

Title: **VCU/VCUHS Enterprise Data Warehouse**

Protocol No: EDW4R_001

Protocol Date: 10/24/2022

F. Gerard Moeller, MD
Principal Investigator
VCU

Richmond, VA, 23298

Phone 804-827-1531

email: frederick.moeller@vcuhealth.org

Alison.montpetit@vcuhealth.org

Alison Montpetit, RN, PhD
Co-Principal Investigator
VCUHS

Richmond, VA 23298

Phone 804-628-4545

email:

ABBREVIATIONS AND DEFINITIONS OF TERMS

AE	Adverse event
CMH	Community Memorial Hospital
CTRIDW	Clinical and Translational Research Informatics Data Warehouse
EDW	Enterprise Data Warehouse
EDW4R	Enterprise Data Warehouse for Research
EHR	Electronic Health Record
MCV	Medical College of Virginia
TDAAC	Tissue and Data Acquisition and Analysis Core
VCU	Virginia Commonwealth University
VCUHS	Virginia Commonwealth University Health System

ABSTRACT

Context: (Background)

Access to electronic health record (EHR) data, other clinical data and administrative data for research is critically important for development of new treatments, best practices for use of current treatments, and understanding the biological and social/behavioral causes of diseases. An enterprise data warehouse for research (EDW4R) is needed to provide secure, compliant, and efficient storage and access to data for research. This protocol establishes the framework for that data warehouse.

EDW4R will be overseen by Wright Center informatics staff who will serve as honest brokers for access to the data through an agreement with VCUHS. Investigators will be required to provide approval or exemption from IRB oversight when they submit their data request through the EDW4R electronic request system. Honest brokers will then work with individual investigators who have IRB approval for access to EDW4R data and will provide them with the appropriate data depending on the IRB approval and legal requirements (de-identified, limited data set, fully identified data) It is important to note that the only IRB protocols that can be associated with or covered by the current EDW4R project are those in which investigators are working with Wright Center Informatics honest brokers and/or Massey Cancer Center honest brokers.

Objectives: (primary and important secondary objectives)

The purpose of this protocol is to provide the infrastructure for compliant, secure, and efficient storage and access to EHR data, other clinical data and administrative data for research.

Study Design:

The oversight of the EDW4R will exclusively involve the vigilant supervision provided by the informatics staff of the Wright Center. These informatics experts will function as honest brokers who have undergone training and certification by the VCUHS Privacy Office. Their primary role will be to facilitate controlled access to the data contained within the warehouse for designated data warehouse studies in strict accordance with the established agreement between the Wright Center and VCUHS.

Data housed in the EDW4R includes electronic health record (EHR) data from VCU Health System (VCUHS) (including, but not limited to, VCU Medical Center, Children's Hospital of Richmond (CHoR), VCU Health Tappahannock, and VCU Health Community Memorial Hospitals), as well as other types of data, such as imaging and scheduling data. The EDW4R resides on a secure server behind our partner health system's secure firewall, with secure Duo two factor authentication. The EDW4R will be overseen by Wright Center informatics staff who will serve as honest brokers for access to the data through an agreement with VCUHS. Honest brokers will work with individual investigators who have IRB approval for access to

EDW4R data when required and will provide them with the appropriate data depending on the IRB approval and legal requirements (de-identified, limited data set, fully identified data).

Setting/Participants:

EDW4R data includes data from VCUHS partner hospitals. VCU Medical Center Hospital, located in Richmond Virginia, has 865 beds with over 37,000 admissions in 2020. Children's Hospital of Richmond had over 265,000 pediatric outpatient visits, over 20,000 emergency room visits, and over 9,000 inpatient admissions in 2019. VCU Health Community Memorial Hospital (CMH) is a nonprofit health care facility treating patients in the rural counties of Brunswick, Lunenburg, and Mecklenburg in Virginia, and adjacent rural areas in North Carolina. Located in South Hill, Virginia (approximately 80 miles south of Richmond), CMH provides care to patients in a large HRSA designated rural area, with over 86,000 emergency room and 850,000 outpatient visits in 2020. VCU Health Tappahannock Hospital (Tappahannock) is located in Tappahannock, Virginia, a HRSA designated rural area of north central Virginia.

Data/Specimen Collection Procedures and Frequency:

Data is collected on a regular basis from VCUHS' EHR (Epic) where it is stored in the EDW4R, which resides in a secure location behind the VCUHS firewall. In addition, the EDW4R provides a secure location for storage of other data, such as death registry data, brain imaging data, genomic data, and other data which can be linked to EHR data with appropriate approvals.

