Fall 2023

Operationalizing the DataGauge Framework in a Health Information Exchange Utilizing Hepatitis C Data

Edward Yao
University of Texas Health Science Center at Houston, Edward.Yao@uth.tmc.edu

Operationalizing the DataGauge Framework in a Health Information Exchange Utilizing Hepatitis C Data

A
Translational Project Paper

Presented to the Faculty of
The University of Texas
Health Science Center at Houston
School of Biomedical Informatics
in Partial Fulfilment of the Requirements for the Degree of
Doctorate in Health Informatics

By

Edward Yao, M.S., B.S.

The University of Texas Health Science Center at Houston

2023

Translational Project Committee:

Angela Ross, DNP, MPH, PMP, DASM, PHCNS-BC, FHIMSS[1], Advisor
Debora Simmons, PhD, RN, CCNS, FAAN[1]
Tiffany Champagne-Langabeer, PhD, RDN, FHIMSS[1]

[1]The School of Biomedical Informatics

Table of Contents

Table of Figures

Table of Tables

Abstract

This project aims to implement the DataGauge framework in a health information exchange (HIE) setting as a proof of concept. The modified DataGauge framework, described by Diaz-Garelli et al. (2019), is utilized to test its functionality and applicability with any dataset. The specific objective of the project is to determine the number of hepatitis C-positive tests within the HIE. The implementation involved multiple iterations following the DataGauge framework's steps for data extraction and analysis. Five iterations were conducted, resulting in both successful and failed queries based on the validity of the data standards. The findings revealed that the HIE, in this case, did not have complete access to the clinical data required to answer the initial question about the number of hepatitis C-positive patients; rather, the HIE only received information from patients who consented to share their health data and were approved by their physicians. To address this limitation, a recommendation is proposed based on Guerrero et al.'s (2019) workflow. The recommendation suggests granting an intermediary actor (referred to as the analyst) access to all clinic data, regardless of patient consent status. The analyst would then gather and deidentify the relevant clinical data, with explicit permission from the clinic, and provide it to the Rio Grande Valley HIE (RGV HIE). This approach would enable the RGV HIE to legally access non-participant data through deidentified datasets or aggregated count/sum data, while ensuring compliance and collaboration. By implementing this recommended process, the RGV HIE can enhance its preparedness for future clinical questions, grants, partnerships, and public health emergencies. Moreover, this model can be applied to other data warehouses and HIEs nationwide.

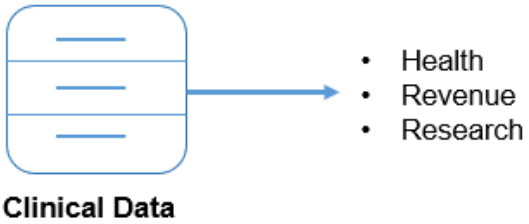
Keywords: clinical entity, health information exchange, HIE, Rio Grande Valley, RGV HIE, iteration, audit, DataGauge, hepatitis

Section 1: Introduction

According to the book Biomedical Informatics by Shortliffe and Cimino (2021), data plays a vital role in various aspects of healthcare, including clinical, administrative, financial, and public health operations. Figure 1 provides an abstract representation of the benefits of clinical data as explained by Shortliffe and Cimino (2021). The visual portrays clinical data, as a server, with an arrow extending towards potential advantages such as healthcare revenue and research. In the clinical setting, data is used for making informed decisions, enabling physicians to rely on information such as lab results, triage assessments, and diagnoses over time. This clinical data can be leveraged within a clinical institution to enhance quality improvement metrics and facilitate better coordination among department staff and administration. Moreover, visit and billing data can be utilized to estimate costs. On a broader scale, population-level clinical data can drive research and inform public health policies as health trends evolve. The capture, extraction, and analysis of clinical data have evolved from paper to electronic records but remain integral to healthcare practices today. The importance of data in healthcare cannot be overemphasized; therefore, the impact of lost or altered data can similarly not be overemphasized.

Figure 1

Clinical Data Benefits



Issues in Data Extraction

Data management issues precede the modern era of electronic platforms. Feinstein (1969) provides an early description of capturing clinical data, noting that the arduous task of manual data extraction occurred through manual abstraction. In the context of Feinstein (1969) and this paper, manual abstraction involves an assembled, dedicated staff that audits records and documents pertinent data in a structured (often tabular) format. This process is time-intensive and prone to human transcription errors. After the passing of the Health Information Technology for Economic and Clinical Health (HITECH) Act in 2009, the law mandated the implementation of electronic health records (EHRs), and those entities who did not follow the statute incurred penalties of a maximum of $1.5 million (U.S. Department of Health and Human Services [HHS], 2021). With the automated reporting of EHRs, data can be extracted significantly quicker than with paper records (Paul et al., 2015); however, despite the reduced time taken to extract data, the process of ensuring that the data complies with data quality standards—before and during the extraction—is also time-intensive and prone to human errors. Some examples of issues in data extraction are outlined in the following sections.

Issue 1: Data Entry Errors

Manual translations depend on various factors, such as the verbiage used by the clinician. If the verbiage in the free text fields contains imprecise language—or if the abstractor perceives it to be imprecise—translations may be hindered. Moreover, misentering information remains a constant danger in human transcription, even in scenarios where the correct data is present.

In the case of electronic data extraction, there may be issues in initially mapping the needed data variable to the data variable stored on the initial server and to other servers with different data standards. One example is if a reporting standard called for abnormal levels of

hepatitis C antibodies; in that case, there would be inconsistencies across multiple clinical institutions with different labs, as they may have different thresholds for abnormal values. Humans both write and interpret reporting standards, which dampens the objectivity of the mapping and extraction processes (Dean et al., 2009).

Issue 2: Labor Shortages and Cost

A survey conducted by Blumenthal (2017) of 152 programs within the National Quality Registry Network found that most registries had three full-time employees and that 88% of said registries relied on manual abstraction for their data collection. In another survey, Islami et al. (2021) questioned nine managers in the National Program of Cancer Registries, and they identified staffing and available technological aid as the limiting factors in their program. Consequently, staffing limitations hinder the number of manual abstractions, and limitations in technological support similarly restrict the ability to provide support work for extractions, such as mapping variables and coding extractions to the appropriate registry.

Issue 3: Disconnected Data

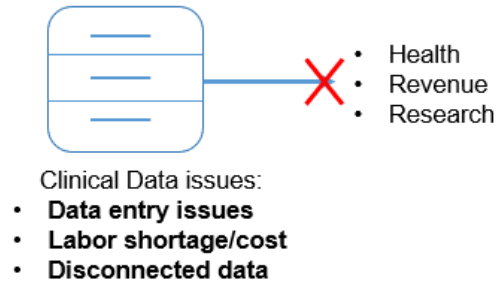
Important information, such as lab and specialist reports, may not be accessible to support a diagnosis sought for a particular patient (Chen et al., 2014). Missing or unavailable data remains an issue for both manual abstractions and electronic extractions (Cowie et al., 2016); for example, data pertinent to a cardiology registry may be found within the cardiology department and within the radiology, outpatient, and intensive care units. If the systems between the various departments are not interoperable, electronic abstraction and manual abstraction would be hindered due to the addition of multiple security credentialing and protocol steps, as well as the translation and mapping of data schemas.

Lack of Quality Assurance: Problems and Solutions

The presence of data issues can have significant negative implications for health, revenue, and research. Inaccurate data can easily lead to erroneous conclusions, resulting in suboptimal decision-making and unfavorable business outcomes. Figure 2 further demonstrates the impact of such failed data, displaying the relationship between data failures and subsequent shortcomings in service delivery. The abstract visualization in Figure 2 represents clinical data using a server, highlighting issues such as data entry problems, labor shortages and costs, and disconnected data. As a result, these issues can lead to a lack of quality in health, a loss of revenue opportunities, and a dearth of research opportunities (Arts et al., 2002; Chen et al., 2014; French & Mykhalovskiy, 2012). Moreover, flawed data carries the potential for both the loss of patient lives and financial value (Arts et al., 2002; Chen et al., 2014; French & Mykhalovskiy, 2012). Establishing a well-defined process to tackle these data issues is of utmost importance, as their failures can have wide-ranging and significant consequences (Arts et al., 2002). These repercussions extend to both the financial stability of clinical institutions and the well-being of patients (Chen et al., 2014; French & Mykhalovskiy, 2012).

Figure 2

Impact of Data Quality Issues

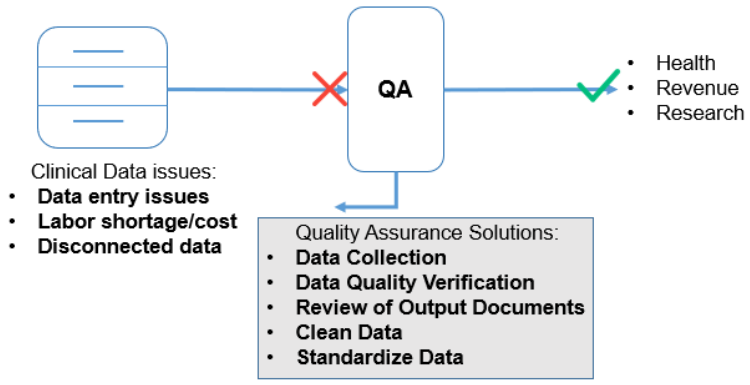


Nonetheless, the implementation of a quality assurance (QA) method can effectively mitigate the adverse impacts caused by potential data quality issues. QA is a comprehensive

process that addresses both social and technical aspects. It involves automated detection of abnormal data values as well as manual audits of technical systems and human processes. By vigilantly monitoring and diligently practicing QA, any data issues can be promptly detected and rectified (Arts et al., 2002; Chen et al., 2014). Figure 3 illustrates the relationship between clinical data issues and their impact on benefits, revenue, and research. QA plays a pivotal role in addressing data issues related to data entry, labor shortages, and disconnected data. Some examples of solutions employed by QA are data collection standardization, data quality verification, and reviews of outputs. These measures aim to ensure clean data through standardization (Arts et al., 2002; Chen et al., 2014). The implementation of a QA process yields improved programs, leading to enhanced health outcomes, increased revenue, and expanded research opportunities (Arts et al., 2002; Chen et al., 2014).

Figure 3

QA Relationship with Data



To successfully implement a QA framework, it is crucial to carefully select a suitable model (Diaz-Garelli et al., 2019). Figure 4 visually depicts the relationship between problem-solving and the application of an evidence-based QA model or framework. QA solutions can address issues such as data entry errors, labor inefficiencies, and disconnected data. However, it

is important to note that the implementation of these solutions through a QA framework is the next critical step. Figure 4 emphasizes that simply acquiring knowledge of the solution is not enough; true distinction lies in its actual implementation (Diaz-Garelli et al., 2019; Zafar et al., 2018). This ongoing project necessitates the implementation of a solution, underscoring the importance of an evidence-based QA model or framework. Consequently, the objective of this endeavor is to identify and adopt a QA model or framework that is substantiated by evidence.

Figure 4

Solution and Implementation Relationship

Project Setting

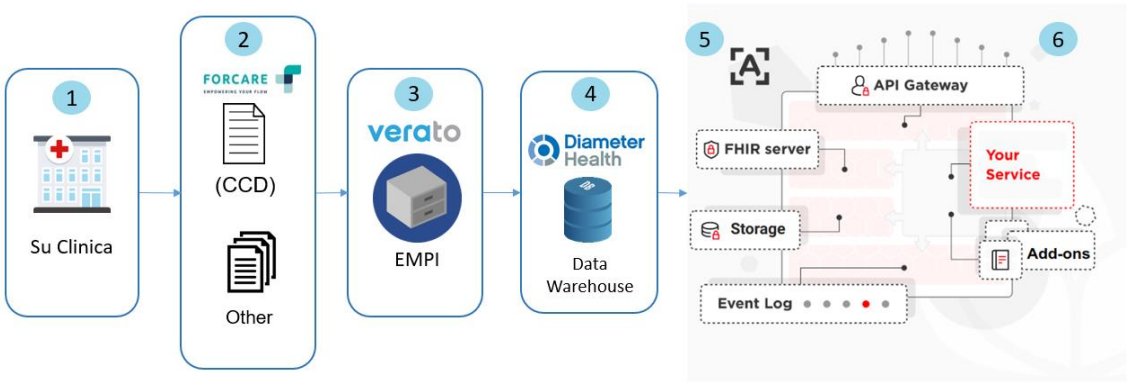
The Rio Grande Valley health information exchange (RGV HIE) is a 501(c)3 non-profit organization that aggregates and delivers clinical data to aid in analysis for academic research, public health reporting, grants, and metric reporting. The RGV HIE currently serves nine counties in South Texas; it ingests data from eight hospitals and 30 clinics within the RGV region and receives an average of 5,000 continuity of care documents (CCDs) per day (RGV HIE, 2023). Figure 5 illustrates the data export, transformation, and storage process in the RGV HIE. The following points provide a detailed explanation of each stage:

- In stage 1, data is housed and maintained within a clinical entity's EHR, which houses clinical data. The clinical entity will be Su Clinica, a clinic partner that

- In stage 2, data is exported as CCDs and other supporting documents and stored with the assigned vendor, Forcare.
- In stage 3, the vendor—Verato—creates an enterprise master patient identifier (EMPI) that serves as the unique identifier for that patient across all documents received from the various clinics.
- In stage 4, Diameter Health standardizes the data so that it can be stored and retrieved through a user interface.
- In stage 5, the data is also sent to Health Samurai's Aidbox fast healthcare interoperability resources (FHIR) server, which translates the available data points from the data warehouse to the FHIR schema. Data on the FHIR server can be interfaced by third parties via the available application programming interface (API; Sharma & Aggarwal, 2018).
- In stage 6, the data is finally sent to a structured query language (SQL) database. Data in the SQL database can be accessed in a standardized manner to allow for human querying but it can also be accessed through interfacing with other applications and automated vendor services (Upadhyay & Upadhyay, 2020).

Figure 5

The Data Export, Transformation, and Storage of Data in the RGV HIE (Yao, 2020)



The RGV HIE currently has five employees in the following roles: executive director, information technology (IT) analyst, data quality analyst, application support analyst, and technical clinical analyst. Inquiries and connections to utilize the RGV HIE's data are conducted in an ad hoc manner with no standardized workflow to evaluate data quality. Efforts to evaluate data quality rely solely on vendors, and the RGV HIE does not have a vendor-independent data QA process.

The Current Problem

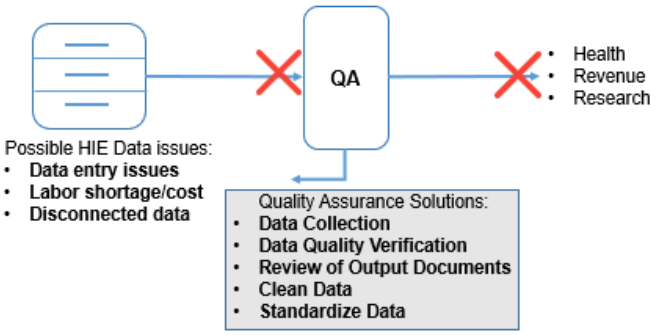
HIEs serve as a repository for structured healthcare data, enabling the analysis of a region's health information (Rudin et al., 2014). However, the RGV HIE failed to fulfill this purpose during the COVID-19 public health emergency due to its insufficient workflow and infrastructure for extracting novel clinical information. To address this limitation, I will use hepatitis C measurements as a proxy for any lab test, including COVID-19, to improve data quality within the RGV HIE. Hepatitis C is a liver disease caused by the hepatitis C virus, which can result in liver function loss (Sinn et al., 2014). The choice of hepatitis C as a proxy

measurement is justified by its relatively low incidence rate of approximately 40 per 100,000 people per year in the United States, reported by the Centers for Disease Control and Prevention (CDC, 2018), compared to the high incidence rate of approximately to a low estimate of 18,250 per 100,000 people per year for COVID-19 (Xu et al., 2022). This low incidence of hepatitis C compared to COVID1-19 allows for a manageable manual audit within a reasonable timeframe, estimated to be less than a year. By employing a proxy measurement that effectively represents other measurements within the exchange, we ensure the generalizability and applicability of the process and workflow used to assess data consistency within the RGV HIE.

Figure 6 illustrates how the absence of a QA process within the RGV HIE leads to inconsistent and unreliable data flow, ultimately hindering its overall effectiveness. All the aforementioned issues make it difficult to extract clinically relevant data both manually and electronically in a logistically feasible manner, thereby impeding essential public health reporting (Chen et al., 2014; Cowie et al., 2016). If a QA process is not implemented, low data quality will hinder the potential benefits associated with usable clinical data (Chen et al., 2014; Cowie et al., 2016). Consequently, the initial inquiries made by researchers in the UTHealth School of Public Health (UTH SPH) to conduct such monitoring and obtain data from the RGV HIE were unsuccessful.

Figure 6

Lack of QA Leads to Lack of Data Quality



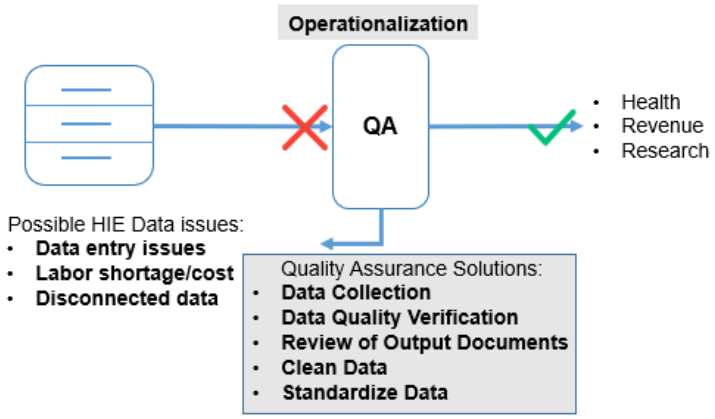
In the case of future public health events, the RGV HIE should be the primary organization to provide data to researchers and public health entities. This is due to its immediate access to clinical data within a specific region (Rudin et al., 2014). According to the systematic study conducted by Rudin et al. (2014), “the use of HIE when it is available is low and likely dependent on context and implementation factors poorly reported in published evaluations.” Through this translational project, I aim to implement an informatics framework that will allow the RGV HIE to assess and improve upon its data quality in order to logistically plan for data deliveries.

Figure 7 illustrates the proposed approach in this translational paper, which aims to enhance the quality of data for the Health Information Exchange (HIE) by implementing a Quality Assurance (QA) framework. Similar to Figure 3, Figure 7 depicts the relationship between QA and clinical data issues. However, it labels the QA process as "operationalization" to highlight how this project aims to improve health outcomes, increase revenue, and provide research opportunities by emphasizing the connection between QA and data quality. In this context, "operationalize" refers to establishing a methodology that is not reliant on specific measurements or conditions, ensuring its reproducibility in various scenarios, and demonstrating

its role as a proof of concept. To achieve this objective, the DataGauge framework developed by Diaz-Garelli et al. (2019) is utilized; this model is designed to be agnostic to specific data variables, making it a suitable framework for evaluating the consistency of data across the servers within the RGV HIE.

Figure 7

Project Intent



In this project, I leverage the Texas Department of State Health Services' (DSHS) reportable disease standard for hepatitis C (Texas DSHS, 2021). Nonetheless, it is important to note that this analysis specifically focuses on a subset of these standards to narrow down the scope and minimize complexity; the primary focus is on the lab value, lab name, and date and time of testing for hepatitis C. By streamlining the focus to these specific elements, it is possible to effectively identify discrepancies in the diagnostic and lab report dates within the RGV HIE. This targeted approach enhances the accuracy and relevance of the analysis while still aligning with the requirements outlined by the DSHS.

SMART Statement

This project aims to implement vendor-independent data QA through the reduction of discrepancies in hepatitis C diagnoses and positive lab results between the data repositories

connected to the RGV HIE by using the DataGauge Framework (Diaz-Garelli et al., 2019) from October 1, 2022 to December 31, 2022.

PICO Statement

- Problem: RGV HIE lacks a standardized vendor-independent data QA process.
- Population: RGV HIE serves a population that is ~85% Hispanic/Latino (Salcedo et al., 2021) and has a prevalence of hepatitis C of 2.3% (Watt et al., 2015).
- Intervention/indicator: I intend to provide an evidenced-based quality assessment/workflow to iteratively assess and improve data.
- Comparison: I intend to find discrepancies within diagnostic and lab report dates for hepatitis C.
- Outcome: The first outcome is an empirical assessment of the data quality of the critical data elements in the RGV HIE. The second outcome is the reduction of discrepancies in the diagnostic counts between the RGV HIE and data sources. The third and final outcome is a return-on-investment analysis of the work conducted in this translational project.

Section 2: Evidence-Based Practice Review

To inform this translational project, I conducted a literature search for publications that attempt to define data quality and provide data QA frameworks. Under the guidance of the librarian services provided by the Texas Medical Center, I narrowed and specified the focus of my practice question to the healthcare field; literature outside this field was included only through a citation search of the reports included for review. The inclusion criteria encompassed publications that were peer-reviewed, written in English, and had an abstract. Publications that exclusively presented data quality definitions and frameworks specific to a disease, schema,

clinical system, or country were excluded. Instead, data definitions and frameworks that attempted to be agnostic to circumstance and generalizable to other situations were desired. In this context, the term “agnostic” refers to the inherent flexibility and adaptability of the employed framework. While the specific results obtained from applying this framework may be particular to the case under investigation, the underlying standardized steps and heuristics can be generalized and applied to a wide range of settings within the field. This characteristic of the framework enables its transferability and applicability across various healthcare contexts, thereby contributing to its robustness and potential for widespread implementation. A dataset-agnostic framework is necessary given the intent of this translational project, as the implementation of the current evidence-based practices should be applicable to all possible data analysis projects required by the RGV HIE.

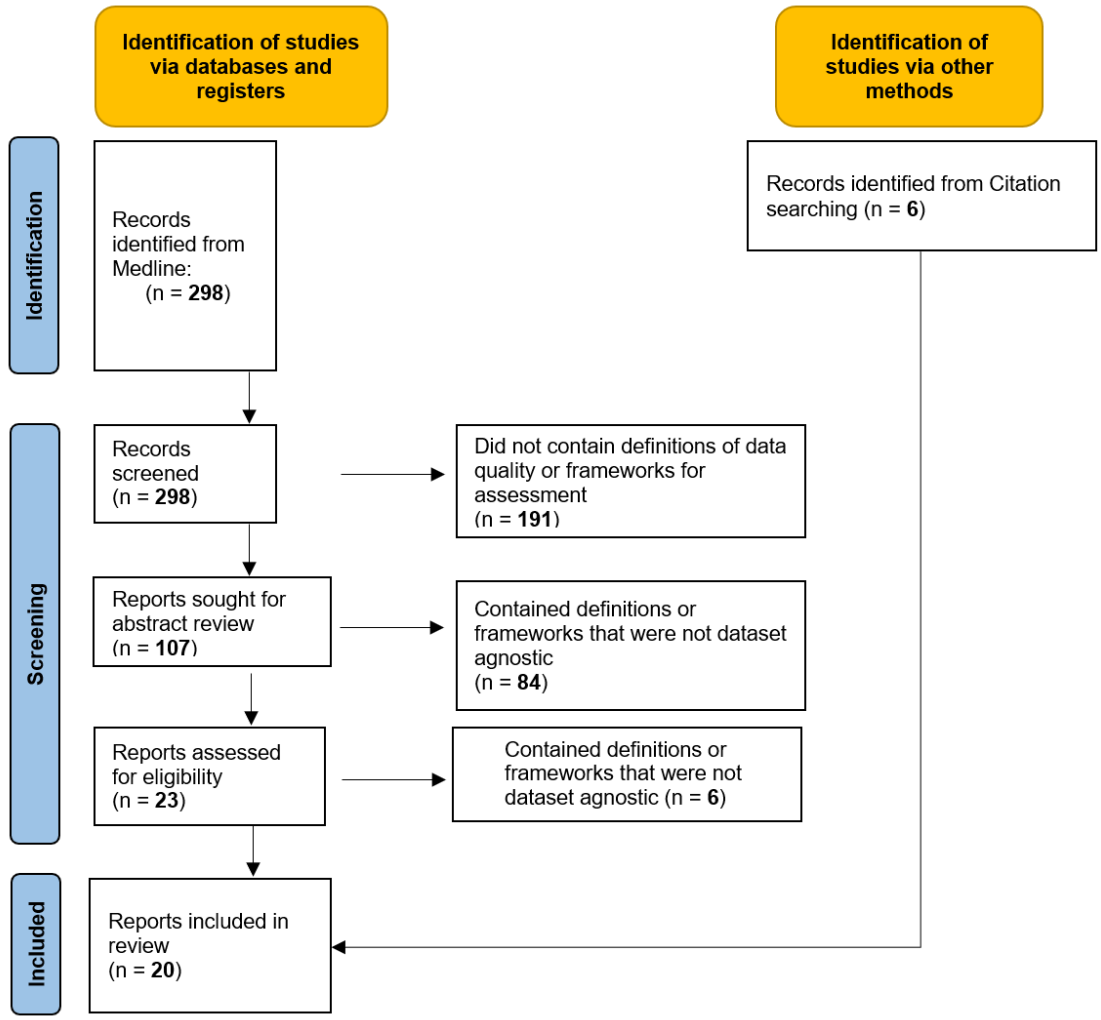
The search was conducted using OVID Medline for publications between 2000 and 2020. The key search terms used were the following: “data,” “quality*,” “framework*,” “accura*,” “science*,” “model*,” “standard*,” domain*,” “defining,” “methodolog*,” “assessing,” “usability,” “utilization,” “manag*,” “govern*,” “collect*,” “need*,” “monitor*,” “extract*,” “format*,” “interoperab*,” “inter-operab*,” “interchange*,” and “inter-chang*.” In addition, terms were used based on the National Institute of Health’s (NIH, 2019) recommendations for search strategies for EHRs. Finally, this search also used the following Medical Subject Headings (MESH) terms: "data accuracy," "data science," "data management," "data mining," "data warehousing," "health information interoperability," and "health information systems."

In order to narrow down the options (illustrated in Figure 8), I utilized a set of exclusion criteria during the review process. An initial search yielded 298 results, which were imported into RefWorks for further examination. These articles were carefully screened to ensure that they

were in English, had an abstract, were peer reviewed by a journal, and contained no duplicates. After conducting a review of the titles, 191 articles were eliminated as they did not provide definitions of data quality or frameworks for data quality assessment. Subsequently, a total of 107 articles remained for abstract review; from this pool, 84 articles were excluded as they did not offer dataset-agnostic data quality definitions or frameworks. The next step involved an in-depth review of 23 selected articles. After careful analysis, six articles were excluded as they did not meet the criteria of containing dataset-agnostic definitions or frameworks. This process resulted in a final selection of 17 articles. Additionally, three more articles were identified through citation searching of the chosen 17 studies; these articles were included in the total, bringing the final number of articles for review and presentation to 20.

Figure 8

Prisma Diagram



Data Quality Problems

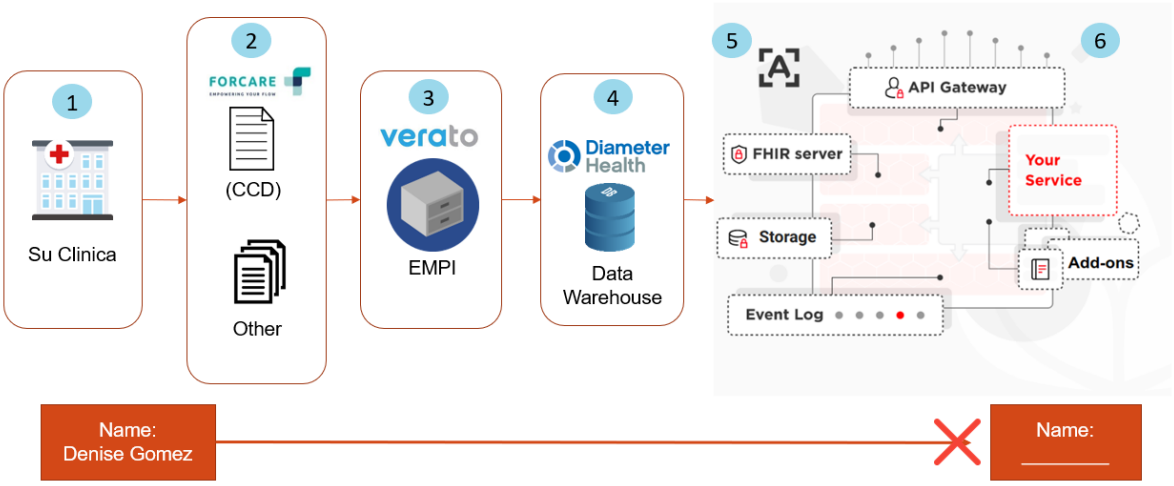
To address the issues related to data quality, it is essential to document both the technical and social factors at play, as these factors interact to create complex issues that require careful consideration and documentation. The following subsections provide a few examples of issues concerning poor data quality.

Example 1: Data Transmission Failures

The first example, illustrated in Figure 9, starts with a hypothetical patient, Denise Gomez, who is in a clinical setting. Files are supposed to be transferred from the clinic to the servers; however, it is later observed that Denise Gomez's data is missing. This situation is clearly not ideal, as the data should have successfully made its way from one end to another. Unfortunately, something went wrong along the way, and some of the potential reasons for this data loss are now outlined. Firstly, there may have been transmission failures, resulting in the data not being sent at all from one server to the next. Secondly, data-matching failures could have occurred, causing the data that should have been associated with a specific individual to become detached, thus leading to its loss. Lastly, data backup failures could have occurred: although the data may have been correctly uploaded and transferred, a natural disaster or catastrophic event might have caused the loss of data. This failure could be due to the backups, which were intended to be a duplicate of the working data but actually contained missing information (Arts et al., 2002; Chen et al., 2014; Leon et al., 2020; Paul et al., 2015).

