service now rather than JIRA.  Any data in CDR goes through an API, with no SQL data analytics. When data is pulled from the file into SQL, data encryption tools are then implemented. Decrypting and querying data over the consumer does impact the SLAs, and for this reason, the team does not utilize encryption/decryption for operational SLAs. Every data is taken care of from the API gateway which is used by operation enterprise.  The platform team has different operational groups and other teams supporting them. They manage the data security aspect which include disk level security and data level security. CVS is the owner and handled by the platform team. Two groups the data protection team works with very closely include Cloud engineering, and cloud security. Data protection team follow the security standards established by them. Kishore noted that his team holds the private keys, whereas the public key is shared with Google. In the Google realm, there are buckets (Similar to a file system) and database. They are configured in the form of tables and columns. When new buckets and databases are created, the keys used are encryption cases. For database query, data sets and databases at a moderate level, encryption is applied; whereas, for big queries, they apply column level security handled by the enterprise security team which is a 3rd level security. |