|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | ***21132 – Clinical Data Repository (CDR) Audit – Data Integrity and Protection Walkthrough***  **Date: 07/12/2021**  **Physical Location: WebEx** | | |
|  | | |  |  |  |
| **Attendees** | | | | | |
| **Customer** | | **Internal Audit Department** | | | |
| Tushar Mohapatra – Director, Data Engineering  Lisa Larsen – Director, Enterprise Systems  Kishore Puvvala – Lead Data Engineer  Olivia Lawson – Intern Associate | | Sol Vazquez – Manager, IT Internal Audit  Jason Nazare – Advisor, IT Internal Audit  Aerozona Obiadazie – Senior Consultant, IT Internal Audit  Tyrell Jarett – Audit Consultant, Internal Audit  Moriah Striegel – Intern Associate, IT Internal Audit  Maria Braun – Senior Consultant, Internal Audit | | | |
|  | | | | | |
|  | | | | | |
| **Business Unit** | Data Integrity Team | | | | |
| **Process** | Data Integrity | | | | |
| **Process Owner(s)** | Tushar Mohapatra – Director, Data Engineering | | | | |
| **Policies and Procedures** | N/A | | | | |

|  |
| --- |
| **Purpose of the process walkthroughs** |
| The purpose of this walkthrough is to get a complete understanding of the data integrity 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** |
| Tushar Mohapatra | Ensure thar all data within CDR is complete and accurate both in transit and at rest |
| Kishore Puvvala | Ensure that all data within CDR is protected both in transit and at rest |

|  |  |
| --- | --- |
| **Supporting Systems** | |
| **System Name** | **System Description** |
| RxConnect | Where clinical data is stored |
| Splunk | Cloud based management software used to store the monitoring logs captured by Google Cloud Logging |
| ServiceNow | Cloud system used to store and access tickets |
| Google Cloud Platform | Cloud based platform where CDR data is stored. Also used for monitoring access and tracking outliers and sends logs to Splunk. |

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

|  |
| --- |
| **Process Walkthrough** |
| In the FHIR store the file specification shows how the data is going to be established or linked, the objective of the CDR is to create the detailed information about a patient medical record. There is a clear FHIR specific documentation that the CDR team is following and the Google Cloud Platform (GCP) API is used.  The foundational layer is called the Provenance, it gives the information of data source; data agent types, transmittal or author of the data, or from clinic office, where it was recorded, all this information are kept in the Providence. Each account is accountable to the Provenance and to the Metadata.  Records are sent to CDR and reordered in Splunk, all the events and accounts goes to Splunk. Data failure occurred last week, when a data comes in it goes through a basic validation, they check to see if it is an empty file or a schema. Each file has an expectation, if the schema is not what is expected it rejects the data and then splits into the bad ones, which goes through re-processing. The good one’s scales through.  From the core pipeline, data is gotten from the data sources which goes through a validation process, then parse and convert to naïve JSON bundle then a transformation and re-conciliation called harmonization is done on the data then passes it to the FHIR store which is a Google managed platform service.  The database is a proprietary called Spanner Database, the CDR team does not have access to the Spanner but they have access to the FHIR store which is a managed service using a multiple way through APIs.  There is an electronic notification mechanism, as part of the pipeline, there are 2 frameworks. One is the Common Ingestion Framework 2.5 (CIF) its functions is to get data from external sources and put it in the GCP area, it could be on-prem to cloud, cloud to cloud, on-prem to dump server etc. After it arrives at the doorstep of the cloud from CDR there is another framework called CHF (Common Harmonization Framework) which takes the data from the landing area to the FHIR store, within that framework there is a notification mechanism, logging mechanism, monitoring mechanism, alerting, depending on the data type and should reach consumer within 1-3 minutes. Notification mechanism for now is on manual mode, but there is a plan in place to automate soon. |