Figure 9

Data Transmission Failures

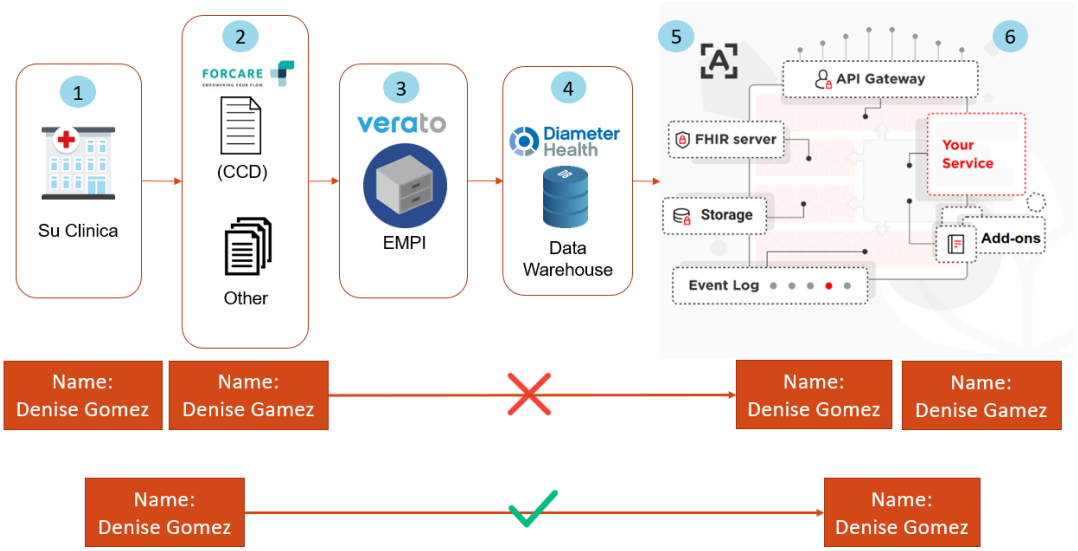


Example 2: Patient Duplicates

Another scenario arises when patients have duplicate records, such as in the clinical setting and servers (Hovenga & Grain, 2013). Figure 10 showcases an example where a single individual's name, Denise Gomez, was spelled in two different ways: Denise Gomez and Denise Gamez. The two records are pushed from the clinical entity to the end of the RGV HIE process when the data is made available through the FHIR servers. The expected outcome was for these names to merge into a single deduplicated identity, specifically Denise Gomez; however, the merging process did not occur as intended, resulting in the persistence of separate identities for Denise Gomez and Denise Gamez. This discrepancy is far from ideal and must be addressed promptly to ensure accurate and streamlined data management (Hovenga & Grain, 2013)..

Figure 10

Patient Duplicates



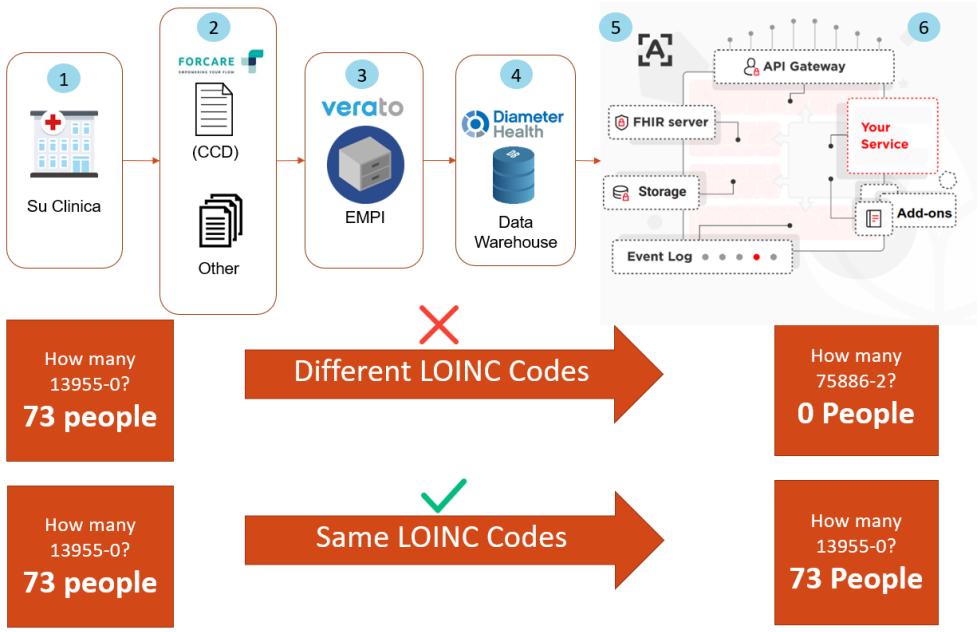
Example 3: Non-Matching Codes

Another potential concern arises from code equivalence, which can create challenges for researchers and analysts. In essence, the codes employed on one server may differ from those

used on another server, resulting in deceptive queries (Feder, 2018; Gruendner et al., 2021; Kahn et al., 2012; Peng et al., 2020). Figure 11 illustrates a hypothetical scenario where the focus was on determining the number of individuals with detectable hepatitis C RNA within a specific timeframe using lab tests. Initially, it seemed like a straightforward inquiry, as only two tests needed to be examined; however, it was discovered that the lab had a multitude of Current Procedural Terminology (CPT) codes, and upon closer inspection, 73 patients were found to be missing in the final step of the process. The root of the problem became evident with the realization that both sides (1 and 6) were searching for different Logical Observation Identifiers Names and Codes (LOINC) codes; these different codes referred to the same test, causing confusion and complications. This issue highlights the importance of ensuring that the codes used for querying align with the codes employed by the clinic. It is crucial to verify the equivalence of codes to avoid any discrepancies and ensure accurate data analysis.

Figure 11

Non-Matching Codes



Data Quality Definitions

The burden of mapping and maintaining a model of data quality requirements is a key issue in software development and computer science (Gruendner et al., 2021; Mezzanzanica et al., 2015; Peng et al., 2020; Taggart et al., 2015). Nonetheless, data quality modeling techniques within healthcare remain siloed from other fields due to differences in both social and technical factors (Diaz-Garelli et al., 2019). In a systematic review, Leon et al. (2020) found that data quality interventions cannot be conclusively proven to be effective due to the variance in approaches. Data quality requirements in healthcare are influenced in large part by non-unified data standards across systems as well as regulations that keep data siloed (Aliabadi et al., 2020).

The literature offered no clear consensus on the exact definition of data quality, although overlapping terms exist, which is consistent with the literature review conducted by Chen et al. (2014). Nonetheless, there is consensus that data quality has a dynamic definition that is determined by the use case at hand for the stakeholders involved. To define data quality, it is first necessary to engage in a sociotechnical iteration process to improve the collection of the critical dimensions necessary for the project to which the data pertains (Arts et al., 2002; Diaz-Garelli et al., 2019; Feder, 2018; Kahn et al., 2012; Weiskopf et al., 2017).

In the review (illustrated in Table 1), five studies attempted to provide granular dimensions of data quality that exhibited the required characteristic of being agnostic to the specific dataset to which they were applied. Data accuracy, which is the alignment with what is understood to be the standard of truth or reality, was mentioned three times (Arts et al., 2002; Diaz-Garelli et al., 2019; Feder, 2018). Data completeness was the one dimension that all five of the studies agreed upon as being necessary in data QA (Arts et al., 2002; Diaz-Garelli et al., 2019; Feder, 2018; Weiskopf et al., 2017). Concordance, a set value of consistency, appeared in

two studies (Diaz-Garelli et al., 2019; Weiskopf et al., 2017), and correctness, the correct reflection of reality and its measurement, also arose twice (Kahn et al., 2012; Weiskopf et al., 2017). The term "believability" only occurred once (Diaz-Garelli et al., 2019) in the studies, despite sharing connotations with both "credibility" (Feder, 2018) and "plausibility" (Weiskopf et al., 2017). No studies agreed upon the nomenclature of frameworks to address data quality (Diaz-Garelli et al., 2019; Feder, 2018; Kahn et al., 2012; Weiskopf et al., 2017).

Table 1

Data Quality Definition Comparison

Diaz-Garelli et al. (2019)	**Feder (2018)**	**Kahn et al. (2012)**	**Weiskopf et al. (2017)**	**Arts et al. (2002)**
accuracy	accuracy			accuracy
believability				
concordance			concordance	
completeness	completeness	completeness	completeness	completeness
timeliness	timeliness			
amount of data				
		integrity		
		flexibility		
		understandability		
		correctness		
		simplicity		
		integration and ability to implement		
			plausibility (could be interpreted as believability and credibility)	
			currency (could be seen as timeliness)	
			granularity	
			fragmentation	

	consistency
	credibility (could be interpreted as believability and plausibility)

The Need for Data Model Development

A solution that addresses data interoperability issues related to transmission failures and non-matching codes is the utilization of a data model (Cowie et al., 2016; Min et al., 2018; Prasser et al., 2018). Such a model serves as a reference point for both systems and users, helping to consolidate semantic inconsistencies and enhance translatability (Min et al., 2018). Consequently, a dynamic and evolving data model becomes an essential component of the quality assurance process. This model must adapt to differing requirements, as various factors such as technologies, staffing, and policies impact how data is stored, monitored, and extracted (Cowie et al., 2016). To be applicable to the use case, data model development must therefore be iterative in its construction of the model for each unique case, ensuring that it accommodates the changes caused by these variables (Prasser et al., 2018).

Data Quality Frameworks

Four of the reviewed papers provided data quality frameworks (Arts et al., 2002; Diaz-Garelli et al., 2019; Kahn et al., 2012; Weiskopf et al., 2017). The various frameworks used to evaluate the definitions of data could be applied in a variety of circumstances, with DataGauge being the most versatile by being agnostic to any dataset and scenario requiring clinical data (Diaz-Garelli et al., 2019). DataGauge achieves this purpose by being a tool for the “secondary use” of clinical data, which refers to the use of clinical data for answering questions outside of direct clinical care, such as quality improvement or practice questions (Diaz-Garelli et al., 2019). DataGauge can be applied to any question by initially defining the scope in its first two steps.

The following two subsections provide a more detailed overview of the process, highlighting that steps 1-4 pertain to model development, while the fifth and final step focuses on data evaluation.

Data Model Development

The need for a data model is explicitly stated in step 1 of the DataGauge process and explicitly asks the implementer to determine the question that needs to be answered from the EHR data (Diaz-Garelli et al., 2019). This approach coincides with Kahn et al. (2012), who emphasize the importance of defining "use cases" before starting a project, which helps establish the scope and purpose of the data quality assurance effort. This process aligns with the concept of a data model, which serves as a definition of what needs to be achieved and helps guide the project's scope. Weiskopf et al. (2017) further emphasize the significance of a structured approach guided by the data model, although only from a specific use case: data points for a patient. Such an approach ensures that queries are consistently executed on the same servers, thereby avoiding discrepancies and generating reliable results.

Subsequently, the second step of the DataGauge process indicates "develop[ing] a data needs model (DNM) that formalizes the data needs" (Diaz-Garelli et al., 2019). In other words, the second step aims to ascertain what data is needed to inform the question established in the previous step. These first two steps work together to ensure that a question is answerable given the data available; for instance, if a researcher were to ask a practice question pertaining to zip code data when zip codes are not available, they would learn during this second step that their question was out of scope concerning the available data.

The third step of the DataGauge process, according to Diaz-Garelli et al.'s (2019) model, involves developing the standards that will be set as requirements, such as (but not exclusive to) acceptable lab codes and ranges from queries (Diaz-Garelli et al., 2019). The fourth step is to

extract the data that matches the standards established in the previous steps. Arts et al. (2002) highlight the necessity of having a clear understanding of where this data is located and how to access it. This knowledge is vital to increasing the reliability and reproducibility of the results; otherwise, repeated queries would present differing results even if the data and its location did not change.

Data Evaluation

The fifth and final step of the DataGauge process is to evaluate the outputs from the previous steps and change any variables if necessary (Diaz-Garelli et al., 2019). All the other models have this iterative aspect (Arts et al., 2002; Kahn et al., 2012; Weiskopf et al., 2017). The dimensions used by all the models and frameworks for evaluation are listed in Table 2 and were comprehensively discussed in the previous section. Any definitions of data quality mentioned are outlined in Table 1. In terms of how definitions interplay within Table 2, Kahn et al. (2012) use the dimensions of completeness, integrity, flexibility, understandability, correctness, simplicity, integration, and implementability; Diaz-Garelli et al. (2019) propose the following dimensions: accuracy, believability, concordance, completeness, and amount of data. However, Diaz-Garelli et al. (2019) uniquely utilize these data definitions after extensive data model development. Only DataGauge iterates over the user requirements in a more granular sense by explicitly evaluating and iterating the previous steps. Whereas DataGauge dedicates four total steps towards data model development, the processes outlined by Kahn et al. (2012), Weiskopf et al. (2017), and Arts et al. (2002) only allocate one step to this development. Arts et al. (2002) and Kahn et al. (2012) suggest working with stakeholders to create diagrams and constraints based on the use case. Weiskopf et al. (2017) are vaguer relative to the other frameworks and do not have a data model development that applies to all use cases but only ask if the data is sufficient for each

patient record. Unlike the other models and frameworks, the DataGauge process outlined by Diaz-Garelli et al. (2019) does not assume that the implementers will know their use case; instead, it outlines a framework to iteratively discover their use case through repeated declarations of the question, standards, and logistics to query the data. DataGauge assumes that through these granular variables, the implementer can begin to change one or many aspects to match a satisfying model for stakeholders through the evaluation step. From this perspective, any data dimension or definition could be applied to the final step of the DataGauge process.

Table 2

Data Quality Assurance Comparison

*Markers: *data model development, **data evaluation*

DataGauge (Diaz-Garelli et al., 2019)	**Definitions for 8 Dimensions of Data Model Quality (Kahn et al., 2012)**	**3x3 Data Quality Assessment (Weiskopf et al., 2017)**	**Quality Assurance Framework (Arts et al., 2002)**
*(1) Define information needs based on the analysis question and analytical methods	* Data modelers work form use cases, which are vignettes illustrating the tasks that an information system needs to support	*1A. There are sufficient data points for each patient.	* Develop the user requirements from the user perspective, which will imply the requirements placed upon the data.
*(2) Develop a data needs model (DNM) that formalizes the data needs	**Completeness—Does the data model contain all user requirements?	**1B. The distribution of value is plausible across patients.	** (1) Data accuracy (the extent to which the registered data is in conformity to the truth) and
*(3) Develop analysis-specific data quality (DQ) requirements based on the analytical purpose, the DNM, and the dimensions of DQ	**Integrity—Does the data model conform to the business rules and processes to guarantee data integrity and enforce policies?	**1C. All data was recorded during the timeframe of interest.	**(2) data completeness (the extent to which all necessary data that could have been registered has actually been registered).
* (4) Extract data from the source dataset to fit the DNM	**Flexibility— Can new data elements and relationships be added if the scope of the project changes?	**2A. There are sufficient data points for each variable.	

**(5) Evaluate the extract according to the DQ requirements and flag all data that infringe upon the DQ assessment standard. Dimensions such as the following may be used to evaluate the data: • accuracy • believability • concordance • completeness • amount of data	**Understandability—Are the concepts and structures in the data model easily understood?	**2B. There is concordance between variables.
	**Correctness—Does the model conform to the rules of the data modeling techniques?	**2C. Variables were recorded in the desired order.
	**Simplicity—Does the data model contain the minimum possibilities and relationships?	**3A. There are sufficient data points for each time.
	**Integration—Is the data model consistent with the rest of the organization’	**3B. The progression of data over time is plausible.
	**Implementability—Can the data model be implemented within the existing time, budget, and technology constraints?	**3C. Data was recorded with the desired regularity over time.

Summary of Evidence-Based Practice Review

I conducted a thorough literature search and discovered various issues that can occur with data in clinical settings. These issues include data deduplication, incomplete data, misspelled names, and inconsistent record matching algorithms (Arts et al., 2002; Chen et al., 2014; Leon et al., 2020; Paul et al., 2015). In order to address these challenges, data quality frameworks recommend developing data models specific to the use case at hand (Arts et al., 2002; Diaz-Garelli et al., 2019; Kahn et al., 2012; Weiskopf et al., 2017). By comparing the extracted data to

the desired use cases, one can assess dimensions such as accuracy, timeliness, and completeness (Gruendner et al., 2021; Mezzanzanica et al., 2015; Peng et al., 2020; Taggart et al., 2015). Among the various data quality frameworks, the DataGauge framework provides the most comprehensive guidance on developing a data model (Diaz-Garelli et al., 2019). It emphasizes the importance of declaring the question, data, standards, and steps needed to pull the data. This framework also highlights the significance of evaluating any assumptions made about the data and ensuring that the chosen standards and data extraction capabilities align with the practice question (Diaz-Garelli et al., 2019). Consequently, I will utilize the DataGauge framework for this purpose.

Section 3: Methodology

Due to security concerns, the Rio Grande Valley Health Information Exchange (RGV HIE) has requested that its comprehensive database schemas and documentation be kept private. As a result, the findings presented in this project focus solely on data diagrams pertaining to queries that examine the consistency of hepatitis C lab results across the servers connected to and within the RGV HIE. The data utilized for this translational project is sourced from various locations, including the partner clinical entity, the RGV HIE clinical document repository, the RGV HIE SQL database, and the RGV HIE FHIR server.

To ensure an evidence-based approach, this project adopts an established framework developed by Diaz-Garelli et al. (2019): DataGauge. This framework enables the documentation of assumptions, the actions taken to test these assumptions, and the results obtained from these actions. Each iteration of queries is presented as rows within a results table, accompanied by a data diagram created using PlantUML, a standard open-source tool for creating diagrams (Version 1.2021.2; GitHub, 2021). Each diagram visually depicts the conceptual model provided

by DataGauge. By adhering to this academic approach, this project aims to audit the discrepancies in hepatitis C data within the RGV HIE and contribute to reducing data discrepancies within its servers.

Methods and Model Framework

DataGauge

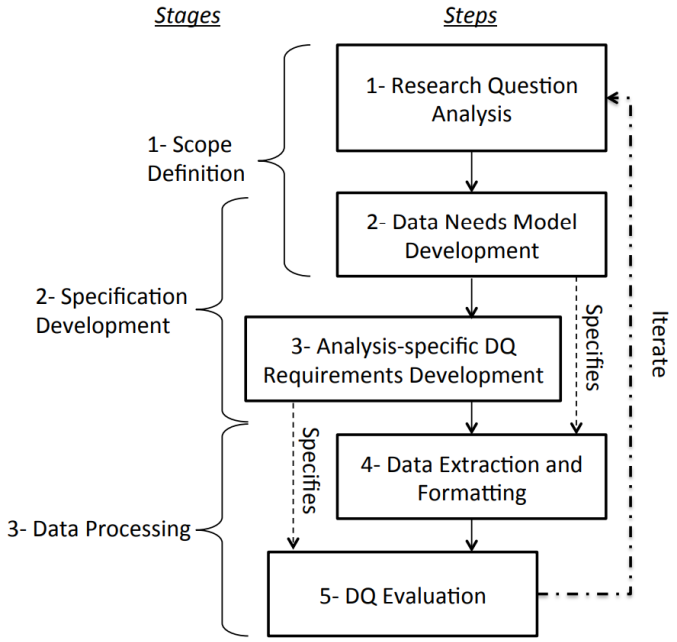
The model guiding this project is the DataGauge framework developed by Diaz-Garelli et al. (2019). According to the Johns Hopkins Nursing Evidence-Based Practice Levels of Evidence, this model is non-experimental and does not include a consensus panel of experts determining the current schema and its abstractions. It is a Level V within the framework, with high-quality evidence because of its definitive conclusions, and provides a logical argument for opinions (Dang et al., 2022). The DataGauge framework was chosen due to its breadth of abstractions, which allow for classifications and mapping of organizational steps to iterate over data quality and reporting; the framework was useful in the initial development phase with stakeholders in this translational project.

DataGauge is described as a tool for the "secondary use" of clinical data, which directly applies to research, as it is not the primary use case of EHR data (Diaz-Garelli et al., 2019) and is illustrated in Figure 12 below. The first step (practice question analysis) requires users of the framework to analyze the pertinent question in terms of its scope and purpose. The second step of DataGauge is data needs model development, which requires users of the framework to make initial presumptions about the data that they might need to answer the pertinent question. The third stage requires analysis-specific DQ requirements development, which are the quality standards for the data to be extracted. The fourth step (data extraction and formatting) asks users to document the expected logistical steps to pull the data and normalize the structure (for

example, conversion from JSON to tabular representation) and data types (text to numerical conversions) for analysis across datasets. Finally, in the fifth step, an evaluation of the previous steps is conducted to assess the suitability of the data and its standards in relation to the practice question and scope. Additionally, it ensures that the necessary abilities, access, and authorizations are in place to effectively retrieve the required data.

Figure 12

DataGauge Steps by Diaz-Garelli et al. (2019)



Based on my interpretation and the goals of this project, I apply my understanding of the DataGauge by Diaz-Garelli et al. (2019) and convert it into a table, which I refer to as the modified DataGauge. This table (illustrated in Table 3 below) allows me to enter, organize, and answer the questions proposed by the DataGauge. A column in Table 3 represents each step, and each row corresponds to an iteration of these steps. The following sections outline these steps:

1. Define the Practice Question. This stage is based on the aforementioned practice question analysis step, which serves as the initial step. Diaz-Garelli et al.'s (2019) DataGauge aims for this step to act as the foundation for measuring all subsequent steps; it involves implementers consulting domain experts to define appropriate inquiries with the necessary assumptions to resolve or address the pertinent practice question. Evaluating project needs and scope before commencing is a standard practice in both project management (Project Management Institute, 2017) and software development (Sommerville, 2016). The rationale behind this practice is that clear objectives provide a definition of completion and facilitate proper documentation of scope changes, which are of utmost importance in stakeholder involvement, resource management, and ultimately risk management (Project Management Institute, 2017; Sommerville, 2016). Consequently, the first step in this process is defining the practice question and thus delineating its scope, given the multitude of factors that it influences.

2. Define the Data Required to Answer Question. The second step of the DataGauge process is data needs model development, which involves creating a data model using unified modeling language (UML) diagrams (Diaz-Garelli et al., 2019). UML diagrams represent the data, its source, and its relationships with other data (Kaliappan & Ali, 2018). By listing the data and its variables, collaboration and iteration are facilitated, allowing for explicit declaration of this data (Redman, 2021); consequently, the UML diagram plays a crucial role in aligning with stakeholder expectations (Bohm et al., 2008). In my proof of concept, I provide UML diagrams in a single diagram and declare the required data in a separate designated table.

3. Define the Data Standards Needed to Validate the Data. The third step of the DataGauge involves establishing data standards for the information gathered in the previous step (Diaz-Garelli et al., 2019). Without these standards, the subsequent steps become unclear,

hindering effective collaboration among stakeholders (Braunstein, 2018). This crucial step defines the acceptable ranges, values, and system responses for future actions (Lin et al., 2018). By having these defined values, collaborators can have a common reference point as they proceed with their work (Bohm et al., 2008; Braunstein, 2018); therefore, without the presence of such standards, effective collaboration becomes unattainable (Wilson et al., 2014).

4. Define the Logistical Steps to Extract the Data. The fourth step in the data analysis process is data extraction and formatting. This stage involves executing queries that align with the data needs model established in step two. The individual who is responsible for performing these queries should have a clear understanding of the data's location and how its sources and keys relate to other tables and servers; their task is to extract the raw data and format it appropriately for the subsequent evaluation phase. It is important to note that errors or misinterpretations during this step can lead to mistakes in the extraction and formatting process; however, it is also possible that any mistakes may have occurred in the previous step, where the data needs model was developed (Zafar et al., 2018).

5. Evaluate the Results and Iterate if Required. The fifth step involves evaluating the data quality using the standards established in the third step. In this phase, the implementer examines the values obtained from the previous step and checks if they conform to an acceptable range or list. Diaz-Garelli et al. (2019) suggest that this evaluation can potentially be automated since it involves measuring the data against a predefined standard; however, for my particular implementation, I assess the steps taken during the extraction process. A case study conducted by Ebad (2020) reveals that errors can arise in both the planning and implementation phases, indicating that mistakes can be made at various stages. This finding aligns with the research by

Zafar et al. (2018), highlighting the universality of issues faced in both the health IT and software development domains.

Table 3

Modified DataGauge Template

	1- Define the question	2- Define data required to answer question	3- Define data standards needed for validating the data	4- Define the logistical steps to extract the data	5- Evaluate the results and iterate if required
Iteration 1 [beginning date – ending date]					
...					

In each iteration of the project, all the practice questions were thoroughly examined and answered, as depicted by the rows in the table. The columns in the table represent the individual steps of the process that challenge and refine the presumptions and assumptions underlying the research inquiry. Table 4 below provides a comprehensive overview of the process involved in gathering patient data from the server. It commences with the initial inquiry in step 1, focusing on determining the number of patients existing on a particular server. Step 2 identifies that the data captured for this query consists of the unique patient IDs associated with each patient. Step 3 establishes the standards for this data, requiring that the patient IDs are not duplicated and that the query does not generate any error messages. In step 4, the query is defined simply as "GET/FHIR/Patient"; however, step 5 highlights an error encountered during the execution of the query, leading to the conclusion that standardized query protocols for the FHIR server are necessary. This finding underscores the importance of developing consistent and reliable methods for retrieving patient data within the RGV HIE.

Table 4

Modified DataGauge Example

	1- Define the question	2- Define the data required to answer the question	3- Define the data standards needed to validate the data	4- Define the logistical steps to extract the data	5- Evaluate the results and iterate if required
Iteration 1 [Beginning Date – Ending Date]	How many patients exist in a server?	Patient IDs	• Patient IDs must not be duplicates • Query must return an error message	Query: GET/FHIR/Patient	Received a 404 error from the query. Need to review query standards in the FHIR server.
Iteration 2					
Iteration 3					

Past DataGauge Implementations

During the preparation of this paper, I discovered two prior works that were not included in my literature review as they were not published in journals. These works, a dissertation by DiazVasquez (2016) published through the UT School of Biomedical Informatics (SBMI) Dissertations (Open Access), and a symposium manuscript by Diaz-Garelli et al. (2021) at the American Medical Informatics Association 2021 Annual Symposium. Although I discovered these manuscripts subsequent to the execution of my project and they had no influence on my work, it is pertinent to acknowledge how my implementation differs from the prior implementations by DiazVasquez (2016) and Diaz-Garelli et al. (2021).

Both DiazVasquez (2016) and Diaz-Garelli et al. (2021) showcase the number of errors they discovered and present a table comparing different data variables to dimensions of data quality, such as completeness, plausibility, correctness, and concordance. However, they do not reveal how the data models changed after each iteration of DataGauge. Also, while both DiazVasquez (2016) and Diaz-Garelli et al. (2021) measure multiple data quality dimensions, my

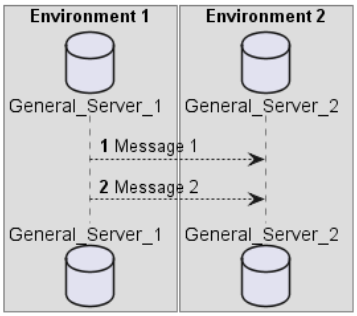
project focuses on measuring a single data dimension, concordance, which refers to the consistency of data between the servers within the RGV HIE and the clinical entity. Additionally, my project aims to iterate over one practice question and demonstrate the evolving data model with each iteration. This distinction sets my project apart from DiazVasquez (2016) and Diaz-Garelli et al. (2021). Furthermore, instead of using MySQL Workbench for visualizing database relationships, I utilize PlantUML, which I will discuss in the next subsection.

PlantUML

Following the DataGauge table, sequence diagrams are used to describe the order of system events, queries, and responses. To indicate these servers, the open-sourced PlantUML (Version 1.2021.2; GitHub, 2021) syntax was used, which allows graphical representations of server relationships to be generated through text input. All the PlantUML code used to generate the sequence diagrams is documented in Appendix D. Figure 13 below provides an example of the graphical outputs: servers are represented by vertical lines, and messages between the servers are represented by horizontal arrows between the servers. A number in the upper-left section of each arrow represents the order in which events occurred (version 1.2021.2; GitHub, 2021).

Figure 13

PlantUML Example 1



The sections can represent different messages (version 1.2021.2; GitHub, 2021). In the below example (Figure 14), “Operation Classification 1” and “Operation Classification 2” are labeled with rectangle framing, with lines crossing the entire picture. They are used to designate messages 1 and 2, respectively. The sectioning of these two messages differentiates them as separate operation classifications, even though they both interact with the same servers and environments.

Figure 14

PlantUML Example 2

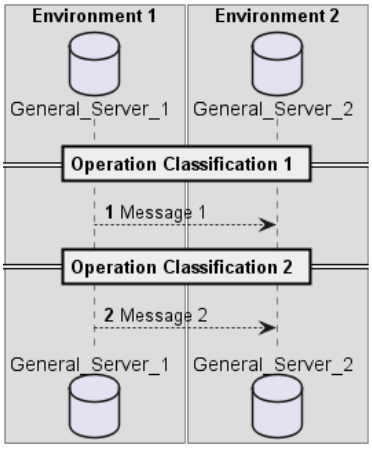


Figure 15 introduces an auditor as an “actor” variable, represented by a human figure. This actor has the ability to interact with the servers by sending their own messages. In this

paper, an X is used to indicate when a message is sent but lacks any accompanying information. In the given example, the auditor attempted to send a message 3 to general server 2, but unfortunately, the message failed to arrive and was incomplete. All of these actions fall under Operation Classification 2.

Figure 15

PlantUML Example 3

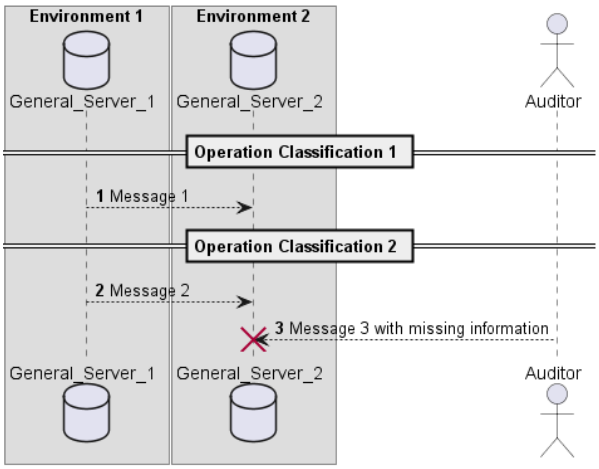
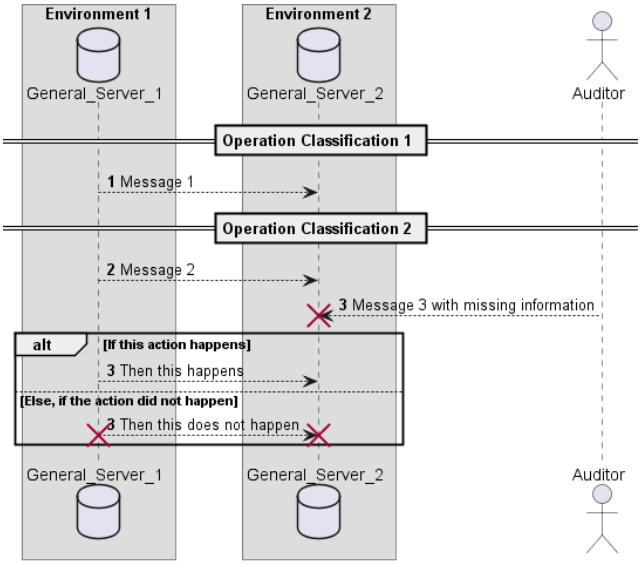


Figure 16 showcases the placement of "if-else" boxes, indicating a message will be sent if a specific event occurs. Conversely, if the event does not take place, no message will be sent (Version 1.2021.2; GitHub, 2021). The utilization of these "if-else" boxes in Figure 16 exemplifies the conditional transmission of messages from General Server 1 in Environment 1 to General Server 2 in Environment 2. The condition within the "if-else" box is contingent upon a hypothetical action: if the action occurs, a message is sent; if the action does not occur, the message is not sent. Messages that are not sent are indicated by a red X on the message line.

Figure 16

PlantUML Example 4



The UML diagrams will be utilized to depict the workflow between servers and the process followed to reach conclusions. In step 2 of their DataGauge framework, Diaz-Garelli et al. (2019) necessitate the communication of the data needs model for describing the data required to fulfill the use case. This is where PlantUML and these diagrams come into play. They are instrumental in conveying the interaction between servers and the presumed actions among different actors in a specific sequence. Additionally, they facilitate effective communication of the work and enable reproducibility, among other things, in the RGV HIE.

Sittig and Singh's Sociotechnical Model

Sittig and Singh (2010) propose a model that encompasses both social and technical variables, offering a formalized description of complex IT systems through these variables. Sittig and Singh's (2010) model serves as a valuable tool for collaborative development and iteration, providing a means to declare and analyze these variables. In this study, I employ Sittig and Singh's (2010) model to elucidate the sociotechnical aspects of my project during its initial and

final iterations. The discussion section also features additional recommendations based on this analysis.

Sittig and Singh (2010) (illustrated in Table 5) describe their sociotechnical model as comprising the following eight key variables: hardware and software pertain to the physical computing components and the instructions stored within those components; clinical content is the information related to the healthcare of the pertinent individuals within the systems; the human-computer interface is a graphical representation of the system's capabilities to allow users to interface through procedural codes or point-and-click interfaces; people represent groups of individuals with various facility affiliations, system access, and authorizations; workflow and communication are the sequence of system and human messages that are sent and received to facilitate processes within the organization; organizational policies and procedures are the rules and standard operating procedures that an organization utilizes to facilitate its processes; external rules, regulations, and pressures, are any legislation and policies that influence the policies and processes in the pertinent systems and organizations; and finally, system measurement and monitoring involve the parsing and dissemination of information pertaining to defined system parameters.

Table 5

Sociotechnical Model Summary

Sociotechnical variable	Definition
Hardware and software computing infrastructure	The physical computers and programs on those computers that are necessary to maintain, enter, and access the pertinent data
Clinical content	The information being maintained, entered, and accessed to answer the pertinent question
Human-computer interface	The methods and technology used by users to access the pertinent data
People	The people performing the tasks
Workflow and communication	The actions taken to accomplish the pertinent tasks and the communication methods used in sequence with those actions
Organizational policies and procedures	The internal/organizational rules that serve to govern and act as constraints of the people and technology within that organization
External rules, regulations, and pressures	The external/state/federal rules that serve to govern and act as constraints of the people and technology within that organization
System measurement and monitoring	The ways in which all the variables in this sociotechnical model are monitored for changes

Summary of Methodology

I will utilize the DataGauge as the evidence-based framework for implementing a quality assurance (QA) process within the RGV HIE. The process will follow the steps outlined by Diaz-Garelli et al. (2019), but I will present a tabular representation of the DataGauge framework as I understand it. The tabular DataGauge representation will consist of iterative rows, each comprising five steps. The first step involves defining the practice question and the specific

objectives of the QA process. In the second step, I will identify the relevant data required to answer the question and present a UML diagram to depict the current understanding of the system. Utilizing the PlantUML standard, I will employ code to generate visual UML diagrams in this step.

Moving on to the third step, I will establish the standards against which the data will be evaluated. Next, in the fourth step, I will define and execute the necessary procedures to extract the data. Lastly, in the fifth step, I will evaluate the data, standards, and extraction steps relative to the use case, which represents the question to be answered for that particular iteration. To describe the interaction of my project with the social and technical aspects of the healthcare system, I will employ the socio-technical model proposed by Sittig and Singh (2010). This model will provide insight into the dynamic relationship between my project and the broader healthcare context.

Section 4: Results

To identify discrepancies or inconsistencies, I used the DataGauge framework and compared the data between the partner clinical entity's EHR (i.e., Su Clinica’s EHR) and the RGV HIE environment. Upon analysis, I found that 30 patients in the clinical entity's EHR had tested positive for hepatitis C; surprisingly, 30 patients were also present in the RGV HIE environment. Nonetheless, there were no recorded hepatitis C results. This discrepancy prompted me to further utilize the DataGauge framework, and the following iterations revealed that the clinical entity had implemented a restriction: they were only transmitting lab results that their physicians had signed off on. Consequently, the lack of physician signatures became the reason for the absence of hepatitis C results in the RGV HIE environment. Until this restriction is addressed, the ongoing issue of discrepancies between the clinical entity's EHR and the RGV

HIE environment will persist, affecting not only hepatitis C but also other lab results. The following subsections will address each iteration and its results separately.

Table 7 in Appendix C provides a concise overview of the DataGauge process I used for my project. The iterations of this process included two successful queries and three unsuccessful ones. The columns represent the different stages of the process. Each row of the table represents a different iteration of these stages and the beginning and end dates of the processes. Sequence diagrams of each iteration are displayed in Figures 32–38 in Appendix E, detailing the presumed order of events that occurred with RGV HIE operations and the associated system queries and responses. The code for the sequence diagrams is shown in Figures 20-31 in Appendix D. The following subsections detail the results as illustrated by Table 7.

Iteration 1

The initial audit aimed to extract encounter data from the FHIR server. Unfortunately, each query returned an error; as a result, the FHIR vendor was consulted to validate subsequent audit queries. The primary focus of this audit was to determine the number of hepatitis C tests conducted at Su Clinica within the RGV HIE between October 1, 2022, and December 31, 2022. To assess the feasibility of answering this question, we initially needed to ascertain if Su Clinica's electronic health record system captures and retains data pertaining to hepatitis C tests within the specified time period. The question presented in its initial iteration and its steps are as follows:

- In stage 1 (define the question), the question I posed was, “Can any encounter data be pulled from the FHIR server?” In this question, I assumed that the FHIR server was present in the HIE and that the encounter data should include lab data.

- In stage 2 (define the data required to answer the question), I searched the FHIR server for placeholder codes (LOINC Code 72376-7 and ICD Code B182) to determine if I could obtain a valid response. I planned to utilize other codes on subsequent servers.
- In stage 3 (define the data standards needed to validate the data), the standards that I placed on the data were that the response from the system must be valid (i.e., not return an error) and that the data must be under the procedure or diagnosis section of the FHIR server. On FHIR servers, sections are called resources (Sharma & Aggarwal, 2018).
- In stage 4 (define the logistical steps to extract the data), I attempted to execute the subsequent queries, building upon the preceding steps taken:
    - REST API Query 1:
        - GET /fhir/Patient?_has:Encounter:procedure-code=http://loinc.org|72376-7*
    - REST API Query 2:
        - GET /fhir/Patient?_has:Encounter:diagnosis code=http://hl7.org/fhir/sid/icd-10|B182*
- In stage 5 (evaluate the results and iterate if required), during the evaluation involving the RGV HIE IT and FHIR vendor, the following error code was generated when attempting to execute queries on the FHIR server:

> Error Code: 1064. You have an error in your SQL syntax'resource Type: OperationOutcome"
> issue:
> - severity: error
> code: invalid diagnostics: No search parameter for Encounter.procedure-code

Upon closer examination, it was discovered that the queries were deemed invalid due to the FHIR vendor's lack of support for FHIR search parameters; instead, the FHIR vendor had

implemented their own proprietary search parameters, which rendered the queries ineffective. Furthermore, it was determined that even if the queries were valid, the user authorizations in place would have prevented their execution. It was therefore advised by the FHIR vendor, for future lab-related tasks, to utilize the observation resource instead of the encounter resource, as the former is a more established and developed resource within the FHIR framework. In terms of the next steps, given my unfamiliarity with the FHIR server, the RGV HIE team suggested selecting the SQL server as the preferred option. Although the FHIR server was presented as a potential alternative, it was recommended to conduct further investigation and training on its capabilities for future projects.

Iteration 2

In the second iteration, the audit's objective was to retrieve data on Su Clinica patients using the SQL server. This iteration was deemed successful, as patient data was retrieved without query errors; however, no hepatitis C tests were found, and this result informed the next iteration.

- In stage 1, the question I posed was, “Using the SQL server, can data on any Su Clinica patients be pulled?”
- In stage 2, on the SQL server, the data required was any patient identifier, with optional data points being any data related to a lab procedure. In order to only pull data from Su Clinica, I used the following facility IDs:
    - Facility IDs: 12, 15, and 18
        - 12 = Su Clinica Familiar Brownsville-Adult
        - 15 = Su Clinica Familiar Harlingen-Adult
        - 18 = Su Clinica Familiar

The first two IDs (12 and 15) correspond to Su Clinica locations in Brownsville and Harlingen, respectively. The remaining ID (18) encompasses other locations.

- In stage 3, the data standards that I placed were that the data must return a valid response with no errors, have a facility ID, and contain a date of visit or creation.
- In stage 4, I attempted to execute the following queries, building upon the preceding steps taken:
    - SQL Query 1:
        - SELECT * FROM data_warehouse.enconter
          WHERE
          (facility_id LIKE '12' or
          facility_id LIKE '15' or
          facility_id LIKE '18') AND
          date between '2022-10-01' and '2022-12-31'
    - SQL Query 2:
        - SELECT * FROM data_warehouse.patient WHERE
          (facility_id LIKE '12' or
          facility_id LIKE '15' or
          facility_id LIKE '18') AND
          date between '2022-10-01' and '2022-12-31'
    - SQL Query 3:
        - SELECT * FROM data_warehouse.problem
          WHERE
          (facility_id LIKE '12' or
          facility_id LIKE '15' or
          facility_id LIKE '18') AND
          date between '2022-10-01' and '2022-12-31'

- In stage 5, the first two queries encountered errors during execution. The RGV HIE IT team discussion revealed that the "patient" and "encounter" tables lacked any data fields, facility fields, or any data point related to a lab or a procedure, thus providing the following error code:

  > Error Code: 1064. You have an error in your SQL syntax'resource Type: OperationOutcome"
  > issue:
  > - severity: error
  > code: invalid diagnostics: No search parameter for Encounter.procedure-code

  Nonetheless, Query 3 was successful, yielding a valid result that confirmed the validity of the facility IDs for the SQL server:

  > 2,580 row(s) found

Upon inspecting the "value_name" and "classification" columns, it was found that no hepatitis C tests were recorded. Further examination of the SQL documentation led to the discovery of another table (the procedure table), which may contain relevant information about lab tests or procedures. It was recommended, after a discussion between the RGV HIE IT team and I, to proceed with an examination of the procedure table to identify any potential tests or procedures that might have been performed.

Iteration 3

The third iteration aimed to fetch data on Su Clinica patients with hepatitis C lab tests performed between October 1, 2022 and December 31, 2022. Although the audit tested the aforementioned procedure table, no hepatitis C tests were found.

- In stage 1, the question I posed was, "Among the patients who are served by the SQL server, is it possible to fetch a patient from Su Clinica from October 1, 2022 to December 31, 2022 with a hepatitis C lab performed?"

- In stage 2, on the SQL server, the data required was any data points related to lab procedures. The date had to be specified in the “date_start” file and be between “2022-10-01” and “2022-12-31”. In order to only pull data from Su Clinica, I used the following facility IDs:
    - Facility IDs: 12, 15, and 18
        - 12 = Su Clinica Familiar Brownsville-Adult
        - 15 = Su Clinica Familiar Harlingen-Adult
        - 18 = Su Clinica Familiar
- In stage 3, the data standards that I placed were that the data must return a valid data response (i.e., no error codes), have a facility ID, contain a date of visit or creation, and contain any data pertaining to hepatitis C.
- In stage 4, I attempted to execute the following query, building upon the preceding steps taken:
    - SQL Query:

```
SELECT * FROM data_warehouse.procedure WHERE
(facility_id LIKE '12' or
facility_id LIKE '15' or
facility_id LIKE '18') AND
date_start between “2022-10-01” and “2022-12-31”
```

- In stage 5, the query was unsuccessful. It yielded a valid result that confirmed the validity of the facility IDs for the SQL server:

    9,914 row(s) returned

Nonetheless, even though a valid response without errors was given, upon further inspection, it was found that the returned results included no hepatitis C tests when conducting

text value inspection on the “value_name” and “classification” columns. The RGV HIE IT team and I discussed the next steps, and it was agreed that an inquiry would be made to Su Clinica to obtain a comprehensive list of relevant tests or procedures. Although access to the database was currently unavailable, the submission of an IT service request ticket was being pursued as a possible means of acquiring said list. Once acquired, the relevant IDs would be cross-referenced with the server to retrieve the necessary data. It was postulated that cross-referencing IDs would prove more efficient than searching through procedure and encounter records.

Iteration 4

The fourth audit iteration focused on identifying Su Clinica patients with hepatitis C within the EHR. The iteration was successful, and 30 patients with Su Clinica IDs and hepatitis C lab tests were identified between October 1, 2022 and December 31, 2022.

- In stage 1, the question I posed was, “Is it possible to obtain data on any patients in Su Clinica between October 1, 2022 and December 31, 2022 who have a hepatitis C diagnosis within Su Clinica’s EHR?”
- In stage 2, in the Su Clinica EHR, the data required was a patient identifier, hepatitis C test name, test date, and whether or not the patient had consented to be part of the HIE.
- In stage 3, the data standards that I placed were that the patient data must be from people who were Su Clinica patients, it must be from patients who had taken a hepatitis C lab test, the test must have been taken between the specified dates, and the patients must have consented to have their data sent to the HIE.
- In stage 4, I executed the queries by creating a ticket through the Su Clinica IT team.
- In stage 5, the query made through the IT ticket was successful. There were 85 patients in Su Clinica who had a hepatitis C test in total, but only 30 consented to be part of the

RGV HIE. Upon consultation with the IT department at Su Clinica, it was determined that laboratory results are not typically stored within the clinic's database; instead, the attending physician signs them before sending them to the relevant parties. Additionally, a set of codes used by Su Clinica was obtained for future reference during data retrieval and analysis:

- CPT:
  - 86803
- LOINC:
  - 48159-8
  - 48159-8
  - 13955-0
  - 11011-4
  - 38180-6
  - 19147-8
  - 62365-2

In terms of the next steps, a workflow diagram was developed to assist in the identification and resolution of issues pertaining to missing laboratory results. Approval from Su Clinica was obtained regarding the problem description and proposed approach. In addition to manual inspection of text values, targeted searches of specific codes will be conducted within the relevant servers to confirm the hypothesis that laboratory reports may not be successfully transmitted.

Iteration 5

The final audit iteration aimed to determine the number of hepatitis C tests that the 30 Su Clinica-identified patients had received within the RGV HIE environment. Unfortunately, the audit was not successful, as no hepatitis C tests were identified in any of the patient records.

- In stage 1, the question I posed was, “How many hepatitis C tests do the 30 identified Su Clinica patients have within the RGV HIE environment?”
- In stage 2, the necessary data from the following servers—Su Clinica, FHIR, SQL, Forcare, and Diameter—includes patient identifiers, the name of the hepatitis C test, and the corresponding test date. For Verato, the EMPI vendor, only patient IDs were required as data.
- In stage 3, the requirement that I set is that the data must include a hepatitis C test in the query results.
- In stage 4, the process involved obtaining a distinct identifier from Su Clinica, which was then used to acquire the EMPI from the FHIR server. Subsequently, the EMPI was utilized to retrieve the required data from the SQL, Diameter, and Forcare databases.
    - Query 1:
        - Forcare user interface:
            - Enter Su Clinica ID in ID section
    - Query 2:
        - Verato user interface:
            - Enter Su Clinica ID in link ID section
    - Query 3:
        - Diameter user interface:
            - Enter Su Clinica ID in MRN extension section
    - Query 4:
        - REST API Query to FHIR ID:
            - Get /fhir/Patient?identifier=[SuClinicaID]

- Query 5:
  - REST API query to obtain labs:
    - Get /fhir/Patient?subject=[FHIR Server ID]
- Query 6:
  - SQL server queries for patient problems:
    - SELECT *
      FROM data_warehouse.patient
      JOIN data_warehouse.problem ON data_warehouse.patient.id = data_warehouse.problem.patient_id
      WHERE (empi LIKE 'EMPI')
- Query 7:
  - SQL server queries for lab results:
    - SELECT *
      FROM data_warehouse.patient
      JOIN data_warehouse.result ON data_warehouse.patient.id = data_warehouse.result.patient_id
      WHERE (empi LIKE 'EMPI')

- In stage 5, queries 1, 2, 5, 6, and 7 yielded no hepatitis C lab results. Queries 3 and 4 yielded 30 patients with 30 IDs.

During the auditor and RGV HIE discussion, analysis of the observed data led to the conclusion that laboratory reports are only transmitted to the RGV HIE when patients have provided explicit consent and attending physicians have provided approval. Although the clinic's database contains records of all patients who have consented to participate in the RGV HIE, laboratory reports are not consistently found within said records. Further investigation is necessary to determine the root cause of this inconsistency.

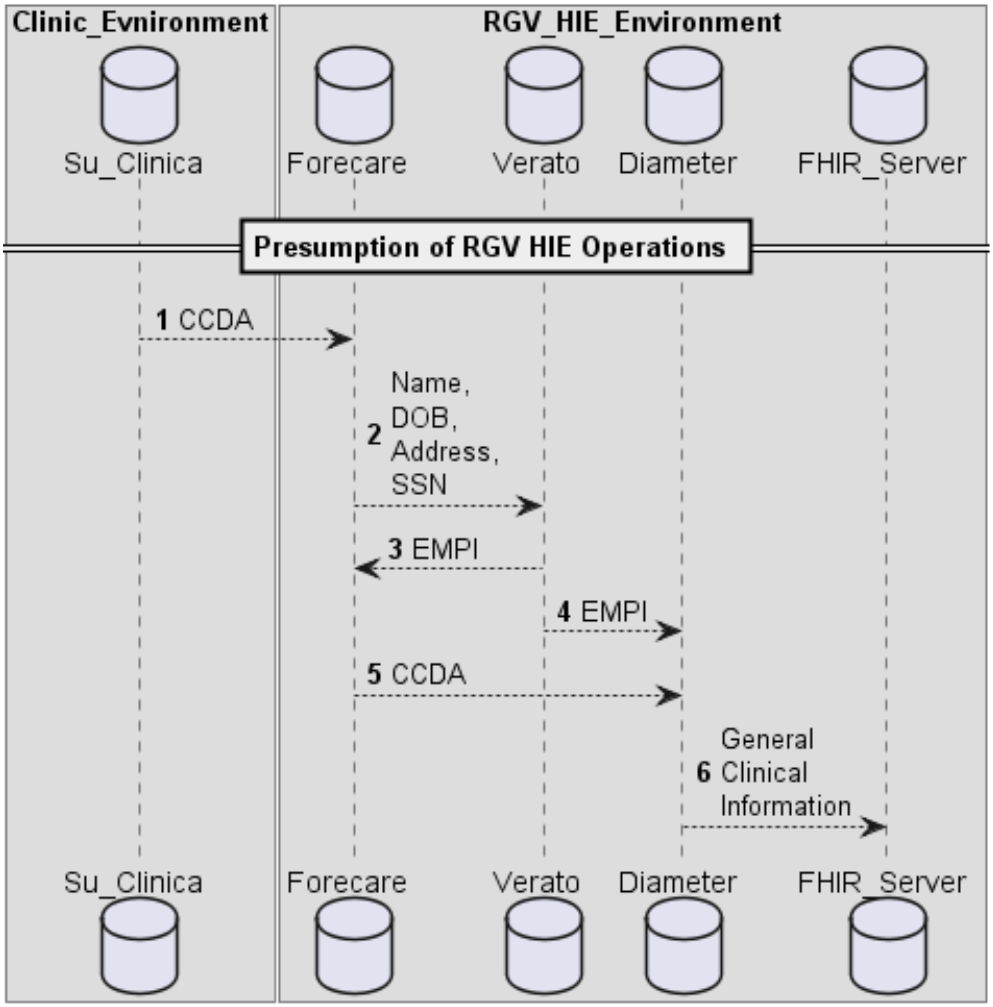
In terms of next steps, the current analysis indicates that the transmission of laboratory reports to the RGV HIE is contingent upon both patient consent and physician approval. Despite the availability of records for consenting patients, laboratory reports are not consistently present within the database, indicating a potential breakdown in the current architecture. A re-evaluation of the architecture is recommended to address this issue and reconcile the need for physician sign-off with the requirement for data accessibility by the RGV HIE. The identification of a scalable and secure solution to this conundrum is crucial for the continued effectiveness of the RGV HIE and its partners.

Initial Presumption Diagram

The initial assumptions made before implementing the modified DataGauge framework are depicted in Figure 17 and explained in this paragraph. Prior to the first iteration, I made the initial assumption that the relevant steps would commence with the consolidated clinical document architecture (CCDA) being pushed to the document repository called Forcare. Subsequently, demographic information (including the patient's name, date of birth, address, and social security number) would be forwarded from Forcare to Verato. The latter, in turn, would utilize this information to generate an enterprise master patient index (EMPI) by comparing it with existing records in the Verato database. This process ensures accurate patient identification across the entire healthcare system. The resulting EMPI is then transmitted to both Forcare and the subsequent vendor, Diameter. Forcare transfers the CCDA to Diameter, which then parses the information and presents it in a human-readable format. Finally, Diameter sends this parsed information to the FHIR server.

Figure 17

Initial Presumption Diagram



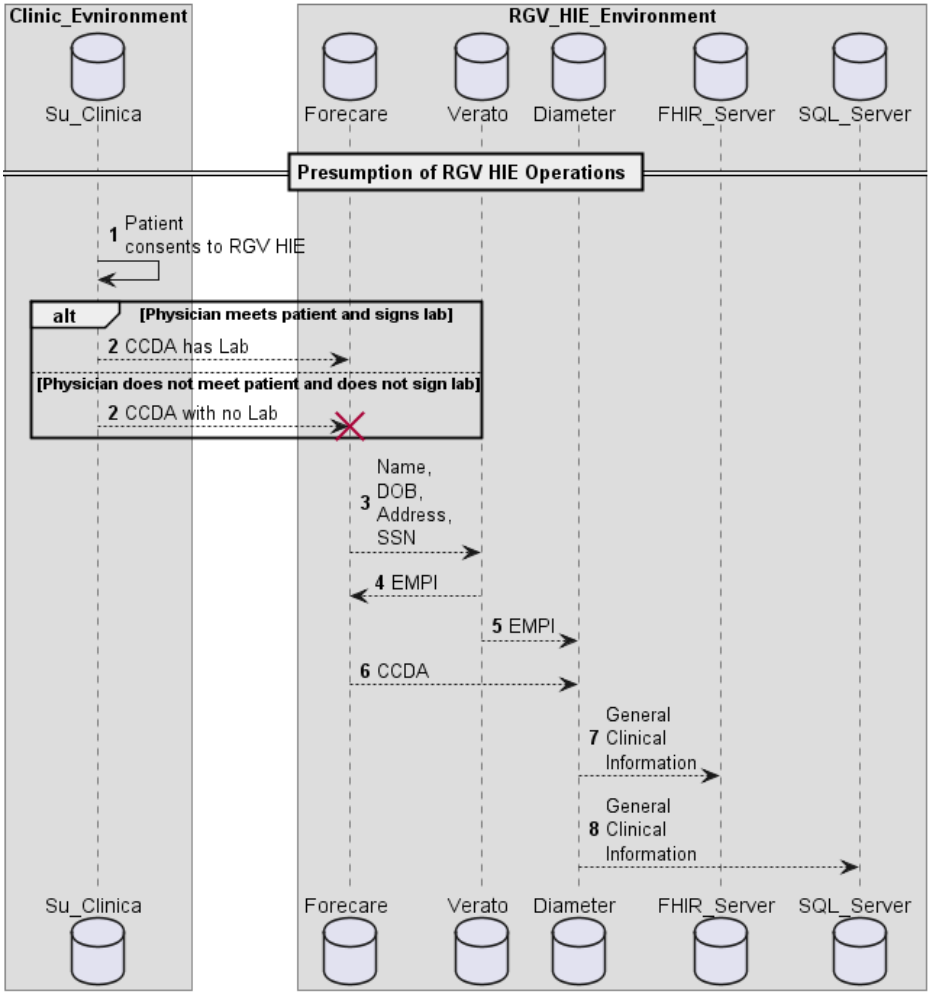
Final Presumption Diagram

The current operations between the RGV HIE and Su Clinica are depicted in Figure 18 and explained in this paragraph. Figure 18 builds upon the initial assumptions in Figure 17 (described in the previous section) but differs in the following key aspects: Firstly, the process commences with the patient providing consent to share their clinical information with the RGV HIE environment. Once consent is obtained, Su Clinica transmits CCDA documents to the RGV HIE environment. Additionally, Figure 18 also illustrates the reason behind the absence of

hepatitis C lab results: the CCDA only includes a lab report if the treating physician authorizes it; if the physician does not approve the lab report, it is not included in the CCDA.

Figure 18

Final Presumption Diagram



Audit Summary and Learnings

The iterations of the modified DataGauge process yielded two successful queries and three unsuccessful ones. Unfortunately, the final query brought the entire process to a halt. The audit found no hepatitis C data received from Su Clinica to the HIE of patients seen at Su Clinica between October 1, 2022 and December 31, 2022, likely because these reports were never sent

and were not signed off by the physician. The audit concluded that this data could not be accessed unless this workflow was addressed. The project was impeded by a change in the practice question, as accessing the necessary data to answer the new question was beyond the scope of this endeavor. The initial motivation for this audit was to determine the number of hepatitis C tests conducted at Su Clinica within the RGV HIE during the specified time frame. However, since there were no hepatitis C tests available for querying in the RGV HIE during those dates, the next steps involved finding solutions within Su Clinica's clinical policy and procedures. These solutions aimed to either eliminate the need for physician sign-off for labs or maintain the requirement while enabling the ability to send the lab test data in a deidentified format. The possibility of sending deidentified lab test data will be discussed in the Discussion section using a process form developed by Guerrero et al. (2019). Ultimately, my Committee and the RGV HIE IT team deemed these next steps to be out of scope, resulting in the project being halted.

Section 5: Discussion

Next Steps

Since the implementation of the DataGauge process for this translational paper, Su Clinica has changed EHRs from GE Centricity to Athena Health. As such, a reevaluation will be necessary to determine if the results are still applicable. Given the new interface between Su Clinica and the HIE, the first question to ask is, “Is data being sent from Su Clinica to the HIE through Athena Health?” Subsequently, if data is being sent, the next question to ask is, “Are labs that are not signed off by physicians still being sent?”

The 21st Century Cures Act was created to address issues in healthcare interoperability and prevent unmercenary data siloes (Black et al., 2018). Information blocking is the blocking of

necessary actions and steps for sharing information and prevents pertinent clinical data for patient health from reaching a setting in which it could be useful (Black et al., 2018). It can occur in various instances, such as clinicians being reluctant to accept or send information (Vest & Gamm, 2010) or EHR vendors obfuscating technical requirements or inflating prices involved in development (Castillo et al., 2018). If Su Clinica and the HIE face such issues, the 21st Century Cares Act may help exert policy pressure for change toward greater interoperability.

Alternative Workflow

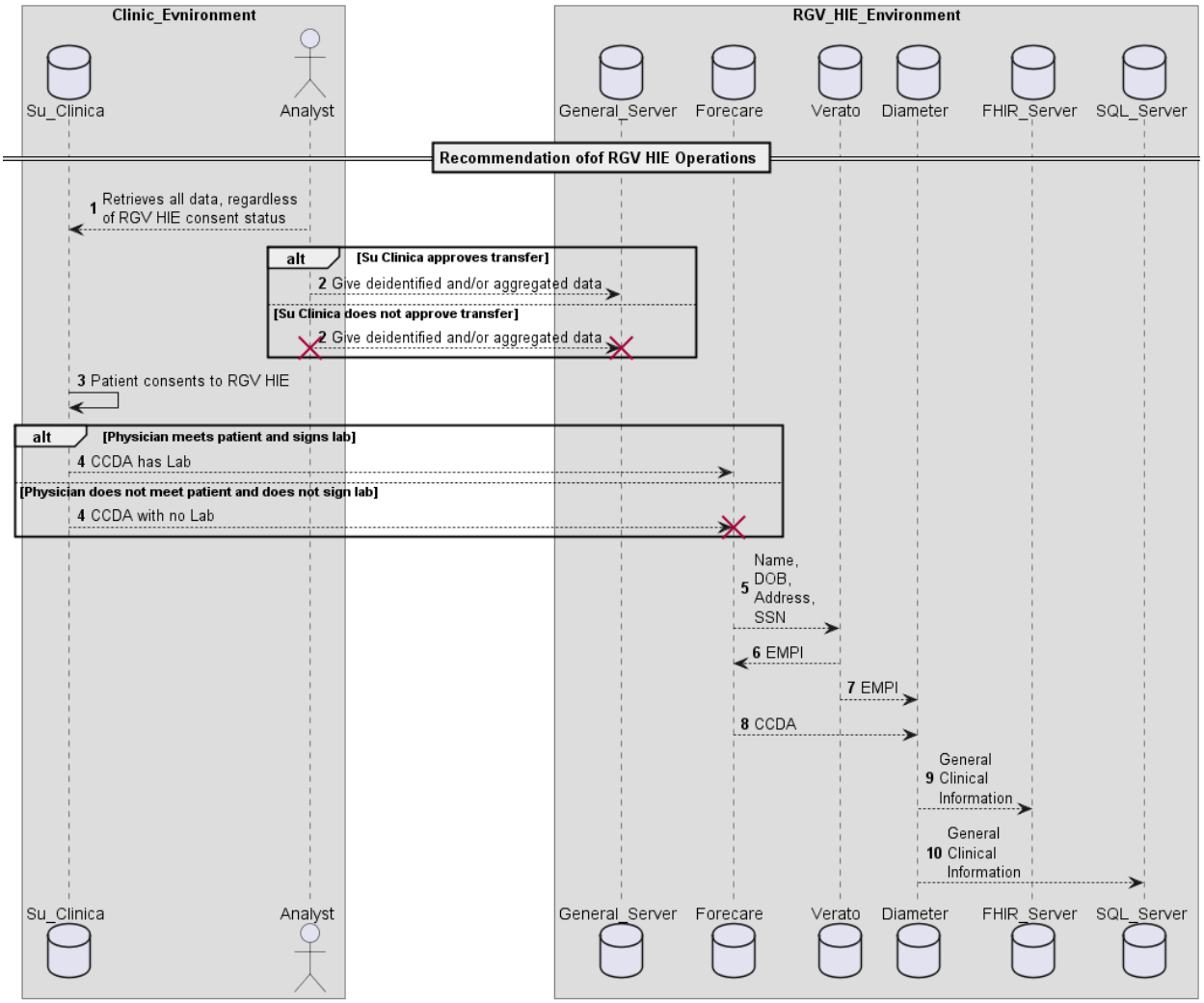
The current implementation of the RGV HIE would not allow the extraction of any data from patients who did not consent to be part of the RGV HIE environment. In addition, even if a patient consents, lab data is only sent if the physician signs off on the lab, as only those reports are sent to the CCDA. If this policy persists, it is possible (with the help of an additional policy) that data could be transferred in the case of public health emergencies (or practice questions) that require these patient populations' clinical information and queued labs for analysis (Guerrero et al., 2019). Consequently, this translational paper recommends the workflow presented by Guerrero et al. (2019), which allows the extraction and analysis of clinical information from patients who are not currently present in the RGV HIE environment or who have not yet consented to the research process. This is possible because, in this process, only deidentified clinical information is presented to those outside the pertinent clinic. Analyzing any identifiable clinical information is only performed within the clinic environment because deidentified protected health information (PHI) no longer carries the same protections as identifiable information under the HIPAA Privacy Rule (Ness, 2007). As such, Guerrero et al. (2019) propose the following constraints to allow for research on deidentified PHI:

1. The data remains on the servers that are maintained by the clinical entity (in this case, Su Clinica)
2. No identifiable data is shared with individual researchers or published, and
3. No contact is made with patients (e.g., to collect additional data; p. 232).

The recommended workflow illustrated in Figure 19, which integrates the constraints proposed by Guerrero et al. (2019), enables RGV HIE to gather the necessary data from the entire patient population or specific subgroups. The process begins with an analyst who works under the auspices of Su Clinica while operating within the clinic environment. This analyst obtains the pertinent data and provides deidentified and/or aggregated count data, subsequently forwarding this data to Su Clinica. The data will only be sent if Su Clinica approves it; otherwise, it will not be transmitted. The remaining steps in the process are the same as those depicted in the preceding diagram.

Figure 19

Recommended Workflow



Sociotechnical Variables: Known and Unknown

In order to monitor the use and evaluation of the implementation, it is necessary to consider all the dimensions in which this project will impact the organization. IT implementations impact both social and technical factors in an organization (Arts et al., 2002; Chen et al., 2014). Hence, the modified DataGauge framework performed in this audit should be expressed in sociotechnical terms if any implementation were to occur. This sociotechnical approach would emphasize the technology, policy, and people elements of the audit and how the

recommendations will affect said variables. The sociotechnical model used is the eight-dimensional model that Sittig and Singh (2010) have described in terms of the following variables:

1. Hardware and software
2. Clinical content
3. Human-computer interface
4. People
5. Workflow and communication
6. Organizational policies and procedures
7. External rules, regulations, and pressures
8. System measurement and monitoring

In the following subsections (illustrated in Table 6), I will address each of these variables and explain their relevance to this project. It is important to note that variables 6 (organizational policies and procedures) and 8 (system measurement and monitoring) are classified as unknown variables under a separate subheading. They fall under this category because they contain ambiguous elements, and due to the project's halt, further investigation within the scope of this project was not possible.

Hardware and software

In the example of the modified DataGauge audit, the hardware and software are the following servers: Su Clinica, Forcare, Verato, Diameter, and FHIR. In the second iteration, the SQL server was added and maintained until the final iteration. In the suggested system sequence, a server labeled “General Server” was added to act as an intermediate point between Su Clinica’s deidentified clinical data and the RGV HIE environment.

Clinical content

The Su Clinica system contains various clinical information, including demographic details, visit dates, care team identities, biometrics, and lab results. This data is transformed into a CCDA document using a modified version of the DataGauge audit tool. The Forcare and Verato servers utilize this CCDA document to provide clinical content; the Verato server focuses on the EMPI, while the Diameter and FHIR servers derive their clinical content from the CCDA. As a final addition, an SQL server was introduced to store and manage the general clinical information extracted from the CCDA document. A "General Server" was incorporated into the suggested system sequence alongside these servers; this server contains Su Clinica's deidentified clinical data, which is awaiting approval before being integrated into the RGV HIE environment.

Human-computer interface

The Su Clinica server was accessed through an EHR user interface (UI), while Forcare and Verato were accessed through their web UIs. The FHIR server was similarly accessed through the web UI, but it also has an application programming interface (API). From the second to the final iteration, the SQL server was accessed through an SQL graphical tool and console called MySQL Workbench. In the suggested system sequence, the “General Server” was added; the human-computer interface would be the UI of the operating system used by the server.

People

The people involved in this paper’s audit are the following: Su Clinica’s IT team, who have access to the Su Clinica server; the RGV HIE IT team and auditor, who have access to all the servers in the RGV HIE environment; and the FHIR vendor staff, who have access to the FHIR server. In the second through final iterations, the SQL server was added, but only the RGV HIE IT team and auditor had access. In the suggested system sequence, the “General Server” was

added. The person responsible for the General Server would be the analyst who works on behalf of Su Clinica to send deidentified clinical information to the RGV HIE. Su Clinica will also have access to the General Server, as they must be able to inspect the data and information before it is sent.

Workflow and communication

In the modified DataGauge process of this paper, the workflow and communication for the first iteration begin with Su Clinica sending a CCDA to Forcare. Forcare will then send identifying information from the CCDA (i.e., name, date of birth, address, and social security number) to Verato, who will generate an EMPI and send it to both Forcare and Diameter. Forcare will then send the CCDA to Diameter, which is linked via the EMPI. Diameter will parse the information and then send the parsed clinical information (denoted as "general clinical information") derived from the CCDA to the FHIR server. In the second through final iterations, the SQL server was added, and it also received general clinical information from Diameter. In the suggested system sequence, the "General Server" was added, which would serve to hold deidentified healthcare data gathered by the analyst and would only be pushed to the RGV HIE environment once Su Clinica agreed to do so.

External rules, regulations, and pressures

In the modified DataGauge process described in this paper, there are three primary pieces of legislation. The first is the HITECH Act (Gold & McLaughlin, 2016), which aimed to support interoperability, address data silos, and thus incentivize and support the creation of HIEs. The next policy and legislation is the HIPAA Privacy Rule (2000), which concerns the use of PHI and how it could be shared among organizations through business associate agreements (BAAs). Finally, in the state of Texas, Su Clinica in particular follows rules regarding HIPAA and the

Texas Administrative Code (1997), which dictate how licensed physicians manage their medical records. These factors remained consistent through each iteration and in the recommended system sequence.

Unknown Variables

The variables 6 (organizational policies and procedures) and 8 (system measurement and monitoring) remain unknown due to their lack of clear definition and understanding. Unfortunately, the limitations of the project's scope prevented further investigation into these variables. While the other variables did not pose any ambiguity at this point in the project, these two require attention. In the following subsections, I will address each variable and identify both the unambiguous and ambiguous elements. Additionally, I will outline how my recommendations align with these variables based on my current understanding.

Organizational policies and procedures

Regarding the initial iteration of the audit, RGV HIE utilizes BAAs with its partners to establish the use of PHI in accordance with the HIPAA Privacy Rule (2000). Nonetheless, by the final iteration, the audit noted that Su Clinica will only send PHI pertaining to patients who have consented to be part of the RGV HIE environment. Lab data that the physician has not signed off will not be pushed to the RGV HIE environment from the clinic environment; however, it is not known whether the physician sending the lab report is part of an explicit procedure of Su Clinica. The procedural rule is within the EHR and likely exists due to the HIPAA Privacy Rule (2000), which describes the need to maintain the clinical accuracy of PHI on the covered entity (in this case, Su Clinica). There are other regulations, such as those in the Texas Administrative Code (1997), which states that all clinical data held by a licensed physician of the board “must contain accurate data and information pertaining to the patient based on actual findings, assessments,

evaluations, diagnostics, or assessments as documented by the physician." Together, both pieces of legislation hold Su Clinica and its physicians liable if any clinical information is misrepresented when sent outside its organization; therefore, it stands to reason that no lab values may be pushed to the RGV HIE without the physician's approval. The information gathered in this audit only confirmed whether Su Clinica IT and RGV HIE IT claim that this is a feature within the Su Clinica EHR; however, it is not known whether this feature can be toggled to submit labs without the physician signing off, nor whether it is an explicit procedure of Su Clinica. In the recommended system sequence, the paper recommends an analyst role that would act as an intermediary between Su Clinica and RGV HIE and, with Su Clinica's explicit approval, deliver deidentified clinical information siloed within Su Clinica. This would allow researchers to gather population information for clinical and research purposes without transferring PHI.

System measurement and monitoring

Currently, the feed of CCDA documents from Su Clinica will be observed in the RGV HIE environment and the servers within it: Forcare, Verato, Diameter, FHIR, and the SQL server. Nevertheless, there is no automated or systematized monitoring of records to ensure that consistency is maintained. This limitation was the reason for embarking on this audit, which aimed to measure concurrency between systems and create a model that could support further system measurement and monitoring. Considering the results of this audit, future work could be conducted to establish automatic concurrency measurements on each server; in other words, an automation of a subset of the methods of this audit. An automation of concurrency measurements would entail a script to run queries on each server and document the results to ensure that data values such as lab values and dates remain consistent within each system.

The syntax used to express the workflow in Table 6 is derived from the PlantUML syntax, which was discussed in the methods section (GitHub, 2021). For conciseness, the workflow is only represented by server names linked to each other using a "→" sequence of characters, followed by a colon and the name of the information sent. Numbers are also added beside each line to indicate where these interactions occurred in the sequential order of messages. Finally, the syntax adopts the convention established in this paper for this table, which indicates that any added variables relative to the previous iteration are denoted with a "+" symbol and are bolded.

Table 6

Sociotechnical Summary

Modified DataGauge Expressed in a Sociotechnical Framework (new variables in iterations denoted with a + symbol and **bolded**)

Sociotechnical variable	First Iteration	Final Iteration	Suggested System Sequence
Hardware and software computing infrastructure	Servers: Su Clinica, Forcare, Verato, Diameter, FHIR	Servers: Su Clinica, Forcare, Verato, Diameter, FHIR, **+SQL**	Servers: Su Clinica, Forcare, Verato, Diameter, FHIR, SQL, **+General Server**
Clinical content	Servers: Su Clinica—CCDA Forcare—CCDA and name, date of birth, address, and social security number Verato—EMPI Diameter—General clinical information (derived from CCDA) FHIR—General clinical information	Servers: Su Clinica—CCDA Forcare—CCDA Verato—EMPI Diameter—General clinical information (derived from CCDA) FHIR—General clinical information (derived from CCDA) **+SQL**—General clinical information (derived from CCDA)	Servers: Su Clinica—CCDA Forcare—CCDA Verato—EMPI Diameter—General clinical information (derived from CCDA) FHIR—General clinical information (derived from CCDA) SQL—General clinical information (derived from CCDA)

	(derived from CCDA)		**+General Server**—Deidentified clinical data in queue to RGV HIE, pending Su Clinica's approval
Human-computer interface	Servers: Su Clinica—EHR UI Forcare—Web UI Verato—Web UI Diameter—Web UI FHIR—Web UI and API	Servers: Su Clinica—EHR UI Forcare—Web UI Verato—Web UI Diameter—Web UI FHIR—Web UI and API **+SQL**—Graphical Console (MySQL Workbench)	Servers: Su Clinica—EHR UI Forcare—Web UI Verato—Web UI Diameter—Web UI FHIR—Web UI and API SQL—Graphical Console (MySQL Workbench) **+General Server**—UI of the operating system used by the server
People	Servers: Su Clinica—Su Clinica IT Forcare—RGV HIE IT and auditor Verato—RGV HIE IT and auditor Diameter—RGV HIE IT and auditor FHIR—RGV HIE IT, auditor, and FHIR vendor	Servers: Su Clinica—Su Clinica IT Forcare—RGV HIE IT and auditor Verato—RGV HIE IT and auditor Diameter—RGV HIE IT and auditor FHIR—RGV HIE IT, auditor, and FHIR vendor **+SQL**—RGV HIE IT and auditor	Servers: Su Clinica—Su Clinica IT and **+Analyst** Forcare—RGV HIE IT and auditor Verato—RGV HIE IT and auditor Diameter—RGV HIE IT and auditor FHIR—RGV HIE IT, auditor, and FHIR vendor SQL—RGV HIE IT and auditor **+General Server**—Analyst and Su Clinica
Workflow and communication	1. Su_Clinica --> Forcare: CCDA 2. Forcare --> Verato: name, date of birth, address, social security number	1. **+Su_Clinica --> Su_Clinica**: Patient consents to RGV HIE **+alt Physician meets patient and signs lab** 2. **+Su_Clinica --> Forcare**: CCDA has lab	1. **+Analyst --> Su_Clinica**: Retrieves pertinent data (depending on clinical question) regardless of RGV HIE consent status **+alt Su Clinica approves transfer**

3. Verato --> Forcare: EMPI
4. Verato --> Diameter: EMPI
5. Forcare --> Diameter: CCDA
6. Diameter --> FHIR_Server: General clinical information

+else Physician does not meet patient and does not sign lab

2. **+Su_Clinica --> Forcare**: CCDA with no lab
3. Su_Clinica --> Forcare: CCDA
4. Forcare --> Verato: name, date of birth, address, social security number
5. Verato --> Forcare: EMPI
6. Verato --> Diameter: EMPI
7. Forcare --> Diameter: CCDA
8. Diameter --> FHIR_Server: General clinical information
9. **+Diameter --> SQL_Server**: General clinical information

2. **+Analyst --> General_Server**: Give deidentified and/or aggregated data

+else Su Clinica does not approve transfer

2. **+Analyst --> General_Server**: Give deidentified and/or aggregated data
3. Su_Clinica -> Su_Clinica: Patient consents to RGV HIE

+alt Physician meets patient and signs lab

4. Su_Clinica --> Forcare: CCDA has lab

+else Physician does not meet patient and does not sign lab

5. Su_Clinica --> Forcare: CCDA with no lab
6. Su_Clinica --> Forcare: CCDA
7. Forcare --> Verato: name, date of birth, address, social security number
8. Verato --> Forcare: EMPI
9. Verato --> Diameter: EMPI
10. Forcare --> Diameter: CCDA
11. Diameter --> FHIR_Server: General clinical information

			12. Diameter --> SQL_Server: General clinical information
Organizational policies and procedures	BAA relationship per HIPAA privacy law	BAA relationship per HIPAA privacy law **+ No lab may be pushed outside Su Clinica that is not signed by physician**	BAA relationship per HIPAA privacy law No lab may be pushed outside Su Clinica that is not signed by physician **+ If relevant to a clinical or practice question, Su Clinica may have an analyst send non-RGV HIE participant data to the RGV HIE if deidentified and explicitly agreed to by Su Clinica**
External rules, regulation, and pressures	HIPAA privacy law, HITECH Act, Texas Administrative Code	HIPAA privacy law, HITECH Act, Texas Administrative Code	HIPAA privacy law, HITECH Act, Texas Administrative Code
System measurement and monitoring	Manual and ad-hoc review	Manual and ad-hoc review	**+ Automated or consistent scheduled manual audits for select measures**

The Need for Iterative Development in Healthcare

The results of the iterations of this audit are consistent with the existing literature. If each iteration was considered a healthcare IT project, the failure rate would be three out of five (60%). Healthcare projects have high failure rates that exceed 50%, with some reports exceeding estimates of 80% (Gesulga et al., 2017; Kaplan & Harris-Salamone, 2009). To address these particular obstacles, iterative development can be used to adapt to specific, unique sociotechnical challenges (Tomoaia-Cotisel et al., 2018). One manner in which iterative development can be brought into the healthcare sector is through the adoption of tactics and strategies associated with software development (Holden et al., 2021). By minimizing the consequences of failures through

test sites and/or limiting scopes, presumptions can be challenged to map ambiguity through iterative frameworks such as agile scrum (Holden et al., 2021).

Additionally, the DataGauge framework (Diaz-Garelli et al., 2019) and other models can be used to map the data, standards, queries, and analysis used to falsify or test relevant claims. The eight dimensions of the sociotechnical model described by Sittig and Singh (2010) can be used to map the social and technical variables of the analysis. Through iterative development, researchers do not expect their initial descriptions and presumptions to be correct; instead, they expect to refine these presumptions through repeated failure. DataGauge introduces software development practices to the healthcare industry by using a use case as a framework. This approach allows for iterative development, enabling continuous improvement in meeting the needs of the use case. Moreover, it recognizes that assumptions underlying the use case may be challenged, leading to potential changes in the use case itself. Through embracing trial-and-error, this method fosters a nuanced approach that refines understanding over time. DataGauge also facilitates the discovery of unknowns and the clarification of ambiguities. This iterative and adaptive process enhances the overall effectiveness of software development in healthcare.

Section 6: Study Limitations

The main limitation of this translational paper is that, although the frameworks and models used to describe the results are generalizable, the specific results cannot be generalized due to the idiosyncrasies of the RGV HIE and the data it collects being particular to the organization and region. Additionally, the application of this methodology may be time-consuming for those seeking to repeat the audit. Incorrect queries, misunderstandings of data standards, query timeouts, and waiting for authorization or access to systems contributed to the time variance of each iteration. Individuals seeking to apply audits within their own systems may

find the time allocation ambiguity difficult to manage, given other impending constraints and deadlines. In addition, vendors were not available for every iteration meeting, which influenced invalid queries in the first iteration but resolved those queries by the fifth. Vendor availability variances may aid or hinder the application of this methodology to other projects.

Section 7: Conclusions

This operationalization implemented the DataGauge framework in a health information exchange (HIE) setting. The purpose was to test the functionality of the model and demonstrate its applicability in any setting with any dataset; specifically, the implementation aimed to determine the number of hepatitis C-positive tests that existed within the RGV HIE. The modified DataGauge framework, as described by Diaz-Garelli et al. (2019), was utilized for this project. The model follows a series of iterations for data extraction and analysis. In each iteration, the auditor defines the claim to be tested, speculates on the relevant data, identifies the standards for evaluation and retrieval, and evaluates the end result.

The proof of concept involved five iterations: two of these iterations successfully received valid queries, while three iterations failed due to invalid data standards. These results highlight the importance of iterative development when working with healthcare data. It was discovered that the RGV HIE, in this case, did not have access to all the necessary clinical data to answer the initial question regarding the number of hepatitis C-positive patients. The RGV HIE only receives information from patients who have consented to share their health information and have been signed off by their physician.To address this limitation, a recommendation has been made based on the workflow presented by Guerrero et al. (2019). The recommendation suggests granting an intermediary actor (referred to as the analyst) access to all the clinic's data, regardless of patient consent status for the RGV HIE. The analyst would then

gather and deidentify the relevant clinical data and, with explicit permission from the clinic, send the information to the RGV HIE. This approach would enable the RGV HIE to legally access non-RGV HIE participant data through deidentified datasets or aggregated count/sum data. The recommendation from this project also emphasizes the need for an RGV HIE employee or script to operate within the clinic's environment on behalf of the clinic, ensuring compliance and providing reports and deidentified data to the RGV HIE.

By implementing this recommended process, the RGV HIE can enhance its preparedness for future clinical questions, grants, partnerships, and public health emergencies. Furthermore, this model can be applied to other data warehouses and HIEs across the nation. In conclusion, this proof of concept has demonstrated the functionality of the DataGauge framework in an HIE setting. It has uncovered the limitations of the RGV HIE's current access to clinical data and proposed a solution to overcome these limitations. The paper also emphasizes the importance of iterative development in healthcare data analysis and implementation, documenting failures, and utilizing data models and sociotechnical models for communication and collaboration.

References

Aliabadi, A., Sheikhtaheri, A., & Ansari, H. (2020). Electronic health record-based disease surveillance systems: A systematic literature review on challenges and solutions. *Journal of the American Medical Informatics Association, 27*(12), 1977–1986. https://doi.org/10.1093/jamia/ocaa186

Arts, D. G. T., de Keizer, N. F., & Scheffer, G. (2002). Defining and improving data quality in medical registries: A literature review, case study, and generic framework. *Journal of the American Medical Informatics Association: JAMIA, 9*(6), 600–611. https://doi.org/10.1197/jamia.M1087

Black, J., Hulkower, R., & Ramanathan, T. (2018). Health information blocking: Responses under the 21st Century Cures Act. *Public Health Reports*, *133*(5), 610–613. https://doi.org/10.1177/0033354918791544

Blumenthal, S. (2017). The use of clinical registries in the United States: A landscape survey. *eGEMs (Generating Evidence & Methods to Improve Patient Outcomes), 5*(1), 26. https://doi.org/10.5334/egems.248

Bohm, M. R., Stone, R. B., Simpson, T. W., & Steva, E. D. (2008). Introduction of a data schema to support a design repository. *Computer-Aided Design, 40*(7), 801–811. https://doi.org/10.1016/j.cad.2007.09.003

Braunstein, M. L. (2018). Healthcare in the age of interoperability: The promise of fast healthcare interoperability resources. *IEEE Pulse, 9*(6), 24–27. https://doi.org/10.1109/mpul.2018.2869317

Cantor, M. (2011). Calculating and improving ROI in software and system programs. *Communications of the ACM*, *54*(9), 121–130. https://doi.org/10.1145/1995376.1995404

Castillo, A., Sirbu, M. A., & Davis, A. (2018). Vendor of choice and the effectiveness of policies to promote health information exchange. *BMC Health Services Research, 18*(1), Article 405. https://doi.org/10.1186/s12913-018-3230-7

Centers for Disease Control and Prevention (CDC). (2023, May 30). *Hepatitis C surveillance in the United States for 2020.* https://www.cdc.gov/hepatitis/statistics/2020surveillance/hepatitis-c.htm

Chen, H., Hailey, D., Wang, N., & Yu, P. (2014). A review of data quality assessment methods for public health information systems. *International Journal of Environmental Research & Public Health, 11*(5), 5170–5207. https://doi.org/10.3390%2Fijerph110505170

Cowie, M. R., Blomster, J. I., Curtis, L. H., Duclaux, S., Ford, I., Fritz, F., Goldman, S., Janmohamed, S., Kreuzer, J., Leenay, M., Michel, A., Ong, S., Pell, J. P., Southworth, M. R., Stough, W. G., Thoenes, M., Zannad, F., & Zalewski, A. (2016). Electronic health records to facilitate clinical research. *Clinical Research in Cardiology, 106*(1), 1–9. https://doi.org/10.1007/s00392-016-1025-6

Dang, D., Dearholt, S., Bissett, K., Ascenzi, J., & Whalen, M. (2022). *Johns Hopkins evidence-based practice for nurses and healthcare professionals: Model and guidelines* (4th ed.). Sigma Theta Tau International.

Dean, B. B., Lam, J., Natoli, J. L., Butler, Q., Aguilar, D., & Nordyke, R. J. (2009). Review: Use of electronic medical records for health outcomes research. *Medical Care Research and Review*, *66*(6), 611–638. https://doi.org/10.1177/1077558709332440

Diaz-Garelli, J.F., Bernstam, E. V., Lee, M., Hwang, K. O., Rahbar, M. H., & Johnson, T. R. (2019). DataGauge: A practical process for systematically designing and implementing quality assessments of repurposed clinical data. *EGEMS (Generating Evidence &*

Methods to Improve Patient Outcomes), 7(1), 32. https://doi.org/10.5334/egems.286

Diaz-Garelli, J. F., Long, A. M., Bancks, M. P., Bertoni, A. G., Narayanan, A., & Wells, B. J. (2021). Developing a Data Quality Standard Primer for Cardiovascular Risk Assessment from Electronic Health Record Data Using the DataGauge Process. *PubMed, 2021*, 388–397. https://pubmed.ncbi.nlm.nih.gov/35308992

DiazVasquez, J. F. (2016). DataGauge: A Model-Driven Framework for Systematically Assessing the Quality of Clinical Data for Secondary Use. *Dissertations (Open Access), 33*. https://digitalcommons.library.tmc.edu/uthshis_dissertations/33

Ebad, S. A. (2020). Healthcare software design and implementation—A project failure case. *Software: Practice and Experience*, *50*(7), 1258–1276. https://doi.org/10.1002/spe.2807

Feder, S. L. (2018). Data quality in electronic health records research: Quality domains and assessment methods. *Western Journal of Nursing Research, 40*(5), 753–766. https://doi.org/10.1177/0193945916689084

Feinstein, A. R. (1969). The epidemiology of cancer therapy: III. The management of imperfect data. *Archives of Internal Medicine*, *123*(4), 448–461. https://doi.org/10.1001/archinte.123.4.448

French, M., & Mykhalovskiy, E. (2012). Public health intelligence and the detection of potential pandemics. *Sociology of Health & Illness, 35*(2), 174–187. https://doi.org/10.1111/j.1467-9566.2012.01536.x

Gesulga, J. M., Berjame, A., Moquiala, K. S., & Galido, A. P. (2017). Barriers to electronic health record system implementation and information systems resources: A structured review. *Procedia Computer Science*, *124*, 544–551. https://doi.org/10.1016/j.procs.2017.12.188

GitHub. (2021). *PlantUML.* https://github.com/plantuml/plantuml

Gold, M., & McLaughlin, C. G. (2016). Assessing HITECH implementation and lessons: 5 years later. *Milbank Quarterly*, *94*(3), 654–687. https://doi.org/10.1111/1468-0009.12214

Gruendner, J., Gulden, C., Kampf, M., Mate, S., Prokosch, H., & Zierk, J. (2021). A framework for criteria-based selection and processing of fast healthcare interoperability resources (FHIR) data for statistical analysis: Design and implementation study. *JMIR Medical Informatics, 9*(4), Article e25645. https://doi.org/10.2196/25645

Guerrero, S. C., Sridhar, S., Edmonds, C., Solis, C. F., Zhang, J., McPherson, D. D., & Bernstam, E. V. (2019). Access to routinely collected clinical data for research: A process implemented at an academic medical center. *Clinical and Translational Science*, *12*(3), 231-235. https://doi.org/10.1111/cts.12614

HIPAA Privacy Rule, 45 CFR § 160-164. (2000). https://www.ecfr.gov/current/title-45/subtitle-A/subchapter-C/part-160?toc=1

Holden, R. J., Boustani, M., & Azar, J. (2021). Agile Innovation to transform healthcare: Innovating in complex adaptive systems is an everyday process, not a light bulb event. *BMJ Innovations*, *7*(2), 499–505. https://doi.org/10.1136/bmjinnov-2020-000574Hovenga, E., & Grain, H. (2013). Health data and data governance. *Studies in Health Technology and Informatics, 193,* 67–92. https://doi.org/10.3233/978-1-61499-291-2-67

Islami, F., Ward, E. M., Sung, H., Cronin, K. A., Tangka, F. K. L., Sherman, R. L., Zhao, J., Anderson, R. N., Henley, S. J., Yabroff, K. R., Jemal, A., & Benard, V. B. (2021). Annual report to the nation on the status of cancer, part 1: National cancer statistics.

JNCI: Journal of the National Cancer Institute, 113(12), 1648–1669. https://doi.org/10.1093/jnci/djab131

Kahn, M. G., Batson, D., & Schilling, L. M. (2012). Data model considerations for clinical effectiveness researchers. *Medical Care, 50*(Suppl.), 60–67. https://doi.org/10.1097/mlr.0b013e318259bff4

Kaliappan, V., & Ali, N. M. (2018). Improving consistency of UML diagrams and its implementation using reverse engineering approach. *Bulletin of Electrical Engineering and Informatics, 7*(4), 665–672. https://doi.org/10.11591/eei.v7i4.1358

Kaplan, B. J., & Harris-Salamone, K. D. (2009). Health IT success and failure: Recommendations from literature and an AMIA workshop. *Journal of the American Medical Informatics Association, 16*(3), 291–299. https://doi.org/10.1197/jamia.m2997

Kushniruk, A., Borycki, E., Kuwata, S., & Kannry, J. (2006). Predicting changes in workflow resulting from healthcare information systems: Ensuring the safety of healthcare. *Healthcare Quarterly, 9*(Sp.), 114–118. https://doi.org/10.12927/hcq..18469

Leon, N., Balakrishna, Y., Hohlfeld, A., Odendaal, W. A., Schmidt, B., Zweigenthal, V., Watkins, J. A., & Daniels, K. (2020). Routine health information system (RHIS) improvements for strengthened health system management. *Cochrane Database of Systematic Reviews, 8,* Article CD012012. https://doi.org/10.1002/14651858.CD012012.pub2

Lin, C. H., Chou, H., & Yang, U. C. (2018). A standard-driven approach for electronic submission to pharmaceutical regulatory authorities. *Journal of Biomedical Informatics, 79*, 60–70. https://doi.org/10.1016/j.jbi.2018.01.006

Mezzanzanica, M., Boselli, R., Cesarini, M., & Mercorio, F. (2015). A model-based evaluation of data quality activities in KDD. *Information Processing & Management, 51*(2), 144–166. https://doi.org/10.1016/j.ipm.2014.07.007

Min, L., Tian, Q., Lu, X., An, J., & In, H. (2018). An open EHR-based approach to improving the semantic interoperability of clinical data registry. *BMC Medical Informatics & Decision Making, 18*(1), 15. https://doi.org/10.1186/s12911-018-0596-8

National Institute of Health (NIH). (2019, May 24). *MEDLINE / PubMed search strategy & Electronic Health Record information resources*. https://www.nInih.gov/services/queries/ehr_details.html

Ness, R. B. (2007). Influence of the HIPAA privacy rule on health research. *JAMA, 298*(18), 2164. https://doi.org/10.1001/jama.298.18.2164

Office Timeline. (2022). *Swimlane diagram* [Illustrations]. https://www.officetimeline.com/swimlane-diagram

Paul, M. M., Greene, C. M., Newton-Dame, R., Thorpe, L. E., Perlman, S. E., McVeigh, K. H., & Gourevitch, M. N. (2015). The state of population health surveillance using electronic health records: A narrative review. *Population Health Management, 18*(3), 209–216. https://doi.org/10.1089/pop.2014.0093

Peng, M., Lee, S., D'Souza, A. G., Doktorchik, C. T. A., & Quan, H. (2020). Development and validation of data quality rules in administrative health data using association rule mining. *BMC Medical Informatics & Decision Making, 20*(1), 75. https://doi.org/10.1186/s12911-020-1089-0

Prasser, F., Kohlbacher, O., Mansmann, U., Bauer, B., & Kuhn, K. (2018). Data Integration for

Future Medicine (DIFUTURE). *Methods of Information in Medicine, 57*(S01), e57–e65. https://doi.org/10.3414/me17-02-0022

Project Management Institute (PMI). (2017). *A guide to the Project Management Body of Knowledge* (6th ed.). Project Management Institute.

Redman, T. (2021). *Data quality: The field guide*. Digital Press.

RGV HIE. (2023). Connected Care Exchange. Retrieved October 1, 2023, from https://www.ccehie.org/

Rudin, R. S., Motala, A., Goldzweig, C. L., & Shekelle, P. G. (2014). Usage and effect of health information exchange. *Annals of Internal Medicine, 161*(11), 803. https://doi.org/10.7326/m14-0877

Salcedo, M. P., Gowen, R., RodríGuez, A. M., Fisher‐Hoch, S. P., Daheri, M., Guerra, L., Toscano, P. A., Gasca, M. L. R., Morales, J., Reyna-Rodriguez, F. E., Cavazos, B., Marín, E., Perez, C. G., Guerra, M. R., Milbourne, A., Varon, M. L., Reininger, B. M., Fernández, M., Ogburn, T., . . . Schmeler, K. M. (2021). Addressing high cervical cancer rates in the Rio Grande Valley along the Texas–Mexico border: a community-based initiative focused on education, patient navigation, and medical provider training/telementoring. Perspectives in Public Health, 143(1), 22–28. https://doi.org/10.1177/1757913921994610

Sharma, M., & Aggarwal, H. (2018). Development of clinical reporting system through HL7 FHIR web services. *International Journal of Sensors, Wireless Communications and Control, 7*(3), 188–197. https://doi.org/10.2174/2210327908666180118101231

Shortliffe, E. H., & Cimino, J. J. (2021). *Biomedical informatics: Computer applications in health care and biomedicine*. Springer.

Sinn, D. H., Gwak, G., Cho, J., Paik, S. W., & Yoo, B. S. (2014). Comparison of clinical manifestations and outcomes between hepatitis B virus- and hepatitis C virus-related hepatocellular carcinoma: Analysis of a nationwide cohort. *PLOS ONE*, *9*(11), Article e112184. https://doi.org/10.1371/journal.pone.0112184

Sittig, D. F., & Singh, H. (2010). A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. *Quality and Safety in Health Care, 19*(3), i68–i74. https://doi.org/10.1136/qshc.2010.042085

Sommerville, I. (2016). *Software engineering*. Pearson Education.

Taggart, J., Liaw, S., & Yu, H. (2015). Structured data quality reports to improve EHR data quality. *International Journal of Medical Informatics, 84*(12), 1094–1098. https://doi.org/10.1016/j.ijmedinf.2015.09.008

Texas Administrative Code, Title 22 §165.1. (1997). https://texreg.sos.state.tx.us/public/readtac$ext.TacPage?sl=R&app=9&p_dir=&p_rloc=&p_tloc=&p_ploc=&pg=1&p_tac=&ti=22&pt=9&ch=165&rl=1

Texas Department of State Health Services (DSHS). (2021, March 2). *IDPS | Notifiable conditions.* https://www.dshs.texas.gov/IDCU/investigation/Notifiable-Conditions.aspx

Tomoaia-Cotisel, A., Farrell, T. M., Solberg, L. I., Berry, C. A., Calman, N., Cronholm, P. F., Donahue, K. E., Driscoll, D. J., Hauser, D., McAllister, J. W., Mehta, S. N., Reid, R. L., Tai-Seale, M., Wise, C. J. E., Fetters, M. D., Holtrop, J. S., Rodriguez, H. P., Brunker, C. P., McGinley… & Magill, M. K. (2018). Implementation of care management: An analysis of recent AHRQ research. *Medical Care Research and Review*, *75*(1), 46–65. https://doi.org/10.1177/1077558716673459

Upadhyay, K. C., & Upadhyay, K. (2020). *D.B.M.S. (database management) with Sql query: Sql (structured query language).* Independently published.

U.S. Department of Health and Human Services. (2021, June 28). *HITECH Act enforcement interim final rule*. Retrieved November 8, 2021, from https://www.hhs.gov/hipaa/for-professionals/special-topics/hitech-act-enforcement-interim-final-rule/index.html

Vest, J. R., & Gamm, L. (2010). Health information exchange: Persistent challenges and new strategies. *Journal of the American Medical Informatics Association, 17*(3), 288–294. https://doi.org/10.1136/jamia.2010.003673

Watt, G. P., Vatcheva, K. P., Beretta, L., Pan, J. Q., Fallon, M. B., McCormick, J. B., & Fisher-Hoch, S. P. (2015). Hepatitis C virus in Mexican Americans: A population-based study reveals relatively high prevalence and negative association with diabetes. *Epidemiology and Infection, 144*(2), 297–305. https://doi.org/10.1017/s0950268815001247

Weiskopf, N. G., Bakken, S., Hripcsak, G., & Weng, C. (2017). A data quality assessment guideline for electronic health record data reuse. *eGEMs (Generating Evidence & Methods to Improve Patient Outcomes)*, *5*(1), 14. https://doi.org/10.5334/egems.218

Wilson, G., Aruliah, D. A., Brown, C. T., Hong, N. C., Davis, M. R., Guy, R., Haddock, S. H. D., Huff, K. D., Mitchell, I. M., Plumbley, M. D., Waugh, B. M., White, E., & Wilson, P. (2014). Best practices for scientific computing. *PLOS Biology, 12*(1), Article e1001745. https://doi.org/10.1371/journal.pbio.1001745

Xu, R., Wang, L., Berger, N. A., Davis, P. B., Kaelber, D. C., & Volkow, N. D. (2022). COVID infection rates, clinical outcomes, and racial/ethnic and gender disparities before and after Omicron emerged in the US. *medRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/2022.02.21.22271300

Yao, E. (2020, June 3). *The data export, transformation, and storage of data in the RGV HIE* [Illustration].

Zafar, A. A., Saif, S., Khan, M., Iqbal, J., Akhunzada, A., Abdul., Almogren, A., & Alamri, A. (2018). Taxonomy of factors causing integration failure during global software development. *IEEE Access*, *6*, 22228–22239. https://doi.org/10.1109/access.2017.2782843

Appendix A: Glossary of Terms

Clinical documentation of hepatitis C – The policies and documentation standards of hepatitis C within the clinical entity.

Clinical workflow analysis – The work performed within clinical entities and the Rio Grande Valley Health Information Exchange to input, store, and retrieve data.

Data – Numerical or text values intended to describe abstract and non-abstract concepts (Redman, 2021).

Dataset agnosticism – A feature of a framework that describes its applicability to any conceivable dataset and is not only applicable to any categorization of datasets.

Data believability/credibility/plausibility – A feature of data that presents information that is deemed (either subjectively through professional acumen or objectively through an empirical assessment) to be within the acceptable range of possibility within the setting from which the data was collected (Diaz-Garelli et al., 2019; Feder, 2018; Weiskopf et al., 2017).

Data completeness – A feature of data that describes the amount of pertinent information that can be derived from said data and is needed for project completion (Diaz-Garelli et al., 2019).

Data concordance – A feature of data that describes its consistency with itself in terms of measurements that are pertinent to the project at hand (Diaz-Garelli et al., 2019).

Data correctness – A feature of data that allows it to represent concepts accurately and reflective of reality in a format that is required of the data (Kahn et al., 2012).

Data quality – A set of empirical measurements and characteristics of the data needed to satisfy the requirements of a project (Redman, 2021).

Data repositories – A set of servers that receive and store data (Redman, 2021).

Data schema – A set of categorizations under which different data may be classified to aid in human and computer readability and retrieval (Bohm et al., 2008).

EHR (electronic health record) – An electronic repository of health-related data that clinical entities use to enter and store their data (Paul et al., 2015).

FHIR (Fast Healthcare Interoperability Resources) – A healthcare-specific data schema used to store and retrieve data (Sharma & Aggarwal, 2018).

Hepatitis C – A liver infection caused by the hepatitis C virus (Sinn et al., 2014).

Hepatitis C Reporting Policy by Health and Human Services – The reporting standard by the Texas Department of Health and Human Services that dictates the timeliness, format, and documentation requirements that define a reportable case of hepatitis C (Texas Department of Health and Human Services [DSHS], 2021).

HIE (health information exchange) – A entity that receives and stores clinical data from clinical entities with the intention of exporting said data for healthcare-related projects (Rudin et al., 2014).

Interoperability – A feature of data to be moved from one data repository or schema to another (Braunstein, 2018).

IT (information technology) – A department in an organization that focuses on developing and maintaining systems reliant on technology (Redman, 2021).

Public health event – An event that pertains to the health of a significant portion of a population (French & Mykhalovskiy, 2012).

Sociotechnical – A description of a process that indicates that involves people and technology working together Sittig & Singh, 2010).

SQL (structured query language) – A specific data schema used to store and retrieve data (Upadhyay & Upadhyay, 2020).

Technical workflow analysis – The work and processes performed within the technological infrastructure of both the clinical entities and the Rio Grande Valley Health Information Exchange.

Workflow – Logical connections of abstractions that represent the work performed in a particular setting (Kushniruk et al., 2006), interactions amongst the people performing the work and technology (Sittig & Singh, 2010).

Appendix B: Project Management Plan

1. Project Overview

The Rio Grande Valley Health Information Exchange (RGV HIE) does not have a vendor-independent data quality assessment process. The lack of such a process does not allow the RGV HIE to view and address data quality issues between data repositories and holistically assess possible issues with export, transform, and load processes. As such, the lack of process has hindered public health reporting in the past and serves as an opportunity to implement the DataGauge Model by Diaz-Garelli et al. (2019), which is designed to be data-agnostic. In this paper, hepatitis C is used as a proxy measurement due to its lower incidence of ~40 out of 100,000 people within the US (Centers for Disease Control and Prevention [CDC], 2018). Data quality metrics used are defined by the Texas Department of State Health Services (DSHS) reportable disease standard for hepatitis C (Texas Department of Health and Human Services [DSHS], 2021).

Data for the translational project will be from various locations: the partner clinical entity (Su Clinica), the Rio Grande Valley Health Information Exchange (RGV HIE) clinical document repository, the RGV HIE SQL Database, and the RGV HIE FHIR Server. The method for evaluation will seek to measure concordance, which is the consistency of values between systems where data is stored (Diaz-Garelli et al., 2019; Weiskopf, Bakken, Hripcsak, & Weng, 2017). The proposed measurement will measure data quality metrics concerning patients diagnosed with hepatitis C or who have positive lab results between data repositories in the RGV HIE ecosystem. Results will be measured by comparing the initial count with the final count. The metric will be expressed as the difference between the initial count and the final count divided by the initial count. The frequency of the measurement will be dependent on the number of changes made. For example, if an alteration in a server's settings is made, the measurement

will be retaken. Measurements will be recorded from electronic queries within Su Clinica, the RGV HIE clinical document repository, the RGV HIE SQL Database, and the RGV HIE FHIR Server. Manual audits of discrepancies between unique patient counts will be conducted in an attempt to resolve said discrepancies.

2. Problem

2.1 Problem Statement and Literature Review

Problem Statement and Literature Review

Summary of Literature
In order to assess data quality, one must engage in a process that measures both the social and technical variables and systems within the organizations related to the data. In addition, it is important to document these variables within the context of stakeholder expectations and constraints. If the expectations and the measurements of data quality do not overlap, then it is suggested to iterate until they do, updating expectations and assumptions based on the data quality fluctuations and changing understanding of the data table and server relationships.

Problem Statement
The Rio Grande Valley Health Information Exchange (RGV HIE) faced challenges in participating in COVID-19 reporting in 2020. This was primarily due to the absence of a standardized vendor-independent data quality assessment process. Without such an assessment, the RGV HIE's ability to effectively respond to future public health events would be compromised. It is important to note that, for this specific project, the RGV HIE IT Team and I deliberately decided to exclude COVID-19 data from the data quality process. Instead, we focused on a lab test with a manageable number of records, allowing us to conduct manual audits within the time frame dedicated to this project. Our emphasis was specifically on hepatitis C, as it has a low incidence rate, according to the CDC (2018).

Review of Evidence

Reference	Year	Key Findings

Diaz-Garelli et al.	2019	The ad-hoc needs of the project define data quality. The burden of mapping and maintaining a model of the data quality requirements can be aided by software development. DataGuage (Diaz-Garelli et al.) is presented as a tool for the "Secondary Use" of clinical data— which directly applies to research as it is not the primary use-case of Electronic Health Record data. The second step of DataGauge (Diaz-Garelli et al.) indicates to "develop a data needs model (DNM) that formalizes the data needs." However, Diaz-Garelli et al. (2019) does not indicate a method to assess the data needs model.
Kahn et al,	2012	Utilizing clinical data for administrative and data analysis purposes requires a robust informatics infrastructure. Said infrastructure needs a Data model suitable for the secondary uses intended for the data.
Weiskopf et al.	2017	Assessing Data Quality requires the incorporation of the intended use of said data. This feature of data quality is described as "task-dependency". It is

		from the associated task the metrics are constructed to measure the quality of the data.
Arts et al.	2002	Clinical data stored in a registry must follow an iterative process of data quality detection and improvement/mitigation. Arts et al. (2002) offers a framework to apply this iterative process to registries, regardless of the analysis needs and data at hand.

3 IT Solution

IT Solutions

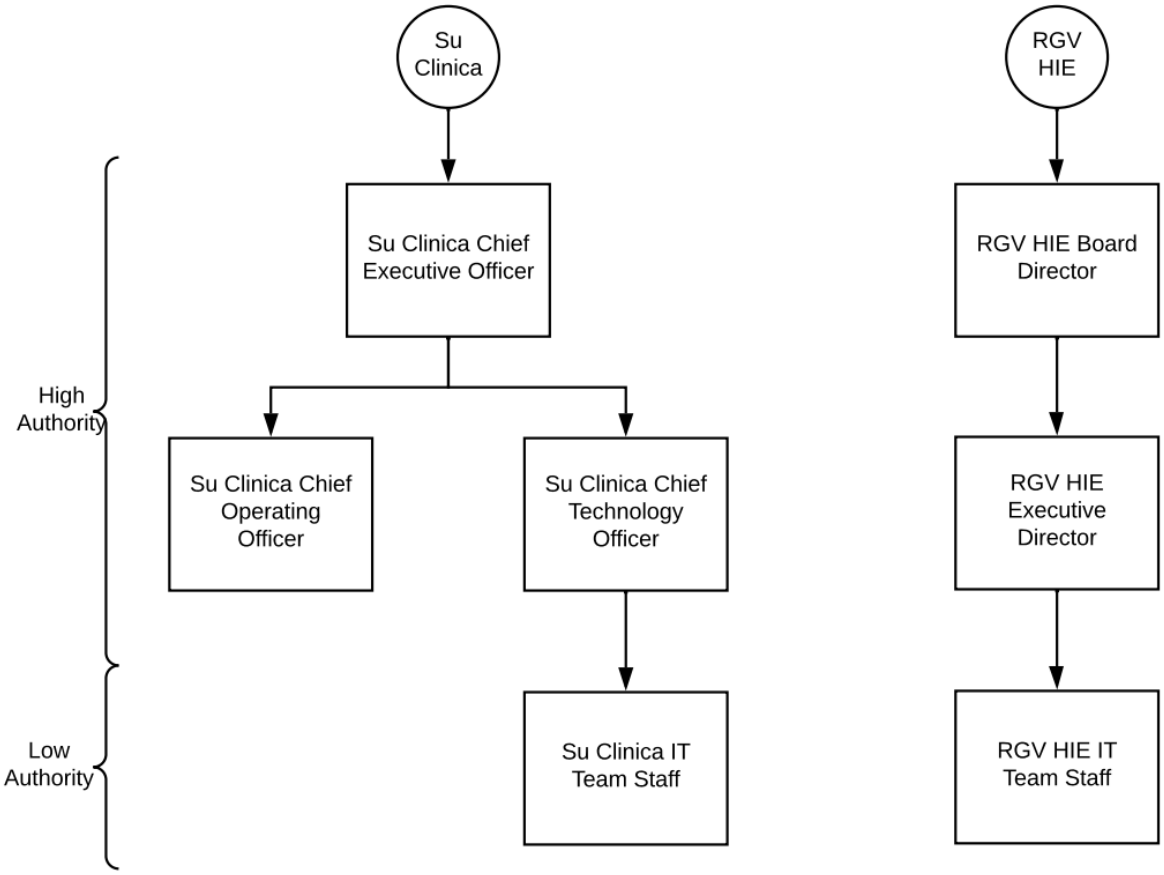
Name IT Solution:	**RGV HIE Independent Data Quality Assessment**
Problem Statement	The Rio Grande Valley Health Information Exchange (RGV HIE) faced challenges in participating in COVID-19 reporting in 2020. This was primarily due to the absence of a standardized vendor-independent data quality assessment process. Without such an assessment, the RGV HIE's ability to effectively respond to future public health events would be compromised. It is important to note that, for this specific project, the RGV HIE IT Team and I deliberately decided to exclude COVID-19 data from the data quality process. Instead, we focused on a lab test with a manageable number of

	records, allowing us to conduct manual audits within the time frame dedicated to this project. Our emphasis was specifically on hepatitis C, as it has a low incidence rate, according to the CDC (2018).
Vendor Name and Website	**NA**
Description of IT Solution	The method for evaluation will seek to measure concordance, which is the consistency of values between systems where data is stored (Diaz-Garelli et al., 2019; Weiskopf, Bakken, Hripcsak, & Weng, 2017). The proposed measurement will measure the count of patients diagnosed with hepatitis C or who have positive lab results between data repositories in the RGV HIE ecosystem. Results will be measured by comparing the initial count with the final count. The metric will be expressed by the difference between the initial count and the final count, divided by the initial count. The frequency of the measurement will be dependent on the number of changes made.

4 Project Integration

4.1 The Organization

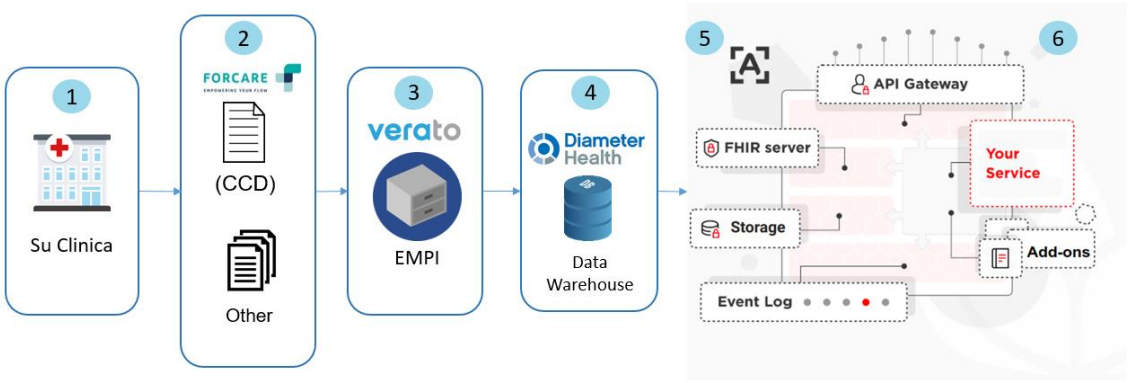
Organization Chart



4.2 Work System

Major Activities or Processes

Major Activities and Processes: The Data Export, Transformation, and Storage of Data in the RGV HIE (Yao, 2020)



In reference to Figure 2: the processes and procedures pertinent to major activities for this project are:

- In stage 1, data is housed and maintained within a clinical entity's Electronic Health Record (EHR), which houses clinical data
- In stage 2, data is exported as CCD's and other supporting documents and stored via the assigned vendor—Forcare.
- In stage 3, the vendor—Verato— creates an Enterprise Master Patient Identifier (EMPI), which will serve as the unique identifier for that patient across all documents received from the various clinics.

- In stage 4, Diameter Health standardizes the data so that it can be stored and retrieved in a systematic manner in a SQL (Structured Query Language) database.

- In stage 5, the data is also sent to Health Samurai's Aidbox FHIR (Fast Healthcare Interoperability Resources) Server, which translates the available data points from the Data Warehouse to the FHIR schema. Data in the FHIR server can be interfaced by third parties via the available API (Application Programming Interface)

Work System

Customers	**Products and Services**

End users receiving and utilizing RGV HIE data
• RGV HIE Staff
• Clinical Entities
  ○ Clinical Staff
  ○ IT Staff
  ○ Community Health Workers
  ○ Administrative Staff
• Third-Party Entities Seeking RGV HIE data
  ○ End Users of said Third-Party Entities
	Products and services involved in data quality assessment are:
	• Logistics
	• Establishing data
	• Inform ETL (Export, Transform, Load) Pipeline

Participants	**Information**	**Technologies**

• Clinical Entity IT Team
• Clinical Entity EHR Vendor Team
• Forcare Vendor
• Verato Vendor
• Diameter Health Vendor
	• Continuity of Care Document
	  ○ Summary of clinical encounter including
	  ▪ Labs
		• Clinical Entity EHR
		• Forcare CCD query and storage
		• Verato unique patient id creation

• Health Samurai Aidbox Vendor • Amazon Web Services Vendor	▪ Medications prescribed ▪ Diagnosis ▪ Visit Date ▪ Visit location ▪ Provider information	• Diameter Health parsing and storage of CCD data • Health Samurai and Aidbox FHIR server • Amazon Web Services servers

Source: Steven Alter, *The Work System Method: Connecting People, Processes, and IT for Business Results,* Work System Press, 2006

5 Project Charter and Scope

5.1 Scope Statement

5.1.1 Project Purpose and Justification.

The Rio Grande Valley Health Information Exchange (RGV HIE) faced significant challenges in 2020 due to the absence of a standardized vendor-independent data quality assessment. This lack of a comprehensive evaluation process had a direct impact on the RGV HIE's ability to actively participate in COVID-19 reporting. Consequently, the RGV HIE's capacity to effectively respond to future public health events is severely impeded unless a robust data quality reporting system is put in place. In order to address this crucial issue of low data quality, the first step is to establish an independent data quality assessment process specifically tailored to the needs of the RGV HIE. By doing so, the RGV HIE will be equipped with a valuable tool to evaluate and verify the accuracy, completeness, and reliability of the data it handles. This assessment process will not only ensure that the RGV HIE can effectively contribute to COVID-19 reporting but will also enhance its overall readiness to tackle future public health events. It is important to note that, for this specific project, the RGV HIE IT Team and I deliberately decided to exclude COVID-19 data from the data quality process. Instead, we

focused on a lab test with a manageable number of records, allowing us to conduct manual audits within the time frame dedicated to this project. Our emphasis was specifically on hepatitis C, as it has a low incidence rate, according to the CDC (2018).

5.1.2 Scope Description.

Auditing will be done between one clinical entity and the RGV HIE servers. Hepatitis C will be used as a proxy measurement for any future desired measurements due to its lower incidence of 3.8 out of 100,000 people within the US (Centers for Disease Control and Prevention [CDC], 2018). This is possible because the proposed framework, the DataGauge Model, which is agnostic to any data measurement (Diaz-Garelli et al., 2019). In addition, I will use the Texas Department of State Health Services (DSHS) reportable disease standard for hepatitis C (Texas Department of Health and Human Services [DSHS], 2021). The DSHS's reportable disease list contains similar reporting standards governing delivery times, accuracy, and format for other diseases to which this methodology may apply within the RGV HIE.

5.1.3 Boundaries/Strategies.

In-Scope: Audits will focus on FHIR and SQL servers on the RGV HIE end, as well as a singular clinical entity connected to the RGV HIE. Hepatitis C lab reports and diagnoses will be counted on a unique individual basis, adhering strictly to the DSHS standard for hepatitis C (see Appendix A). The workflow and ROI analysis will only apply to actions performed under this project. Out of Scope: Audits will not include patient charts that contain implicit diagnoses or elevated hepatitis C lab reports recorded in free text. The implementation of the proposed workflow for this use case will not be conducted.

5.1.4 Assumptions.

I assumed that the HIE would be able to provide credentials and access to their relevant servers. If I had not gotten the credentials, I would not have been able to proceed with my analysis. A significant assumption was made regarding the incident rate of hepatitis C, with the expectation that it would result in a manageable number of cases. This assumption was essential for the feasibility of conducting audits within a reasonable timeframe. If the number of hepatitis C cases surpasses a certain threshold, it would impact the time required for reviewing all cases. Consequently, this could impede the progress of the project and hinder the comprehensive analysis of discrepancies. Furthermore, it was assumed that the Health Information Exchange (HIE) would be able to provide the necessary credentials and access to their relevant servers. Without obtaining these credentials, proceeding with the analysis would not have been possible.

5.1.5 Constraints.

One constraint was the need for more funding to support any additional costs that may have arisen during the project. If any process required additional funding, it would have halted the project. Another constraint was the limited access to teams. Throughout the project, only the HIE IT team, the Clinic IT Team, and the vendor consultants were available. Ideally, regular meetings with clinical staff, such as nurses and physicians, could address data discrepancies caused by user input. A constraint is the need for more funding to initiate and maintain operations that may incur stakeholder costs.

5.2 Requirements/Characteristics

Requirements/Characteristics

Numbered	Desired Functionality	Existing Functionality	Change / New	Justification for the Desired Functionality	Stakeholders / Business impacted	Priority

1.	Measured discrepancies of hepatitis C lab test counts between the RGV HIE and a singular clinical entity	RGV HIE can only report what is in their system	Change	Measuring outcomes directly from the clinical entity will bypass the vendors and provide a vendor independent data quality assessment	RGV HIE Su Clinica (clinical entity) Health Samurai (RGV HIE Vendor) Diameter Health (RGV HIE Vendor)	High
2.	Workflow description of actions taken under this project through the DataGauge Model/Framework	NA	New	Provides an initial workflow through a data agnostic governance framework that will allow the RGV HIE to implement and change if they so choose	RGV HIE	High
3.	UML diagram of the current state of the HIE as it relates to Hepatis C Labs from Su Clinica	NA	New	Will provide a data model of the current state, and will allow for collaboration and communication amongst stakeholders	RGV HIE	High
4.	Reasons why discrepancies exist (if any)	NA	New	Will allow the HIE to begin to address the issues, if they so desired, and investigate further	RGV HIE	Medium
5.	Recommendations on how to proceed given the current state	NA	New	Will give a starting point for the HIE to begin to address the issues at hand	RGV HIE	Medium
6.	Estimated costs of staffing workflow	NA	New	This will inform the ROI analysis for the RGV HIE to consider feasibility of implementing such a workflow	RGV HIE	Low
7.	Return on Investment assuming	NA	New	This will provide an empirical number for the	RGV HIE	Low

	workflow is staffed	RGV HIE to use as guide on whether staffing and scaling such a workflow is feasible and/or beneficial for them

5.3 Acceptance Criteria

The acceptance criteria define the boundaries of the user requirements and will be used to confirm that the proof-of-concept of the DataGauge is working as intended. The DataGauge will provide an output of assumptions, queries, and evaluations of those queries and assumptions in a step-wise, iterative manner. As such, the HIE will be able to reproduce the results and the logic behind why those queries were made. The proof-of-concept DataGauge:

- The recommendations must be high-level enough to apply to any proposed data measurement but low enough to allow them to be translated to staffing duties.
- There must be an initial and final state of the data models created using the DataGauge, indicating both the initial and final understanding of the data and its relationships.
- The work done must not have incurred additional costs for the HIE unless agreed upon ahead of time.
- Each task done in order to conduct the discrepancy review must be able to be performed by the HIE.
- A list of queries, outputs, and errors from the DataGauge process will be provided and listed.
- No patients will be contacted to obtain any data or information.

5.4 Project Deliverables

The deliverables produced from the implementation of DataGauge within the HIE environment. These variables will be assessed by the HIE IT Team and the Executive Director using the criteria described here. If these deliverables are not satisfied, then the project will not be deemed a success.

1. Deliver the numerical discrepancy (if any exists) of hepatitis C diagnosis counts between the Clinical System
    a. Results will be measured by comparing the initial count with the final count.
    b. The metric will be expressed as the difference between the initial count and the final count divided by the initial count.
2. Deliver the presumed and final understanding of the data model of hepatitis C lab data and its relationships to the various servers in the HIE
3. Deliver recommended workflow of the process to allow the RGV HIE to adopt
4. Deliver Return on Investment report if the workflow is staffed

5.5 SWOT Analysis

SWOT Analysis

Problem Statement

The Rio Grande Valley Health Information Exchange (RGV HIE) faced challenges in participating in COVID-19 reporting in 2020. This was primarily due to the absence of a standardized vendor-independent data quality assessment process. Without such an assessment, the RGV HIE's ability to effectively

respond to future public health events would be compromised. It is important to note that, for this specific project, the RGV HIE IT Team and I deliberately decided to exclude COVID-19 data from the data quality process. Instead, we focused on a lab test with a manageable number of records, allowing us to conduct manual audits within the time frame dedicated to this project. Our emphasis was specifically on hepatitis C, as it has a low incidence rate, according to the CDC (2018).

Internal Factors

Strengths (+)	**Weaknesses (-)**
Experts in FHIR and SQL languages are available for query consultation Dedicated RGV HIE Team to assist in access and authorization to data repositories within the HIE. Experts in data analytics are available to evaluate the results of queries Support of the project within the Executive and Administrative teams of the RRGV HIE and Su Clinica.	Project is not the highest priority within the contexts of other initiatives. Clinical Entity will not allow remote EHR access and all reporting from the EHR must be done within the clinical wall of the entity. Queries for hepatitis C have never been created for the FHIR and SQL servers. As such, a review of data variables and database schemas is needed to begin work on optimizing said queries.

	There are no funds available that can be solely dedicated to this project; all funds dedicated are primarily serving other initiatives.
External Factors	
Opportunities (+)	**Threats (-)**
HIEs are aware they may serve as a data source for public health reporting and monitoring public health reporting in the region. The workflow of hepatitis C data quality may be applicable to other reportable diseases. Data quality measurements can aid in continuous improvement by serving as a reference point.	If the project requires costs outside the budget, the project may be stalled by the RGV HIE. The RGV HIE heavily relies on the availability of several vendors; if the companies representing these vendors become unavailable or financially insolvent, then the project may be halted. Data Quality concerns that are connected to vendor proprietary operations will be

Increased surveillance support can lead to further funding to aid in the treatment and payment of diseases affecting the region.	heavily dependent on vendor response time in order to resolve. Delayed responses may hinder the project timeline. If another public health event occurs, priority could be shifted, and the current project may also be shifted or halted depending on the priorities of Su Clinica and the RGV HIE.

6 Project Schedule Management

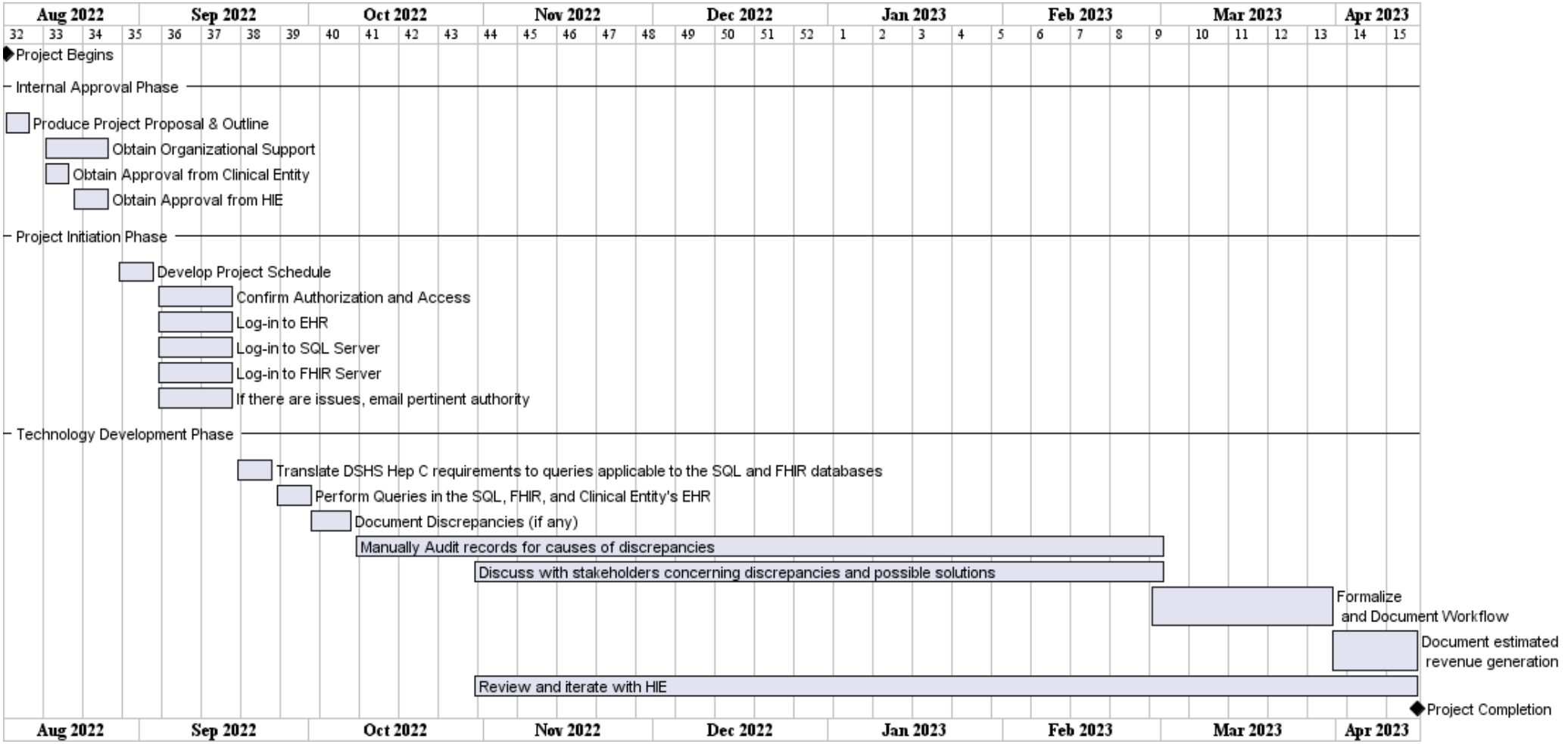
6.1 Schedule Development

Schedule Development

Task Name	Duration	Scheduled Start	Scheduled Finish
- Internal Approval Phase			
-- Project Proposal Development			
--- Produce Project Proposal & Outline	4	8/8/2022	8/12/2022
--- Obtain Organizational Support	11	8/15/2022	8/26/2022
---- Obtain Approval from Clinical Entity	4	8/15/2022	8/19/2022
---- Obtain Approval from RGV HIE	6	8/20/2022	8/26/2022
- Project Initiation Phase			
-- Develop Project Schedule	7	8/27/2022	9/3/2022
-- Confirm Authorization and Access	13	9/4/2022	9/17/2022

--- Log-in to EHR	13	9/4/2022	9/17/2022
--- Log-in to SQL Server	13	9/4/2022	9/17/2022
--- Log-in to FHIR Server	13	9/4/2022	9/17/2022
--- If there are issues, email pertinent authority	13	9/4/2022	9/17/2022
- Technology Development Phase			
-- Translate DSHS Hep C requirements to queries applicable to the SQL and FHIR databases	6	9/18/2022	9/24/2022
-- Perform Queries in the SQL, FHIR, and Clinical Entity's EHR	6	9/25/2022	10/1/2022
-- Document Discrepancies (if any)	7	10/1/2022	10/8/2022
-- Manually Audit records for causes of discrepancies	143	10/9/2022	3/1/2023
-- Discuss with stakeholders concerning discrepancies and possible solutions	122	10/30/2022	3/1/2023
-- Formalize and Document Workflow	32	2/27/2023	3/31/2023
-- Document estimated revenue generation	15	3/31/2023	4/15/2023
-- Review and iterate with RGV HIE	167	10/30/2022	4/15/2023
- Project Completion			

Gannt Chart Illustration



6.2 Schedule Control

The student will hold weekly project status meetings with the RGV HIE to review:

- The current issues documented in the issue log
- Updates to deliverables and acceptance criteria
- Updates to task completion schedule as well as phases of the project
- Review of closed issues/problems and next issues/problems that will be dealt with by next meeting

7 Budget

Proposed 1 Year Total Cost of Ownership (TCO) 2022-2023

Vendor Cost	**Year 1**	**Total**
Amazon Web Services	$72,000.00	**$72,000.00**
Vendor 1	$65,000.00	**$65,000.00**
Vendor 2	$50,000.00	**$50,000.00**
Vendor 3	$75,000.00	**$75,000.00**
Vendor 4	$15,000.00	**$15,000.00**
Vendor 5	$10,000.00	**$10,000.00**
Organizational Cost	**Year 1**	**Total**
Data Management Team	$14,287.50	**$14,287.50**
Vendor and Organizational Total	**Year 1**	**Total**
Vendor Total	$287,000.00	**$287,000.00**
Organizational Total	$14,287.50	**$14,287.50**
Taxes	$24,856.22	**$24,856.22**
Grand Total	$326,143.72	**$326,143.72**

Estimated Cost Breakdown

Data Management Team	No. of Required Build Hours per Year	Rate per Hour	Total
Amazon Web Services	50	$150.00	**$7,500.00**
FHIR Consultant	50	$75.00	**$3,750.00**
HIE Integration Analyst	50	$34.31	**$1,715.50**
HIE Data Analyst	50	$26.44	**$1,322.00**
Total	200	**$285.75**	**$14,287.50**

8 Quality Assurance

Prepared by:	Edward Yao
Date (MM/DD/YYYY):	3/8/2022

8.1 Deliverables and Acceptance Criteria

Deliverables	Acceptance Criteria / Applicable Standards
1. Deliver the empirical discrepancy (if any exists) of hepatitis C diagnosis counts between the Clinical System	Access must be done in a HIPAA-compliant manner and within pertinent organizational policies. As such, Analytics should be done within permitted devices within the RGV HIE ecosystem. All Results and data transfers outside of the RGV HIE ecosystem should be preapproved by the RGV HIE.
2. Deliver documented workflow of process to allow the RGV HIE to adopt	The workflow must apply to the RGV HIE organizational structure. The RGV HIE must approve the proposed workflow.
3. Deliver Return on Investment report if workflow is staffed	The Return on Investment must be applicable to the RGV HIE. The RGV HIE must approve the proposed Return on Investment report.

8.2 Quality Assurance Activities

- What steps will you take to ensure that Quality is built into the production processes?

I will work directly with the HIE IT Team, the Clinic IT Team, and the Vendor Consultants to assess the data quality from the DataGauge iterations. The proof-of-concept DataGauge process calls for the pertinent question to be declared, the data needed to answer the question, the standards to be placed on that data, the logistics to pull data, and finally, the validation of the results obtained from the data. In order to ensure that Quality is within the production process, quality assurance activities will be placed in the final step, which is the validation of results. In that step, the HIE IT Team, the Clinic IT Team, and the Vendor consultants will review and analyze deliverables at the end of each iteration.

- Will the test team work from a Test Plan? Do they understand their responsibilities?

The test team will work with a Test Plan based on the DataGauge framework. This plan will dictate the data and tasks they will oversee. The HIE Team will provide guidance and support for all HIE operations, including the document repository, enterprise master index vendor, clinical data repository, and APIs. The Clinic IT Team will consult on the Clinic's EHR and its interaction with the HIE to ensure seamless integration and functionality of data operations, specifically in pushing documents from the EHR to the HIE. The FHIR consultant will assist with any querying and data storage questions related to the FHIR Server. All evaluations will be conducted within the context of the DataGauge fifth process, where the entire team will analyze the question, data, data standards, and logistics to extract data from their respective systems and understand how these systems interact at key integration points.

- How will you ensure that Requirements are correct, complete and accurately reflect the needs of the Customer?

Throughout the iteration phases, the team will convene weekly to address the reports and procedures implemented. Compliance will be consistently reviewed during these feedback sessions. Each session will involve the team assessing the progress made and discussing any necessary adjustments to ensure compliance. If adjustments are deemed necessary, a collaborative plan of action will be developed to address them. These considerations will occur at the end of each iteration within the final phase of the DataGauge framework.

- How will you verify that Specifications are an accurate representation of the Requirements?

Multiple team members will thoroughly review both the specifications and requirements to ensure their alignment and prevent any oversight during the creation of specifications. Testing will involve a comprehensive examination of the DataGauge framework in relation to its relevant integration points, as well as validating the integration between systems. As previously mentioned, the requirements will be continuously reviewed in accordance with HIPAA compliance and their applicability to the HIE environment. Each data point must undergo verification to confirm its relevance to the queries. The data must adhere to the necessary standards for compliance, and it should only be sourced from systems with appropriate user access and authorization controls.

8.2 Quality Assurance Activities

- What steps will you take to ensure that the project plan (e.g. Risk Management Plan, Change Management Plan, Procurement Plan) is followed?

Weekly meetings will be scheduled to ensure that updates are thoroughly documented and discussed with all relevant parties. A comprehensive document will be maintained using the proof-of-concept DataGauge table to record information from past meetings and current assumptions that inform the data analysis questions. At the conclusion of each iteration of the DataGauge framework, all comments and evaluation results will be diligently documented in the fifth step of the process prior to commencing the next iteration.

- Describe how *Requirement – Specification – Test Plan* traceability is managed:

The management of this process will be centralized under the project manager, with input from relevant team members. Specifications and requirements will undergo continuous review and iteration, incorporating feedback from stakeholders. The test plans will be flexible, allowing for adjustments in response to changes or updates in the specifications and requirements while ensuring compliance and applicability. This adaptability is made possible through the utilization of the DataGauge Framework, which guides the identification of relevant questions, data investigation, adherence to data standards, and formulation of data queries. The DataGauge Framework enables the test plans to accommodate shifts in presumptions about the specifications and requirements while the team ensures compliance and applicability.

- What audits and reviews are required and when will they be held?

The team will conduct weekly reviews to discuss documented empirical measurements. However, evaluations will only take place at the end of each iteration. The duration of each iteration is uncertain due to the DataGauge framework's purpose of investigating ambiguities in assumptions, data standards, and data queries. These ambiguities make it difficult to determine timelines. Nevertheless, the weekly meetings will track progress and serve as a consistent documentation process, unlike the evaluations. During the evaluations, discrepancies between data repositories will be examined in relation to the query code and/or procedures used. Any issues related to these discrepancies will be addressed and documented in a master continuous document, presented in a tabular format.

- What steps will you take to ensure that the Vendor is supplying deliverables of adequate quality?

During the evaluation process or in the weekly meetings with the team, if any discrepancies are discovered, appropriate tickets will be created and sent to the responsible vendor for resolution. Vendors are consulted through ticketing systems to track and address issues within their organization over time effectively. The RGV HIE team members will promptly submit tickets and arrange meetings with the relevant vendors to resolve any outstanding issues.

8.2 Quality Assurance Activities

- What will you measure to determine if the project is out of Scope?

The evaluation process involves weekly meetings to discuss any discrepancies that have been observed. If any issues are found, tickets are created and sent to the relevant vendor. Each vendor has their own ticket. Additionally, the evaluation process is the final step in the DataGauge process at the end of each iteration, before any repetition of previous steps occurs. This process entails evaluating the data, its standards, and the queries related to the question we are trying to answer. As each iteration focuses on a different question, there will be different variables, such as data, standards, and queries. Throughout each iteration, the team not only evaluates the relevance of these variables but also ensures compliance.

- What will you measure to determine if the project is within budget?

The Cost Sheet will be reviewed and addressed throughout the project to identify areas of overspending in relation to the predetermined budget. As this project does not anticipate any additional costs beyond what has already been allocated, the team will evaluate and monitor for new line items or unexpected expenses that may arise. This can be achieved by retrospectively reviewing monthly expenses and proactively considering upcoming tasks on the to-do list that may result in additional expenses.

- What will you measure to determine if the project is within schedule?

The team and I will assess the project's progress by referring to the Project Schedule and Gannt chart. Our evaluation will involve checking if any tasks are behind or ahead of schedule. If any modifications are necessary, the team and I will make adjustments to the schedule and Gantt Chart during the evaluation stage of the DataGauge process. This stage will also include determining the next steps and addressing any questions regarding data, data standards, and authorization and control access for querying the data.

8.3 Project Monitoring and Control

Define the following:

- How will you ensure that adequate testing is done? How do you define “adequate”?

The weekly meeting reviews will address the specific process and actions done by the student and if both querying and manual audits were done within compliance. The weekly meeting reviews will focus on the specific process and actions carried out by the student, ensuring compliance through both querying and manual audits. The DataGauge framework is specifically designed to facilitate comprehensive testing and adherence to the necessary standards and protocols. Each iteration will be thoroughly tested and observed for compliance. Additionally, the team will utilize the appropriate standards determined for each specific iteration, evaluating them during the fifth stage of the process.

- How will you report and resolve variances from acceptance criteria?

8.3 Project Monitoring and Control

The acceptance criteria may vary depending on the cause of the variance. For instance, if the variance is a result of a misunderstanding of the acceptance criteria, it can be resolved through clarification and effective communication with the relevant stakeholders. On the other hand, if the variance is caused by a deviation in the implementation, the root cause needs to be identified and necessary adjustments made to align with the acceptance criteria. In cases where resolutions or issues require additional funding, the project may be canceled due to the lack of budget to support any associated costs.

- At what milestones will testing and reviews take place – who and how will they do them?

Reviews will be conducted weekly by the HIE team and I. Testing will take place at the end of each DataGauge iteration. For every practice question, we will thoroughly analyze the data to ensure compliance with data standards, authorization, and access requirements. If necessary, we will determine the appropriateness of accessing any clinical information or data not relevant to this project. The proof-of-concept DataGauge will be utilized by the team to review past requirements and evaluate compliance metrics and standards specific to each iteration.

- What action by the Sponsor constitutes acceptance of deliverables at each phase?

The sponsor will actively participate in the weekly meetings and iteration evaluations. They will rely on the HIE IT team and I to assess whether the outputs meet the established quality standards and compliance requirements agreed upon at the beginning of each iteration. All data accessibility, including the queries and their results, will be thoroughly reviewed using this criteria. It is important to note that the sponsor can halt the project at any given time.

9 Stakeholder Analysis

Tier 1 Stakeholders: Senior Leaders and Key Decision Makers

Ensuring project feasibility	**Name of person/group**	**Why exactly is this person/group important?**
Who can help fund the initiative?	• Dr. Elena Marin, Clinical Entity (Su Clinica) CEO	• Dr. Elena Marin can obtain funding from internal sources for experimentation projects within the IT (Information Technology) department
	• Dr. Sheila Magoon, RGV HIE Board Director	• Dr. Sheila Magoon can leverage her position to seek funding internally to address possible events and raise awareness within the RGV HIE board
Who can provide additional resources?	• Humberto (Bert) Gonzalez, Clinical Entity (Su Clinica) Chief Technology Officer	• Humberto Gonzalez has discretionary funds for IT projects that could aid quality improvement projects
	• Andrew Lombardo, Executive Director (RGV HIE)	• Andrew Lombardo can allocate consultant and RGV HIE staff time. Andrew also has discretionary funds under the RGV HIE
Who can decide whether or not the project can proceed, be terminiated or put on hold?	• Dr. Elena Marin, Clinical Entity (Su Clinica) CEO	• Dr. Marin must give the go-ahead to all projects taking place in Su Clinica
	• Cristina Perez, Clinical Entity (Su Clinica) Chief Operating Officer	• Cristina Perez oversees the operational logistics of major projects in Su Clinica and must approve all new projects

	• Andrew Lombardo, Executive Director (RGV HIE) • Humberto (Bert) Gonzalez, Clinical Entity (Su Clinica) Chief Technology Officer	• Andrew Lombardo must approve all operations within the RGV HIE • Humberto Gonzalez must approve and be involved in the planning of data operations
Who can remove obstacles and barriers that are beyond the project team's control?	• Andrew Lombardo, Executive Director (RGV HIE) • Humberto (Bert) Gonzalez, Clinical Entity (Su Clinica) Chief Technology Officer • Cristina Perez, Clinical Entity (Su Clinica) Chief Operating Officer	• Andrew Lombardo has the authority to request action items and allocate work hours. • Humberto Gonzalez is a board member of the RGV HIE and the CTO of the Clinical Entity. He both internally allocates resources but also influences the RGV HIE Board as well as the Clinical Entity Leadership • Cristina Perez has the authority and relationships to change work hours and assign tasks to staff.
Who needs to approve/sign-off on deliverables?	• Dr. Elena Marin, Clinical Entity (Su Clinica) CEO • Cristina Perez, Clinical Entity (Su Clinica) Chief Operating Officer • Andrew Lombardo, Executive Director (RGV HIE)	• Elena Marin will have the authority to approve/sign off on deliverables as they pertain to the Clinica Entity • Cristina Perez will have project authority as it pertains to the Clinical Entity and approval of deliverables • Andrew Lombardo will have overall authority to approve/sign off on

	• Humberto (Bert) Gonzalez, Clinical Entity (Su Clinica) Chief Technology Officer	deliverables as they pertain to the Clinica Entity • Humberto Gonzalez will review the procedures and workflow to ensure he can support the activities and that they stay within compliance
Who can help build additional senior level political support?	• Andrew Lombardo, Executive Director (RGV HIE)	• Andrew Lombardo, as the RGV HIE Executive Director, advocates for the project amongst the board and the administrators of clinical partners.
	• Cristina Perez, Clinical Entity (Su Clinica) Chief Operating Officer	• Cristina Perez directly oversees all significant operations within Su Clinica and is the primary trusted source with Dr. Elena Marin and the executive team
	• Humberto (Bert) Gonzalez, Clinical Entity (Su Clinica) Chief Technology Officer	• Humberto Gonzalez, as the CTO, can provide additional data reporting and outcome contexts.

Tier 2 Stakeholders: Project Contributors

Ensuring the quality of deliverables and activity execution:	**Name of person/group**	**Why exactly is this person/group important?**
Where can we find the required project resources	• Andrew Lombardo, Executive Director (RGV HIE)	• Andrew will manage the project from a human resource perspective and allocate necessary staff and consultant time
	• Cristina Perez, Clinical Entity (Su Clinica) Chief Operating Officer	• Cristina will provide staffing from the clinical entity to aid authorization and access to clinical data
Where can we find required SMEs?	• Clinical Entity (Su Clinica) IT Team	• The Su Clinica IT team will be the SME regarding the interoperability of clinical data with the RGV HIE. They will also provide access to their EHR directly and monitor for compliance
	• RGV HIE Team	• The RGV HIE Team will oversee all data extracts and compare between data silos contained within the HIE. Queries and results of those queries will be documented for such comparison
Who can provide support in the areas of training and competency development?	• RGV HIE Team • Clinical Entity (Su Clinica) IT Team	• Both the RGV HIE and Su Clinica IT teams manage the training and competency development within their respective organizations
What groups can help us publicize/communicate this initiative	• Andrew Lombardo, Executive	• Andrew has the contact of clinical partners and the board. As the Executive

	Director (RGV HIE) • Cristina Perez, Clinical Entity (Su Clinica) Chief Operating Officer	Director, he sends updates and coordinates communication campaigns with RGV HIE stakeholders • Cristina is in direct contact with the executive and administrative team of Su Clinica. Within her weekly meetings, she can directly communicate relevant details of the project
Who can help us support the initiative once it is deployed?	• RGV HIE Team	• The RGV HIE Team provides analysis and technical support through their services to the clinical partners

Tier 3: Stakeholders Recipients

Areas where people/groups may be impacted:	**Name of person/group**	**Why exactly is this person/group important?**
Who is the intended audience for the project outputs or the change?	• Clinical Entity (Su Clinica) • RGV HIE	• Su Clinica and the RGV HIE will have a quality assessment of the data feed and mapping operations
Will the change have any effect on secondary groups or individuals?	• Providers • Nurses	• Data quality measurements may inform changes to Su Clinica's workflows and standard operating procedures for clinical staff

10 Human Resource Management Plan

Roles, Responsibilities, and Authority

Role	Responsibility	Authority
Su Clinica Chief Operating Officer	The highest-ranking individual within the institution is ultimately responsible for administrative and managerial decisions	High Authority
Su Clinica Chief Operating Officer	Responsible for the planning, implementation and sustainability of projects within the organization	High Authority
Su Clinica Chief Technology Officer	Responsible for the logistical implementation of IT solutions within the organization	High Authority
RGV HIE Board Director	High active and influential member of the RGV HIE Board who approves projects	High Authority

RGV HIE Executive Director	The person responsible for the implementation of goals and projects approved by the RGV HIE board	High Authority
RGV HIE IT Team Staff	RGV HIE team works under the supervision of the RGV HIE Executive Director; they ensure the storage, transfer, and retrieval of data within compliance	Low level of authority
Su Clinica IT Team Staff	The Su Clinica IT Team Staff team works under the supervision of the Su Clinica Chief Technology Officer to implement IT projects within the organization	Low level of authority

11 RACI Chart

RACI Chart

R = Responsible A = Accountable C = Consult I = Inform

Activity	Su Clinica Chief Executive Officer	Su Clinica Chief Operating Officer	Su Clinica Chief Technology Officer	Su Clinica IT Team Staff	RGV HIE Board Director	RGV HIE Executive Director	RGV HIE IT Team Staff	Edward Yao (Translational Student
Project Proposal Development	I	I	C	C	I	A	C	R
Produce Project Proposal & Outline	I	I	C	C	I	A	C	R
Obtain Approval from Clinical Entity	R	R	R	I	I	I	I	R
Obtain Approval from RGV HIE	I	I	I	I	R	R	R	R
Develop Project Schedule	I	I	R	I	I	R	I	R
Confirm Authorization and Access to FHIR Server	I	I	I	I	I	A	R	R
Confirm Authorization and Access to SQL Server	I	I	I	I	I	A	R	R
Confirm Authorization and Access to EHR	I	I	A	R	I	I	I	R
Translate DSHS Hep A Requirements to SQL and FHIR Queries	I	I	I	I	I	I	R	R
Perform Queries	I	I	I	C	I	I	C	R
Document Discrepancies	I	I	I	C	I	I	C	R

Audit Discrepancies	I	I	I	C	I	I	C	R
Discuss Solutions to Discrepancies	I	I	I	R	I	I	R	R
Formalize and Document Workflow	I	I	I	C	I	I	C	R
Document Revenue Generation	I	I	C	C	I	C	C	R
Review and iterate over solutions, workflow, and revenue generation analysis	I	I	C	C	I	C	C	R

12 Communication Plan

Communication Plan

#	Recipient	Message	Assumptions	Timeline	Channel	Recipients Response	Responsible
1	Dr. Elena Marin	Progress of the project milestones and adverse events	Communication will be high level concerning major institutional priorities and constraints	On as-needed basis	email	Response not required, in the case of adverse events a response is required on the strategies and timeline of resolving the issues	Edward Yao
2	Dr. Sheila Magoon	Progress of the project milestones and adverse events	Communication will be high level concerning major institutional priorities and constraints	On as-needed basis	email	Response not required, in the case of adverse events a response is required on the strategies and timeline of resolving the issue	Edward Yao
3	Humberto (Bert) Gonzalez	All technical and compliance issues and requests during the project	Communication will be low-level and specific in terms of authorization use, data flow, and compliance as it pertains to the clinical entity	On as-needed basis	email	Approvals and disapprovals of current list tasks, strategies, and results	Edward Yao
4	Cris Perez	All managerial and administrative barriers issues; any requests for resource allocation changes	Communication will involve current staff time and resource usage, along with any contact information of people who need to be contacted	On as-needed basis	email	Response not required; in the case of resource requests and adverse events, response is required to discuss and resolve issues	Edward Yao

5	Nelda Garza	Scheduling Queries for Su Clinica Staff	Communication will involve inquiries on the availability of Su Clinica Staff	On as-needed basis	email	Response not required	Edward Yao
6	Joanne Benavides'	Scheduling Queries for Dr. Magoon	Communication will involve inquiries on the availability of Su Clinica Staff	On as-needed basis	email	Response not required	Edward Yao
7	Andrew Lombardo	Any technical issues, data-related questions, exchanged questions, and logistical resource allocations	Communication will be medium-level specific updates and progress of the project, focusing on staff and resource usage	Regular Weekly Meetings	email phone call/web-based meeting	Approvals and disapprovals of current list tasks, strategies, and results	Edward Yao
8	RGV HIE Team	Technical and data mapping issues within the exchange	Communication will be low-level and task-specific requirements agreed on by Andrew Lombardo	Regular Weekly Meetings	email phone call/web-based meeting	Updates and feedback on current strategies	Edward Yao
9	Clinical Entity IT Team	Technical issues within the clinic	Detailed requests and requirements for fulfilling tasks agreed to by Bert Gonzalez	On as-needed basis	email	Required response to confirm the issue and additional discussion times to resolve ambiguities	Edward Yao

10	Edward Yao	Questions regarding project accomplishments, objectives, and adverse events pertaining to the doctoral translational project	Provide translational project's past accomplishments, current progress, and future goals, as well as adverse events/learnings to stakeholders	On as-needed basis	email phone call/web-based meeting	Required response to pertinent stakeholders	Edward Yao

13 Risk Management

13.1 Risk Plan Overview

Identifying and addressing possible adverse events related to goals is a necessary step in project completion. The list of said adverse events will be documented with their associated risk of occurring, the impact on the project, and the mitigation measures. It will serve as the risk management documentation, which will be edited on an ongoing basis throughout the project's life.

13.2 Risk Identification

In evaluating the quality assessment of the hepatitis C discrepancies between systems, possible adverse events were identified and documented. All stakeholders will participate in the process of risk identification and contingencies in collaboration with the project manager, who will execute the specified protocol in response to the risks.

This section describes some of the key project risks and their potential impact on the success of the project. This list of risks should be regarded as provisional rather than complete because risks are usually identified and dealt with continuously during the course of the project.

Risk Identification

Risk	Possible impacts on the project
Data transfer issues	Lack of interoperability between systems can provide a significant amount of technical work to resolve that could set back project timelines and goals. The project would be reduced in scope, but deficiencies will be documented.
Clinical Entity priority change	Resources dedicated to the project may diminish, and access issues remain persistent. The project would not be able to be completed if it was not resolved.
Vendor priority change	A lack of vendor support could significantly hinder the amount of work directed toward resolving or addressing issues. As a result, technical issues and inaccessible data would hinder audits and, therefore, reduce the scope of the translational project, but deficiencies will be documented.
RGV HIE priority change	Lack of RGV HIE support could bring issues involving data access, authorization, and querying. The RGV HIE not being able to support the translational project would lead to the incompletion of the project.
Insufficient funding from the RGV HIE	Insufficient funding would detract from Vendor, consultant, and RGV HIE support. There would be no dedicated staff and work hours dedicated to resolving issues related to the RGV HIE environment. The overall project will be impacted.
Insufficient training	Insufficient training in technical knowledge of accessing data within available repositories would delay the entire methodology of the translational project. The overall project will be impacted and could hinder completion.

13.3 Risk Analysis

The likelihood of occurrence is also tracked, along with the impact and risk rating. The table will be used to determine and contextualize follow-up responses to said risks.

Risk Table

Risk	Probability	Impact	Overall Risk Rating	Type
Data transfer issues	Moderate	Major	Medium Risk	Technical Risk
Clinical Entity priority change	Unlikely	Major	Low Risk	Management Risk
Vendor priority change	Moderate	Major	High Risk	Commercial Risk
RGV HIE priority change	Unlikely	Extreme	Low Risk	Management Risk
Insufficient funding from the RGV HIE	Rare	Extreme	Low Risk	External / Management Risk
Insufficient training	Likely	Extreme	High Risk	Management Risk

13.4 Risk Mitigation

The likelihood of occurrence is also tracked, along with the impact and risk rating. The table will be used to determine and contextualize follow-up responses to said risks.

Risk Mitigation Table

Risk	Response
Data transfer issues	Mitigation: The project manager will work closely with all stakeholders to characterize and solve issues. The project itself will not be dependent on successful data transfers.
Clinical Entity priority change	Mitigation: The project manager's role within Su Clinica is tied to multiple revenue-generating projects. Duplicate role involvement will ensure that changes in Su Clinica priority will be less likely to effect authorization and access.
Vendor priority change	Mitigation: The project manger will work with multiple data repositories to mitigate access failures. Vendor failure will also be documented, and compliance is not necessary for project completion.
RGV HIE priority change	Mitigation: The project manager will work with multiple data repositories to mitigate access failures. Vendor failure will also be documented, and compliance is not necessary for project completion.
Insufficient funding from the RGV HIE	Mitigation: The project manager's role within the RGV HIE is tied to multiple revenue-generating projects. Duplicate role involvement will ensure that changes in RGV HIE priority will be less likely to affect authorization and access.
Insufficient training	Mitigation: FHIR and SQL training will be undergone by the program manager before the initiation of the project. In addition, available FHIR and SQL consultants will be made available and part of the translational project team.

14 Plan Project Procurement

The standard procurement process within the organization that involves the Executive Director of the RGV HIE creates a statement of work (SOW) to be presented to the RGV HIE Board of Directors. Once it is approved by the board, the Request for Proposals (RFPs) can be submitted to pertinent vendors in current use and to those seeking offers in the public market. Upon the arrival of the predetermined RFP deadline, proposals will be reviewed. The RGV HIE team will then submit bidding documentation to receive cost proposals, which are examined for quality and a cost evaluation. Accommodations and agreements are formalized in the contract

between the vendor and the organization to finalize the procurement. Plan project procurement will not be used as the purpose of the project is to assess the RGV HIE in its current state.

15 Conduct Procurement

The organization currently has all the pertinent vendor relationships pertinent to the translational project; in addition, all services and staffing are available. As such, the procurement process is not necessary. The FHIR and SQL vendor servers and consultants have been previously contracted, and this translational project will not need additional amendments, resources, or obligations. The procurement process for vendors, for human resources, all staff and contractors have been procured for this translational period. Therefore, no additional recruitment, hiring, or time allocation is needed.

16 Control Procurement

The procurement process within the organization follows a standard procedure. If, by any chance, procurement becomes necessary for this project, it would be essential to discuss and review the process with the HIE Executive Director. The Executive Director would then seek approval from the board for any procurement activities. However, it is crucial to emphasize that in the event of any procurement needs arising, the project cannot proceed immediately due to the absence of a budget allocated for this project. Therefore, careful consideration and financial planning would be required before initiating any procurement activities.

17 Implementation and Deployment Strategy

The DataGauge implementation strategy involves a structured approach to enhance data quality and reporting. Each iteration of the project is presented as rows within a result table, accompanied by a data diagram created using PlantUML. The framework allows for classifications and mapping of organizational steps to iterate over data quality and reporting. The

implementation process includes steps such as practice question analysis, data needs model development, establishing data quality requirements, and measuring data quality against those requirements. In the practice question analysis step, users analyze the pertinent question in terms of scope and purpose. In the data needs model development step, users make initial presumptions on what data they might need to answer the pertinent question. Quality assurance activities are placed in the final step to ensure quality within the production process. Collaboration among stakeholders is facilitated through the establishment of data standards, which define acceptable ranges, values, and system responses for future actions. The implementation of the DataGauge framework for hepatitis C within the RGV HIE ecosystem will occur in phases. The deployment will take place within the Su Clinica clinical system, which is connected to the RGV. The deployment strategy includes querying relevant data silos, manually auditing discrepancies, iterating and testing with the RGV HIE team, documenting and analyzing workflows, and conducting an ROI analysis to estimate revenue generation relative to implementation costs.

17.1 Product Verification Testing.

The product verification testing phase will be used to verify the empirical impact of the changes conducted by the project manager, RGV HIE, and Clinical IT Team.

Product Testing Team

Name	Function	Contact
Edward Yao	Project Manager	956-466-7131
Andrew Lombardo	Project Sponsor	956-622-5801
RGV HIE Team	Functional Testing and Output/End User Testing	956-622-5801
Clinical Entity IT Team	Functional Testing	(956) 365-6000

Goal. To reduce the discrepancies in hepatitis C diagnosis counts within the RGV HIE ecosystem.

General Guidelines.

1. The queries and methodology of each instance of data extraction must be documented in order to recreate and verify empirical results.
2. All issues should be reported to Andrew Lombardo and the RGV HIE team in order to manage pertinent stakeholders (Vendors and Clinical Partners) to resolve said issues.
3. Andrew Lombardo and the RGV HIE Team will provide final approval for the workflow readiness and results of the hepatitis C improvement project.

17.2 Final Actions.

- January 30th, 2023: RGV HIE announces that authorization and access have been granted.

- March 1st, 2023: RGV HIE announces that workflow and ROI analysis of data governance has been created and will be available for consideration by the RGV HIE Board.

18 Return on Investment (ROI)

18.1 Quantitative ROI

A return on investment (ROI) calculation compares the difference between the amount invested and the amount gained divided by the expected return (Cantor, 2011).

ROI Equation

$$ROI = \frac{Gain\ from\ Investement - Cost\ of\ Investment}{Cost\ of\ Investment}$$

Even though this project does not require additional investment, I will assume that resources are diverted to this project as an investment. These resources can be quantified as the number of expected hours to be given by staff per their dollar rate per hour. Salaried employees will be assumed to be paid hourly for the purposes of this project. Here are the costs associated with the team I will work with for one hour per week for consultation:

Estimated Cost Breakdown

Data Management Team	No. of Required Build Hours per Year	Rate per Hour	Total
Amazon Web Services	50	$150.00	**$7,500.00**
FHIR Consultant	50	$75.00	**$3,750.00**
HIE Integration Analyst	50	$34.31	**$1,715.50**
HIE Data Analyst	50	$26.44	**$1,322.00**
Total	200	**$285.75**	**$14,287.50**

With a total cost of $285.75 per hour, I can assume that if my project takes 32 weeks to complete, the initial investment will cost $9,144.00 ($285.75 * 32). If the grant received to

complete the project is $10,505, we have an ROI of 14% (0.14 *100) for the period of 32 weeks. Below displays the ROI with these pertinent values calculated.

ROI Applied

$$ROI = \frac{10{,}505.00 - 9{,}144.00}{9{,}144.00} = 0.14$$

There are no indirect costs, as there are no additional costs or divergence of resources for electricity, overhead, rent, or utilities. The only instance of resources diverging and, therefore, affecting resource allocation is labor in the form of one-hour meetings per week during the duration of the project..

18.2 Qualitative ROI

18.2.1 Workflows

A proposed quantitative measure is a workflow analysis detailing the steps that will be taken to complete the project. The steps will be represented by rectangular denominators with an appropriate title, from which they will be connected to other steps by arrows to indicate the next step. The workflow arrows represent the sequential flow of actions and steps taken with these title linkages (Tilley, 2019). These actions and steps will be represented through a data diagram, which categorizes said actions and steps with designated horizontal lines that will hold the pertinent categorization of titles (Diaz-Garelli et al., 2019; Tilley, 2019). I will use the server categorizations to represent servers within the Rio Grande Valley Health Information Exchange (RGV HIE), the clinical entity (Su Clinica). Therefore, the proposed data diagrams will describe the sequential steps and actions relative to the responsible departments capable of performing these steps and actions. Finally, a key feature of the data diagrams is the indication of data storage and retrieval (Braunstein, 2018; Tilley, 2019). This feature will be useful in identifying

redundancies and possible technical and labor constraints for the workflow conducted within the project. Data diagrams offer an opportunity to explicitly document the processes undertaken by an organization (Diaz-Garelli et al., 2019; Tilley, 2019; Wilson et al., 2014).

The explicit documentation serves as a visual aid for collaboration and discussion to improve processes and discuss issues. Processes can start in one department and end in another. Because of this, a workflow analysis can bring together a multi-departmental team to observe relationships between different groups and identify points for improvement. (Tilley, 2019). It is important to collaborate specifically with the staff performing the work to avoid miscommunication between outputs and intentions from management. A process may be planned, but the amount of variance can change the process over time, leading to unintended consequences and process outputs. A workflow analysis allows managers and staff to identify the actual steps undertaken and how to reduce variance to achieve a reliable outcome (Redman, 2021). In addition to aiding in collaboration, the creation of workflow analysis may allow for explicit role assignments and clarity on responsibilities. Staff and managers will be able to see which departments are affected by which actions (Redman, 2021). If there are further changes to the processes, an understanding of the old workflow analysis may aid in change adaptation by setting a precedent for future workflow discussions (Tilley, 2019).

18.2.2 Data Quality Measurement

Data quality measurement is a necessary step prior to using the data for analytical purposes. Even though data quality may be a quantitative measurement, there are qualitative reasons for the returns on investments that the RGV HIE will receive from the proposed project. All such aspects improve upon the current contextualization of success and failure as defined by the stakeholders. Successes cannot be acknowledged and problems cannot be solved if they are

not defined and measured (Chen et al., 2014 & Cowie et al., 2016). Data may not be acceptable for analysis by stakeholders in its current state. As such, defining the data quality variable and understanding the initial measurements are necessary to determine if the data can be used. In addition, if the data is improved, it must be determined at what point it will be deemed worthy of analysis. Since an analysis may involve contextualizing the data in a manner different from its initial use, this secondary use of the data must be carefully considered to understand if it can be used in other settings (Chen et al., 2014 & Cowie et al., 2016).

Data transferred from one database to another may undergo changes such that the same data is not comparable between databases. Data consistency—the proposed measurement for this project—between databases is important for stakeholders to understand, as data variance between sources may influence their confidence in data extraction and analysis. In addition, understanding the consistency between distributed and linked databases will also inform stakeholders' desire to improve communication between databases in the future (Yoshihashi & Hoyt, 2017). Having a systematized process for measuring data will allow the measurement to be repeatable over time. Having key measurements stored longitudinally will aid in identifying when problems occurred and what measurements were associated with said problems (Chen et al., 2014 & Cowie et al., 2016). Understanding the initial state informs the quality of future states of the system. It is not only important to identify to what magnitude a problem exists but also how it compares to the previous states of the system. (Chen et al., 2014 & Cowie et al., 2016)

Appendix C: DataGauge Results Table

Table 7

DataGauge Results Table

		1- Define the question	2- Define the data required to answer the question	3- Define the data standards needed to validate the data	4- Define the logistical steps to extract the data	5- Evaluate the results and iterate if required
Iteration [10/31/2022–1/20/2023]	**1**	Can any encounter data be pulled from the FHIR server? Presumptions: 1. FHIR server is in the RGV HIE. 2. Encounter should have lab data in the FHIR server.	In the FHIR server, placeholder data points to test query capabilities of the FHIR server: 1. LOINC Code: 72376-7 2. ICD Code: B182 Other codes will be searched in future iterations.	In the FHIR server: 1. must return valid response (no error codes) 2. must be under procedure or diagnosis in the encounter resource	Queries attempted in the FHIR server: REST API Query 1: GET /fhir/Patient?_has:Encounter:procedure-code=http://loinc.org|72376-7* REST API Query 2: GET /fhir/Patient?_has:Encounter:diagnosis-code=http://hl7.org/fhir/sid/icd-10|B182*	**Unsuccessful** **Queries 1 and 2:** "Error Code: 1064. You have an error in your SQL syntax'resourceType: OperationOutcome" issue: - severity: error code: invalid diagnostics: No search parameter for Encounter.procedure-code **Auditor, RGV HIE IT, and FHIR vendor discussion:** When attempting to execute queries on the FHIR server, an error code was generated. Upon further examination, it was discovered that the queries were deemed invalid due to the lack of support for FHIR search parameters by the FHIR vendor. The vendor in question implemented their own proprietary search parameters, rendering the

						queries ineffective. Additionally, it was determined that even if the queries were valid, the user authorizations in place would have prevented their execution. Furthermore, it is advisable to utilize the observation resource in lieu of the encounter resource for future lab-related tasks, as the former is a more established and developed resource in the FHIR framework. **Next steps:** Due to the unfamiliarity with the FHIR server, the RGV HIE team recommended that the SQL server was selected as the preferred option. Although the FHIR server was presented as a potential alternative, further investigation and training on its capabilities are recommended for future projects.

Iteration [1/24/2023–1/26/2023]	**2**	Using the SQL server, can any Su Clinica patients be pulled?	Found location IDs in documentation in SQL server:
			• Facility IDs: 12, 15, and 18
			  ○ 12 = Su Clinica Familiar Brownsville-Adult
			  ○ 15 = Su Clinica Familiar Harlingen-Adult
			  ○ 18 = Su Clinica Familiar
			• Optional: Any data point relating to a lab or procedure
				In SQL server:
				1. must return valid response (no error codes)
				2. must have facility ID field
				3. must have date visit or creation field
					Queries attempted in SQL server:
					SQL Query 1:
					SELECT * FROM data_warehouse.enconter
					WHERE
					(facility_id LIKE '12' or
					facility_id LIKE '15' or
					facility_id LIKE '18') AND
					date between '2022-10-01' and '2022-12-31'
					SQL Query 2:
						Successful
						
						Queries 1 and 2:
						
						“Error Code: 1054. Unknown column ‘facitily_id’ in ‘where clause’”
						
						Auditor discussion
						
						Tables for “patient” and “encounter” did not have data fields, facility fields, or any data point relating to a lab or procedure.
						
						Query 3:

			• Required: Any patient identifier		SELECT * FROM data_warehouse.patient WHERE (facility_id LIKE '12' or facility_id LIKE '15' or facility_id LIKE '18') AND date between '2022-10-01' and '2022-12-31' SQL Query 3: SELECT * FROM data_warehouse.problem WHERE (facility_id LIKE '12' or facility_id LIKE '15' or facility_id LIKE '18') AND date between '2022-10-01' and '2022-12-31'	"2,580 row(s) returned" **Auditor discussion:** The problem table was able to yield a valid result with all three standards: data fields, facility fields, or any data point relating to a lab or procedure. The valid result proved that the facility IDs are valid for the SQL server. A result was returned with no hepatitis C tests when conducting text value inspection on the "value_name" and "classification" columns. While inspecting the SQL documentation, I noticed another table that could potentially hold the lab information: the procedure table. **Next steps:** An examination of the procedure table is recommended to identify any relevant tests or procedures that may have been performed.
Iteration [1/27/2023]	**3**	Among the patients served by the SQL server, is it possible to fetch a patient from Su Clinica from October 1, 2022 to December 31, 2022 with hepatitis C?	Found location IDs in documentation in SQL server: • Facility IDs: 12, 15, and 18 • 12 = Su Clinica Familiar Brownsville-Adult • 15 = Su Clinica Familiar Harlingen-Adult • 18 = Su Clinica Familiar	In SQL server: 1. must return data 2. must have facility ID field 3. must have date visit or creation field 4. must have hepatitis C data	Queries attempted in SQL server: Query: SELECT * FROM data_warehouse.procedure WHERE (facility_id LIKE '12' or facility_id LIKE '15' or facility_id LIKE '18') AND date_start between "2022-10-01" and "2022-12-31"	**Unsuccessful** **Query:** 9,914 row(s) returned **Auditor discussion:** A result was returned but with no hepatitis C tests when conducting text value inspection on the "value_name" and "classification" columns. **Next steps:**

			• date_start between "2022-10-01" and "2022-12-31" • Any data point relating to a lab or procedure is needed			An inquiry will be made to Su Clinica to obtain a comprehensive list of relevant tests or procedures. Although access to the database is currently unavailable, the submission of an IT service request ticket is being pursued as a possible means of acquiring said list. Once acquired, the relevant IDs will be cross-referenced with the server to retrieve the necessary data. It is postulated that cross-referencing IDs will prove more efficient than searching through procedure and encounter records.
Iteration [1/30/2023–2/6/2023]	**4**	Is it possible to obtain any patients in Su Clinica between October 1, 2022 and December 31, 2022 who have a hepatitis C diagnosis within the Su Clinica's EHR?	In Su Clinica: • ID • hepatitis C test name • test date • RGV HIE status	In Su Clinica: 1. must be a Su Clinica patient 2. must have a hepatitis C test 3. test must be taken between the relevant dates 4. must have RGV HIE status	Requested data from Su Clinica IT through IT ticket	**Successful** **Query:** 85 in total, but only 30 consented to be part of the RGV HIE. **Auditor evaluation:** Upon consultation with the IT department at Su Clinica, it was determined that laboratory results are not typically stored within the clinic's database. Instead, the attending physician signs them before sending them to the relevant parties. Additionally, a set of codes used by Su Clinica was obtained for future reference in data retrieval and analysis.

						CPT: 86803 -LOINC: 48159-8, 48159-8, 13955-0, 11011-4, 38180-6, 19147-8, 62365-2 **Next steps:** A workflow diagram was developed to assist in the identification and resolution of issues pertaining to missing laboratory results. Approval from Su Clinica was obtained regarding the problem description and proposed approach. In addition to manual inspection of text values, targeted searches of specific codes will be conducted within the relevant servers to confirm the hypothesis that laboratory reports may not be successfully transmitted.
Iteration [2/7/2023–2/27/2023]	**5**	How many hepatitis C tests do the 30 identified Su Clinica patients have within the RGV HIE environment?	In Su Clinica, FHIR, SQL, Forcare, and Diameter, the audit will match the IDs provided by Su Clinica: • ID • test name • test date In Verato: • EMPI must be present	Must be present and have a hepatitis C test present in the results of the queries	The process involves retrieving a unique identifier from Su Clinica, using said identifier to procure the EMPI from the FHIR server, and subsequently utilizing the EMPI to retrieve the necessary data from the SQL, Diameter, and Forcare databases. Query 1: Forcare user interface: Enter Su Clinica ID in ID section	**Unsuccessful** **Query 3** 30 Su Clinica IDs returned 30 EMPIs. **Query 4** 30 Su Clinica IDs returned 30 FHIR server IDs. **Queries 1, 2, 5, 6, and 7** No hepatitis C lab results were found.

	Query 2: Verato user interface: Enter Su Clinica ID in link ID section Query 3: Diameter user interface: Enter Su Clinica ID in MRN extension section Query 4: REST API query to FHIR ID: Get /fhir/Patient?identifier=[SuClinicaID] Query 5: REST API query to obtain labs: Get /fhir/Patient?subject=[FHIR Server ID] Query 6: SQL server queries for patient problems: SELECT * FROM data_warehouse.patient JOIN data_warehouse.problem ON data_warehouse.patient.id = data_warehouse.problem.patient_id WHERE (empi LIKE 'EMPI') Query 7:	**Auditor and RGV HIE discussion:** Analysis of the observed data has led to the conclusion that laboratory reports are only transmitted to the RGV HIE when patients have provided explicit consent and attending physicians have provided approval. Although the clinic's database contains records of all patients who have consented to participate in the RGV HIE, laboratory reports are not consistently found within said records. Further investigation is necessary to determine the root cause of this inconsistency. **Next steps:** The current analysis indicates that the transmission of laboratory reports to the RGV HIE is contingent upon both patient consent and physician approval. Despite the availability of records for consenting patients, laboratory reports are not consistently present within the database, indicating a potential breakdown in the current architecture. A re-evaluation of the architecture is recommended to address this issue and reconcile the need for physician sign-off with the requirement for data accessibility by the RGV HIE. The identification of a scalable and secure solution to this

	SQL server queries for lab results: SELECT * FROM data_warehouse.patient JOIN data_warehouse.result ON data_warehouse.patient.id = data_warehouse.result.patient_id WHERE (empi LIKE 'EMPI')	conundrum is crucial for the continued effectiveness of the RGV HIE and its partners.

Appendix D: UML Code for Iterations

Figure 20

PlantUML Example 1

```
@startuml

box "Environment 1"
database General_Server_1
end box

box "Environment 2"
database General_Server_2
end box

autonumber
General_Server_1 --> General_Server_2: Message 1
General_Server_1 --> General_Server_2: Message 2

@enduml
```

Note: Servers are represented by the “database” variable, and messages between the servers are represented by the “-->” syntax between the servers. The “autonumber” input will place the number in the upper left section of the arrow in the graphical output, which represents the order in which events occurred (version 1.2021.2; GitHub, 2021).

Figure 21

PlantUML Example 2

```
@startuml

box "Environment 1"
database General_Server_1
end box

box "Environment 2"
database General_Server_2
end box

== Operation Classification 1 ==
autonumber
General_Server_1 --> General_Server_2: Message 1

== Operation Classification 2 ==
General_Server_1 --> General_Server_2: Message 2

@enduml
```

Note: The sections can represent different messages (version 1.2021.2; GitHub, 2021). In this example, “=Operation Classification 1=” for Message 1 and “=Operation Classification 2=” for Message 2 are used to represent different sections.

Figure 22

PlantUML Example 3

```
@startuml

box "Environment 1"
database General_Server_1
end box

box "Environment 2"
database General_Server_2
end box

actor Auditor

== Operation Classification 1 ==
autonumber
General_Server_1 --> General_Server_2: Message 1

== Operation Classification 2 ==
General_Server_1 --> General_Server_2: Message 2
Auditor --> General_Server_2: Message 3 with missing information
destroy General_Server_2

@enduml
```

Note: An auditor serves as an actor with the “actor” variable, represented by a human figure in the graphical output who can also act on the servers with their own messages (Version 1.2021.2; GitHub, 2021). For the purposes of this paper, the syntax “destroy” will have a graphical output of X, which is denoted at the end of the arrow to indicate that the message was sent but with no information present.

Figure 23

PlantUML Example 4

```
@startuml

box "Environment 1"
database General_Server_1
end box

box "Environment 2"
database General_Server_2
end box

actor Auditor

== Operation Classification 1 ==
autonumber
General_Server_1 --> General_Server_2: Message 1

== Operation Classification 2 ==
General_Server_1 --> General_Server_2: Message 2
Auditor --> General_Server_2: Message 3 with missing information
destroy General_Server_2

autonumber 3
alt If this action happens
        General_Server_1 --> General_Server_2: Then this happens
autonumber 3
else Else, if the action did not happen
        General_Server_1 --> General_Server_2: Then this does not happen
        destroy General_Server_1
        destroy General_Server_2
end


@enduml
```

Note: “If-else” boxes were placed to denote that if an event did occur, a message would be sent; if the event did not occur, then a message would not be sent (Version 1.2021.2; GitHub, 2021).

Figure 24

Initial Presumption Diagram

```
@startuml

autonumber

box "Clinic_Evnironment"
database Su_Clinica
end box

box "RGV_HIE_Environment"
database Forecare
database Verato
database Diameter
database FHIR_Server
end box

== Presumption of RGV HIE Operations ==

Su_Clinica --> Forecare: CCDA

Forecare --> Verato: Name, \nDOB, \nAddress, \nSSN

Verato --> Forecare: EMPI

Verato --> Diameter: EMPI

Forecare --> Diameter: CCDA

Diameter --> FHIR_Server: General \nClinical \nInformation


@enduml
```

Figure 25

Final Presumption Diagram

```
@startuml


box "Clinic_Evnironment"
database Su_Clinica
end box

box "RGV_HIE_Environment"
database Forecare
database Verato
database Diameter
database FHIR_Server
database SQL_Server

end box


== Presumption of RGV HIE Operations ==
autonumber

Su_Clinica -> Su_Clinica: Patient consents to RGV HIE

alt Physician meets patient and signs lab
        Su_Clinica --> Forecare: CCDA has Lab
autonumber 2
else Physician does not meet patient and does not sign lab
        Su_Clinica --> Forecare: CCDA with no Lab
        destroy Forecare
end

Forecare --> Verato: Name, \nDOB, \nAddress, \nSSN

Verato --> Forecare: EMPI

Verato --> Diameter: EMPI

Forecare --> Diameter: CCDA

Diameter --> FHIR_Server: General \nClinical \nInformation

Diameter --> SQL_Server: General \nClinical \nInformation

@enduml
```

Figure 26

Iteration 1 Code

```
@startuml

autonumber

box "Clinic_Evnironment"
database Su_Clinica
end box

box "RGV_HIE_Environment"
database Forecare
database Verato
database Diameter
database FHIR_Server
end box

actor Auditor

== Presumption of RGV HIE Operations ==

Su_Clinica --> Forecare: CCDA

Forecare --> Verato: Name, \nDOB, \nAddress, \nSSN

Verato --> Forecare: EMPI

Verato --> Diameter: EMPI

Forecare --> Diameter: CCDA

Diameter --> FHIR_Server: General \nClinical \nInformation

== Audit Operations ==

Auditor --> FHIR_Server : Data Request
note right of FHIR_Server: **REST API Query 1**:\nGET
[base]/Patient?_has:\nEncounter:procedure-code=\nhttp://loinc.org|72376-7*
note right of FHIR_Server: **REST API Query 2**:\nGET /fhir/Patient?_has:\nEncounter:diagnosis-
code=\nhttp://hl7.org/fhir/sid/icd-10|B182*

FHIR_Server --> Auditor : Data Response
note right of FHIR_Server: **REST API Response 1**:\ncode: invalid. diagnostics: \nNo search
parameter for Encounter.procedure-code
note right of FHIR_Server: **REST API Response 1**:\ncode: invalid. diagnostics: \nNo search
parameter for Encounter.procedure-code

@enduml
```

Figure 27

Iteration 2 Code

```
@startuml

autonumber

box "Clinic_Evnironment"
database Su_Clinica
end box

box "RGV_HIE_Environment"
database Forecare
database Verato
database Diameter
database FHIR_Server
database SQL_Server

end box

actor Auditor

== Presumption of RGV HIE Operations ==

Su_Clinica --> Forecare: CCDA

Forecare --> Verato: Name, \nDOB, \nAddress, \nSSN

Verato --> Forecare: EMPI

Verato --> Diameter: EMPI

Forecare --> Diameter: CCDA

Diameter --> FHIR_Server: General \nClinical \nInformation

Diameter --> SQL_Server: General \nClinical \nInformation

== Audit Operations ==

Auditor --> SQL_Server: Data Request
note right of SQL_Server: **SQL Query 1**:\nSELECT * FROM \ndata_warehouse.enconter
\nWHERE (facility_id LIKE '12' OR \nfacility_id LIKE '15' OR \nfacility_id LIKE '18') AND \ndate
between '2022-10-01' and '2022-12-31'

note right of SQL_Server: **SQL Query 2**:\nSELECT * FROM \ndata_warehouse.patient
\nWHERE \n(facility_id LIKE '12' OR \nfacility_id LIKE '15' OR \nfacility_id LIKE '18') AND \ndate
between '2022-10-01' and '2022-12-31'
```

```
note right of SQL_Server: **SQL Query 3**:\nSELECT * FROM \ndata_warehouse.problem
\nWHERE \n(facility_id LIKE '12' OR \nfacility_id LIKE '15' OR \nfacility_id LIKE '18') AND \ndate
between '2022-10-01' and '2022-12-31'

SQL_Server --> Auditor: SQL Response
note right of SQL_Server: **SQL Response 1**:\nError Code: 1054. \nUnknown column facitily_id in
where clause

note right of SQL_Server: **SQL Response 2**:\nError Code: 1054. \nUnknown column facitily_id in
where clause

note right of SQL_Server: **SQL Response 3**:\n2580 row(s) returned; \nbut no Hepatitis C Tests
when doing text \nvalue inspection on the value_name and \nclassification columns.

@enduml
```

Figure 28

Iteration 3 Code

```
@startuml

autonumber


box "Clinic_Evnironment"
database Su_Clinica
end box

box "RGV_HIE_Environment"
database Forecare
database Verato
database Diameter
database FHIR_Server
database SQL_Server
end box

actor Auditor

== Presumption of RGV HIE Operations ==

Su_Clinica --> Forecare: CCDA

Forecare --> Verato: Name, \nDOB, \nAddress, \nSSN

Verato --> Forecare: EMPI

Verato --> Diameter: EMPI

Forecare --> Diameter: CCDA

Diameter --> FHIR_Server: General \nClinical \nInformation

Diameter --> SQL_Server: General \nClinical \nInformation

== Audit Operations ==

Auditor --> SQL_Server: Data Request
note right of SQL_Server: **SQL Query**:\nSELECT * FROM \ndata_warehouse.procedure
\nWHERE (facility_id LIKE '12' OR \nfacility_id LIKE '15' OR \nfacility_id LIKE '18') AND \ndate
between '2022-10-01' and '2022-12-31'

SQL_Server --> Auditor: Data Response
note right of SQL_Server: **SQL Response**:\n9914 row(s) returned; \nNo Hepatitis C Tests found

@enduml
```

Figure 29

Iteration 4 Code

```
@startuml

box "Clinic_Evnironment"
database Su_Clinica
end box

box "RGV_HIE_Environment"
database Forecare
database Verato
database Diameter
database FHIR_Server
database SQL_Server
end box

actor Auditor

== Presumption of RGV HIE Operations ==
autonumber

Su_Clinica -> Su_Clinica: Patient consents to RGV HIE

alt Physician meets patient and signs lab
        Su_Clinica --> Forecare: CCDA has Lab
autonumber 2
else Physician does not meet patient and does not sign lab
        Su_Clinica --> Forecare: CCDA with no Lab
        destroy Forecare
end

Forecare --> Verato: Name, \nDOB, \nAddress, \nSSN

Verato --> Forecare: EMPI

Verato --> Diameter: EMPI

Forecare --> Diameter: CCDA

Diameter --> FHIR_Server: General \nClinical \nInformation

Diameter --> SQL_Server: General \nClinical \nInformation

== Audit Operations ==

Auditor --> Su_Clinica: Data Request
note right of Su_Clinica: **IT Ticket**:\nHow many patients exist in Su Clinica \nwith Hep C tests
between those dates who are \npart of the RGV HIE between '2022-10-01' and '2022-12-31'?
```

```
Su_Clinica --> Auditor: Data Response
note right of Su_Clinica: IT Ticket Response**:\n85 were in total but only 30 consented to be part of the RGV HIE. \nReceived Su Clinica IDs of the 30 patients.

@enduml
```

Figure 30

Iteration 5 Code

```
@startuml

box "Clinic_Evnironment"
database Su_Clinica
end box

box "RGV_HIE_Environment"
database Forecare
database Verato
database Diameter
database FHIR_Server
database SQL_Server


end box

actor Auditor

== Presumption of RGV HIE Operations ==
autonumber

Su_Clinica -> Su_Clinica: Patient consents to RGV HIE

alt Physician meets patient and signs lab
        Su_Clinica --> Forecare: CCDA has Lab
autonumber 2
else Physician does not meet patient and does not sign lab
        Su_Clinica --> Forecare: CCDA with no Lab
        destroy Forecare
end

Forecare --> Verato: Name, \nDOB, \nAddress, \nSSN

Verato --> Forecare: EMPI

Verato --> Diameter: EMPI

Forecare --> Diameter: CCDA

Diameter --> FHIR_Server: General \nClinical \nInformation

Diameter --> SQL_Server: General \nClinical \nInformation

== Audit Operations for Each of the 30 Su Clinica IDs ==

Auditor --> Forecare: Data Request
```

```
note right of Forecare: **Query 1-- User Interface**:\nEnter Su Clinica ID in ID section

Forecare --> Auditor: Data Response
note right of Forecare: **Response 1-- User Interface**:\n0 Hepatitis C Records

Auditor --> Verato: Data Request
note right of Verato: **Query 2-- User Interface**:\nEnter Su Clinica ID in Link ID section

Verato --> Auditor: Data Response
note right of Verato: **Response 2-- User Interface**:\n30 EMPIs

Auditor --> Diameter: Data Request
note right of Diameter: **Query 3-- User Interface**:\nEnter Su Clinica ID in MRN Extension section

Diameter --> Auditor: Data Response
note right of Diameter: **Response 3-- User Interface**:\n0 Hepatitis C Records

Auditor --> FHIR_Server: Data Request
note right of FHIR_Server: **Query 4-- REST API Query to get FHIR Server ID**:\nGet
/fhir/Patient?identifier=[EMPI]

FHIR_Server --> Auditor: Data Response
note right of FHIR_Server:**Response 4-- REST API Response for FHIR Server ID**:\nFHIR Server
ID

Auditor --> FHIR_Server: Data Request
note right of FHIR_Server:**Query 5-- REST API Query to get labs**:\nGet
/fhir/Observation?subject=[FHIR Server ID]

FHIR_Server --> Auditor: Data Response
note right of FHIR_Server:**Response 5-- REST API Response for labs**:\n0 Hepatitis C Records

Auditor --> SQL_Server: Data Request
note right of SQL_Server: **Query 6-- SQL Query**:\nSELECT * FROM data_warehouse.patient
\nJOIN data_warehouse.problem \nON data_warehouse.patient.id =
data_warehouse.problem.patient_id \nWHERE ([EMPI Number] LIKE 'EMPI')

SQL_Server --> Auditor: Data Response
note right of SQL_Server: **Response 6-- SQL Query**:\n0 Hepatitis C records

Auditor --> SQL_Server: Data Request
note right of SQL_Server: **Query 7-- SQL Query**:\nSELECT * FROM data_warehouse.patient
\nJOIN data_warehouse.result \nON data_warehouse.patient.id = data_warehouse.result.patient_id
\nWHERE ([EMPI Number] LIKE 'EMPI')

SQL_Server --> Auditor: Data Response
note right of SQL_Server: **Response 7-- SQL Query**:\n0 Hepatitis C records

@enduml
```

Figure 31

Recommended Sequence Diagram PlantUML

```
@startuml

box "Clinic_Evnironment"
database Su_Clinica
actor Analyst
end box

autonumber

box "RGV_HIE_Environment"
database General_Server
database Forecare
database Verato
database Diameter
database FHIR_Server
database SQL_Server
end box

== Recommendation ofof RGV HIE Operations ==
autonumber 1
Analyst --> Su_Clinica: Retrieves all data, regardless \nof RGV HIE consent status
autonumber 2
alt Su Clinica approves transfer
        Analyst --> General_Server: Give deidentified and/or aggregated data
autonumber 2
else Su Clinica does not approve transfer
        Analyst --> General_Server: Give deidentified and/or aggregated data
        destroy General_Server
        destroy Analyst
end

autonumber 3

Su_Clinica -> Su_Clinica: Patient consents to RGV HIE

alt Physician meets patient and signs lab
        Su_Clinica --> Forecare: CCDA has Lab
autonumber 4
else Physician does not meet patient and does not sign lab
        Su_Clinica --> Forecare: CCDA with no Lab
        destroy Forecare
end
```

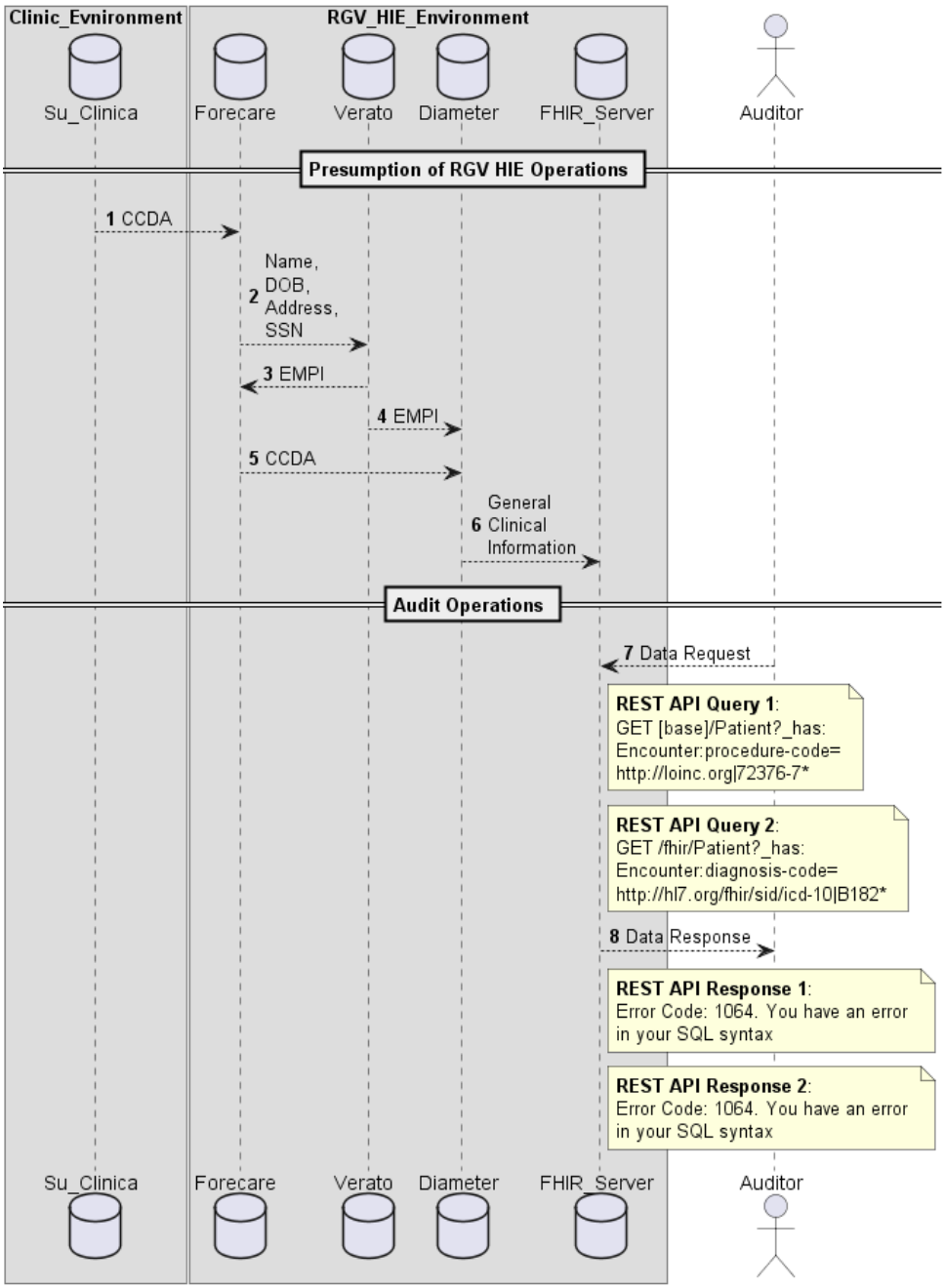
```
Forecare --> Verato: Name, \nDOB, \nAddress, \nSSN

Verato --> Forecare: EMPI

Verato --> Diameter: EMPI

Forecare --> Diameter: CCDA

Diameter --> FHIR_Server: General \nClinical \nInformation

Diameter --> SQL_Server: General \nClinical \nInformation

@enduml
```

Appendix E: UML Diagrams

Iteration 1

Figure 32

Iteration 1 Diagram



Note: The initial iteration of RGV HIE operations involves sending a CCDA from Su Clinica in the clinic environment to Forcare in the RGV HIE environment, which then sends relevant patient variables (such as name, date of birth, address, and social security number) to Verato. Verato subsequently transmits the information to the EMPI, which is used by both Forcare and Diameter. Forcare pushes the CCDA to Diameter, and Diameter parses the CCDA data with respect to the known EMPI and transmits it to the FHIR server. During the audit, two REST API queries were executed, yielding identical responses:

“resourceType: OperationOutcome

issue:

- severity: error

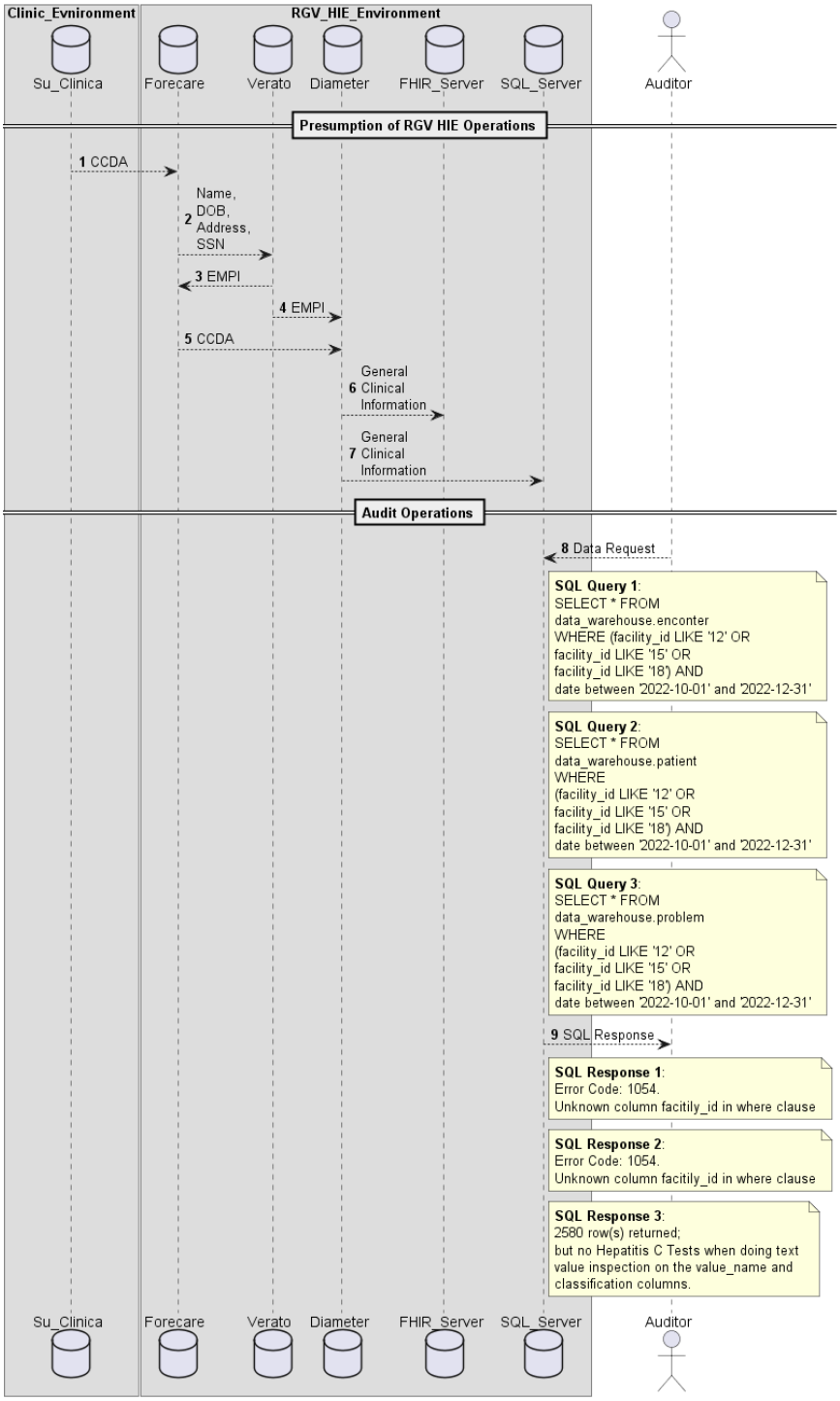
  code: invalid

  diagnostics: No search parameter for Encounter.procedure-code”

Iteration 2

Figure 33

Iteration 2 Diagram

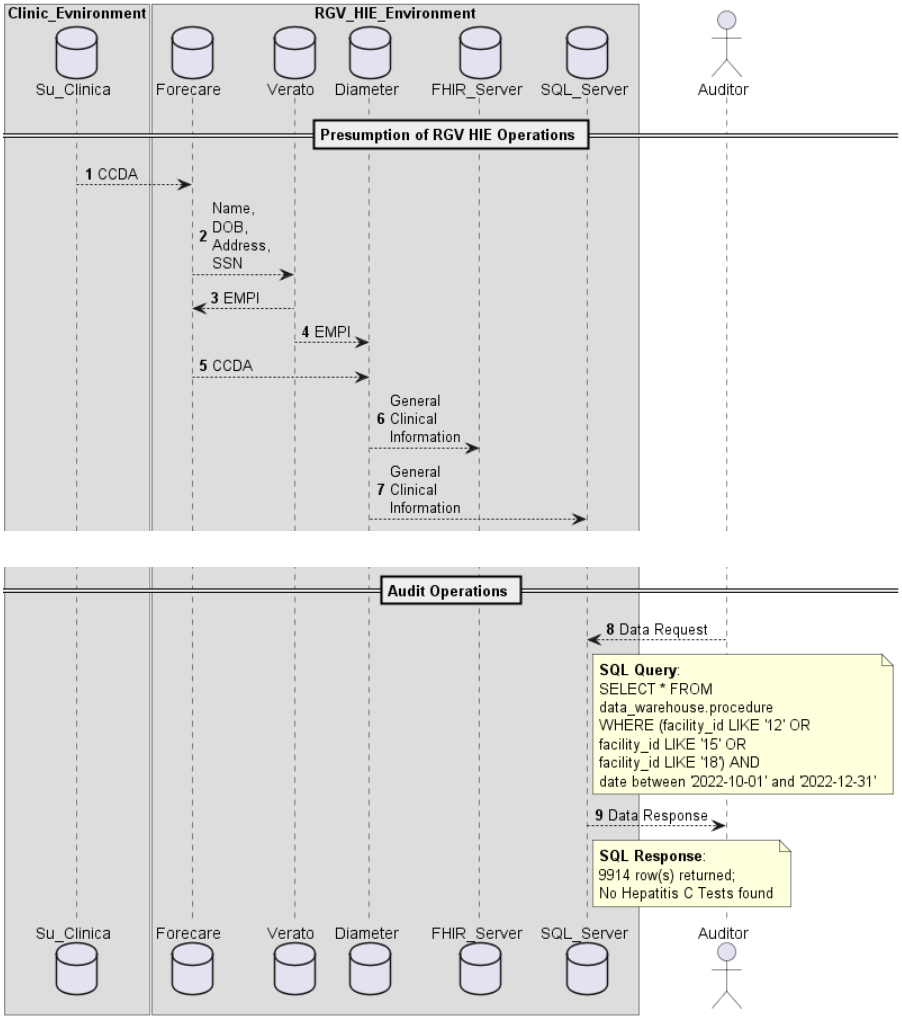


Note: In the second iteration, an SQL server was discovered and integrated into the existing sequence of server communications. On the FHIR server, Diameter sent the parsed general clinical information to the SQL server. As part of the audit, three SQL queries were submitted: SQL Query 1, SQL Query 2, and SQL Query 3. The resulting SQL responses were labeled SQL Response 1, SQL Response 2, and SQL Response 3.

Iteration 3

Figure 34

Iteration 3 Diagram

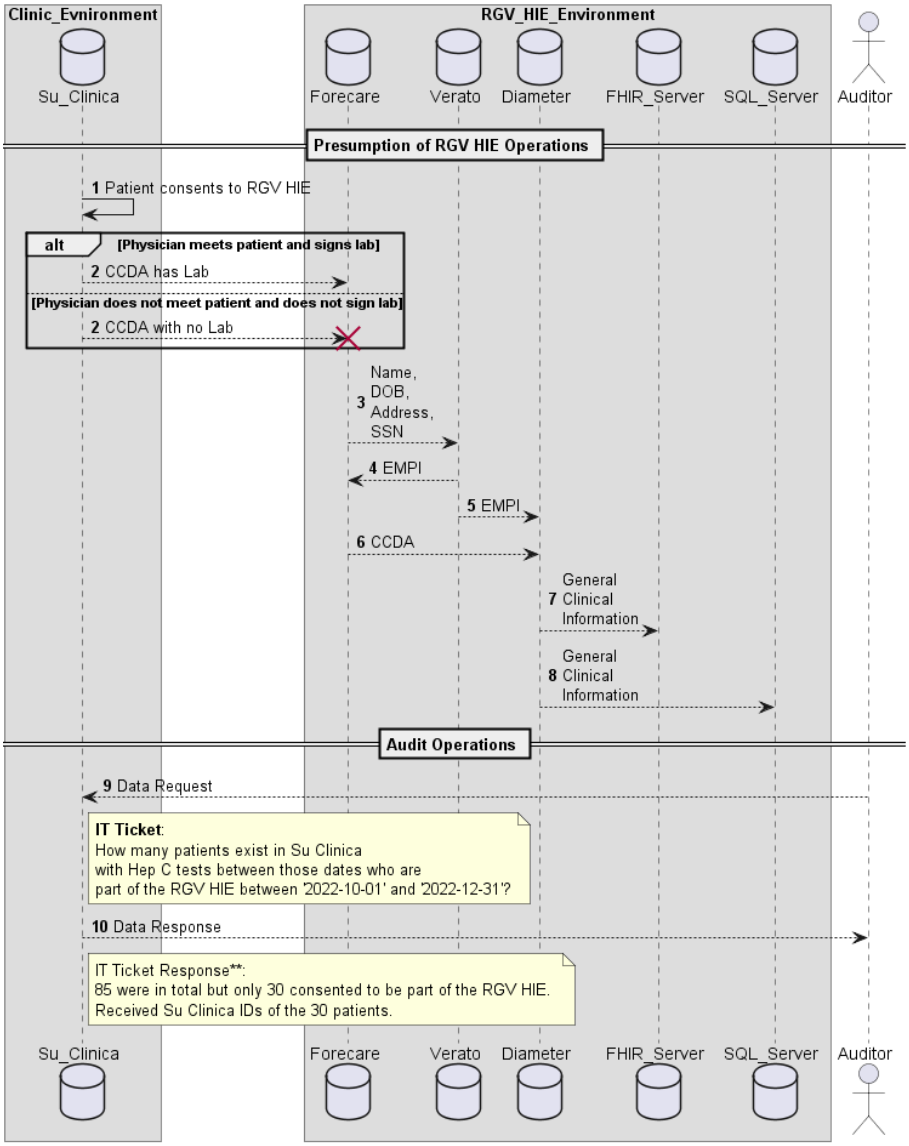


Note: The operations of RGV HIE in iteration 3 maintained the same presumption as the previous iteration; however, there is a distinction in that the audit now submitted a distinct SQL query and received a varied SQL response.

Iteration 4

Figure 35

Iteration 4 Diagram



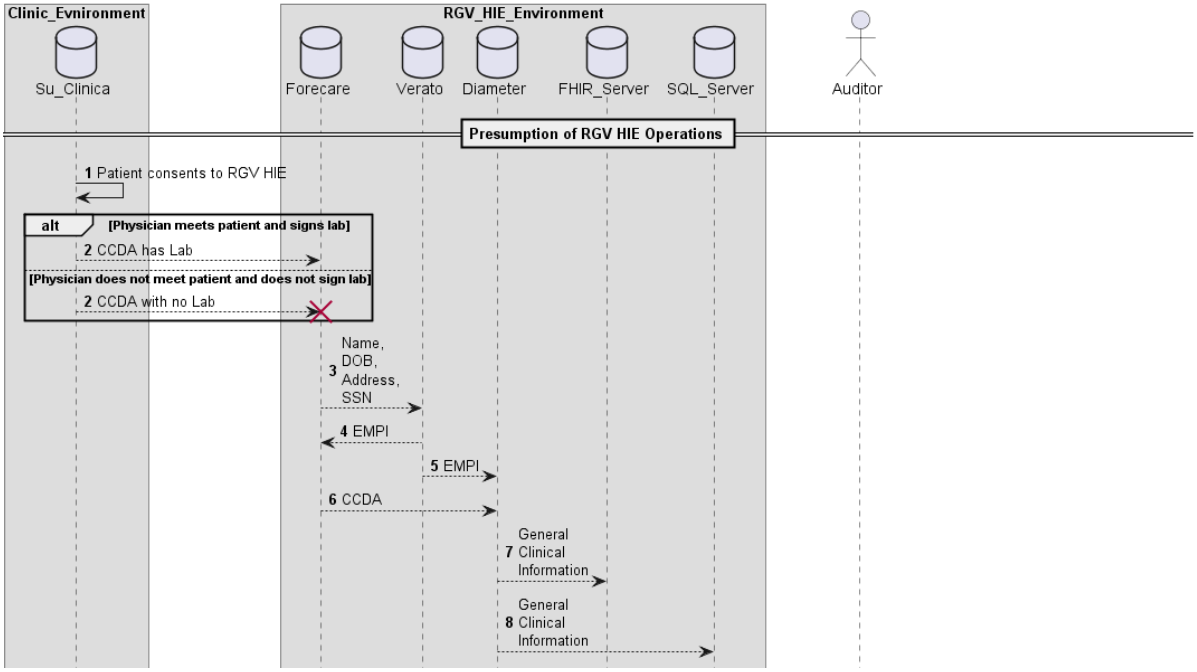
Note: In the fourth iteration, the RGV HIE operations underwent a significant change after consulting with Su Clinica IT. The updated operations now specify that the RGV HIE process commences only after the patient has granted consent to transmit their protected health information (PHI) to the RGV HIE. Additionally, the audit results revealed that labs are included

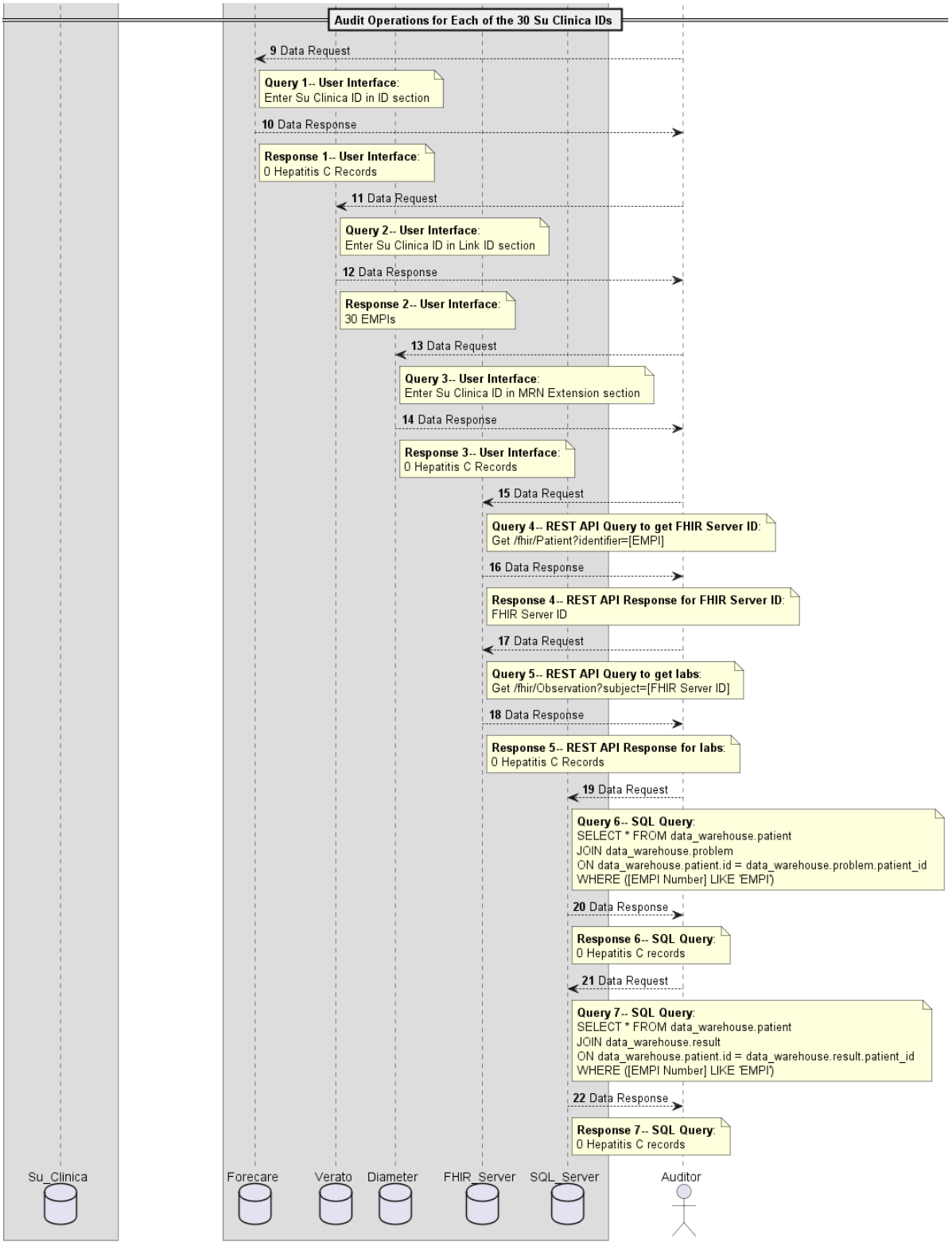
in the CCDA report only if the physician has met with the patient and approved the labs; if the physician does not approve the labs, they are not included in the CCDA report. All other operations remain unchanged. The audit called for the submission of an IT ticket, which yielded a response indicating 85 participants. Nonetheless, the audit ultimately received only 30 IDs, as only those participants had given their consent to the RGV HIE.

Iteration 5

Figure 36

Iteration 5 Diagram





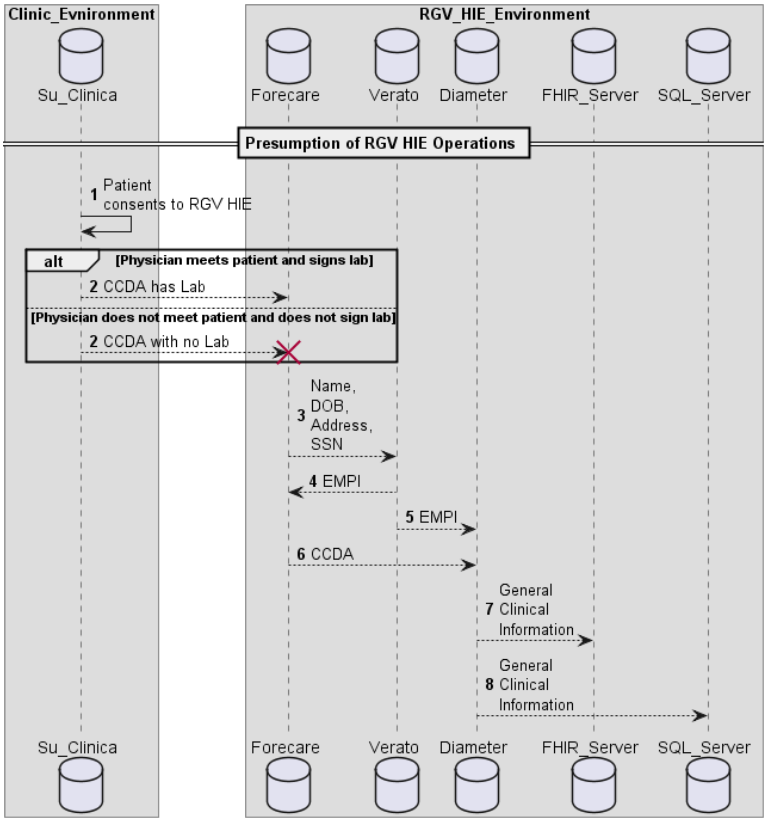
Note: In the fifth iteration, the RGV HIE's presumption remained unchanged from iteration 4. During the audit operations, the Su Clinica IDs were inputted into Forcare, but no records of

hepatitis C were found. Subsequently, the same IDs were submitted to Verato and obtained 30 EMPIs. Upon entering the Su Clinica IDs into Diameter, no hepatitis C results were found. The EMPIs were then utilized to query the observation resource on the FHIR server, but they still found no hepatitis C results. Lastly, the audit referenced the patient EMPI relative to both the problem and result tables using the JOIN statement in the SQL database but found no records of hepatitis C.

As-Is Diagram

Figure 37

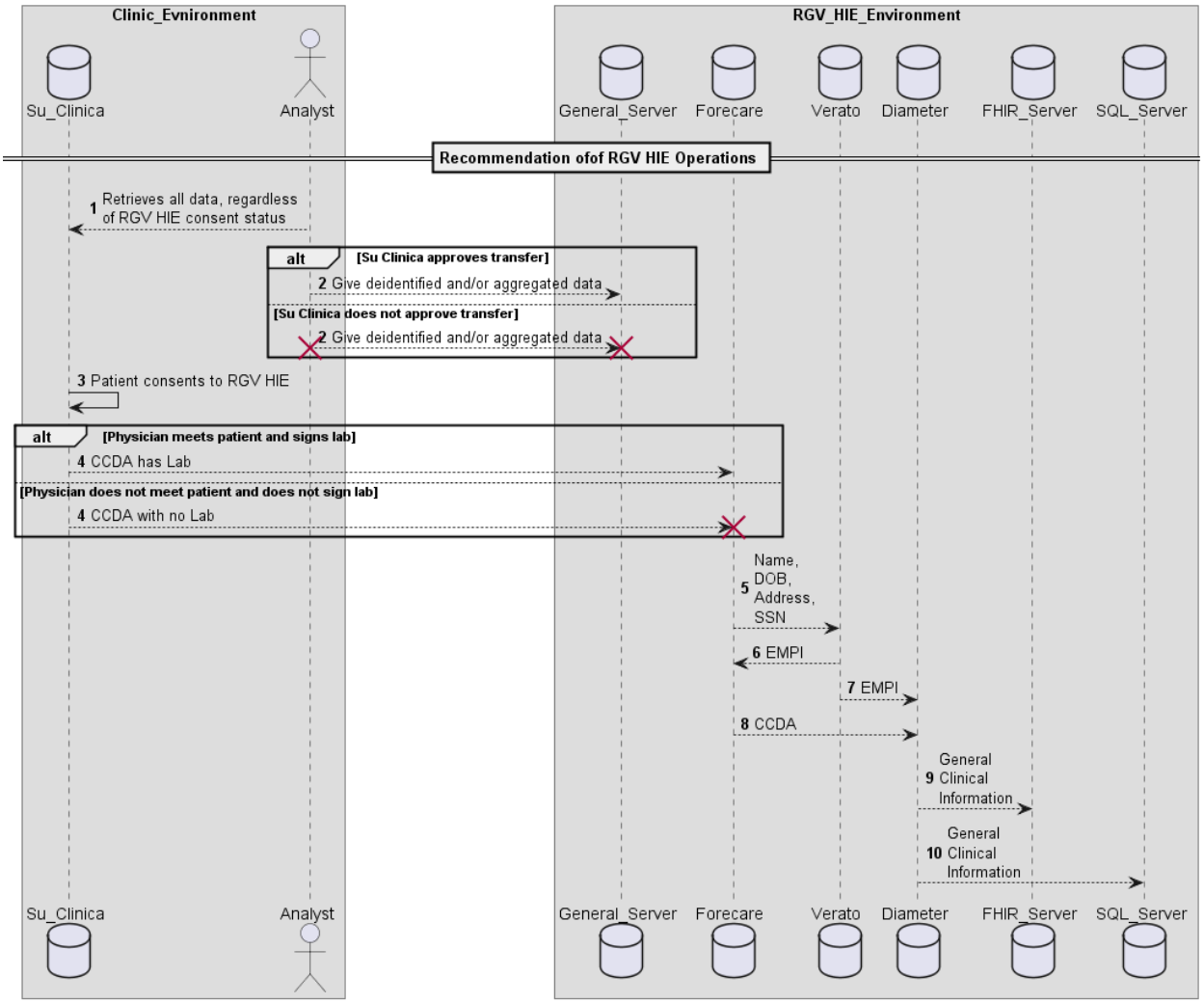
As-Is Diagram



Note: The current operations between the RGV HIE and Su Clinica are depicted in the above diagram. It begins with the patient consenting to share their clinical information with the RGV HIE environment. Once consent is given, Su Clinica sends CCDA documents to the RGV HIE environment. Finally, the reason why hepatitis C labs were missing was because the CCDA only includes a lab report if the physician who treated the patient authorizes it; if the physician does not sign off on the lab report, it is not included in the CCDA.

Figure 38

Recommended Workflow



Note: The recommended workflow, which integrates the constraints proposed by Guerrero et al. (2019), as illustrated in the above diagram, enables the RGV HIE to gather the necessary data from the entire patient population or specific subgroups. The process begins with an analyst who works under the auspices of Su Clinica while operating within the clinic environment. This analyst obtains the pertinent data and provides deidentified and/or aggregated count data, subsequently forwarding the data to Su Clinica. The data will only be sent if Su Clinica approves

it; otherwise, it will not be transmitted. The remaining steps in the process are the same as those depicted in the preceding diagram.