1 BACKGROUND INFORMATION AND RATIONALE

Electronic patient data are critical to the conduct of clinical and translational science, and experts have noted that “an enterprise data warehouse (EDW) and related informatics core services for research purposes are no longer optional components for a robust translational research enterprise.” [1]. Among Clinical and Translational Science Award (CTSA) hubs funded by the National Institutes of Health (NIH) National Center for Advancing Translational Science (NCATS), adoption of EDWs containing electronic health record (EHR) data for research has increased from 64% in 2008 [2] to 94% in 2017 [3]. Potential benefits of EDW for research (EDW4R) services include more effective and efficient clinical and translational research support through access to EHR data that have been curated by informatics experts for scientific purposes [4]. For example, EDW4R have enabled acceleration of drug development pipelines [5], phenome-wide association studies when linked with genotyped biospecimens [6], rapid pharmacovigilance [7], characterization of treatment pathways for 250 million patients in 11 countries [8], multisite clinical trial recruitment [9], automated data collection [10], and delivery of real-world evidence at the point of care to actualize the learning healthcare system [11]. With respect to technology, the literature contains several descriptions of EDW4R infrastructure at CTSA hubs [12-21]. From this literature comes a wide variety of functions, capabilities, and needs, including data integration, management, education, support, tooling, governance, optimization, and alignment across missions [22].

This protocol describes the data resources implemented at VCUHS to support Virginia Commonwealth University's (VCU) clinical and translational science.

Potential uses of the EDW4R include COVID-19 research, cancer research, drug overdose research, and cardiovascular and liver disease research.

The EDW4R is funded by VCUHS and the Wright Center institutional funds as well as NIH funds through our Clinical and Translational Science Award (CTSA).

1.1 Compliance Statement

This study will be conducted in full accordance with all applicable VCU and VCUHS Policies and Procedures and all applicable Federal and State laws and regulations including the Common Rule at 45 CFR Part 46, and the HIPAA Privacy Rule at 45 CFR Part 160, 164 and 21 CFR Part 11. Any episode of noncompliance will be documented and reported as required.

The investigators will perform the study in accordance with this protocol and will report unexpected problems in accordance with VCU IRB Policies and Procedures and all Federal requirements. Collection, recording, and provisioning of data will be accurate and will ensure the privacy, health, and welfare of research subjects.

All key study personnel will complete the CITI Biomedical Research training and the VCUHS HIPAA for Research Training. VCU employees who will be utilizing VCUHS data must complete a Data Use Agreement (DUA) annually and have completed appropriate VCUHS forms to have access to PHI for research purposes. The Wright Center honest broker team is responsible for keeping documentation of all training,

forms, and certifications. This will be maintained on a secure VCU server. All requests for data access will be submitted by individual investigators through the data access portal on the Wright Center website. The central Honest Broker will assure that all requests for protected health information (PHI) are IRB approved and will be responsible for assuring that released data follows the Honest Broker processes, defined by the Honest Broker SOPs.

2 STUDY OBJECTIVES

One of VCU's main missions is to support "research and discovery that advances knowledge, inspires creativity and improves human health" (quote from VCU's Mission Statement). To achieve this goal, the Wright Center for Clinical and Translational Research in collaboration with the Massey Cancer Center created a shared Clinical and Translational Research Informatics resource including a data repository to support VCU's research enterprise. The goal of this resource is to provide easy access to biomedical data for research purposes. At the same time this shared resource supports secure and compliant processes to make sure that patient privacy is preserved, and relevant regulations are observed.

The VCU Clinical and Translational Research Informatics shared resource focuses on three capabilities:

- 1) Support VCU's clinical and translational research enterprise by providing access to biomedical data for research purposes.
- 2) Support compliant and secure data sharing between VCU and VCUHS.
- 3) Provide these informatics services in accordance with VCU and VCUHS policies, as well as with relevant laws and regulations.

2.1 Primary Objective (or Aim)

The VCU/VCUHS Enterprise Data Warehouse for Research (EDW4R) is designed to house biomedical data in a secure way to produce valuable information for researchers. EHR data is provided through data feeds from VCUHS, and all processes and procedures are overseen and monitored by VCUHS compliance and security personnel.

Data is stored at the “atomic” level (lowest level of detail prior to any transformation to specific data formats) (Visweswaran 2021). Prior to data provisioning to investigators, data is converted to the appropriate format (OMOP, i2b2) and identification level (de-identified, limited data set, fully identified) by the honest broker team based on the IRB approval and legal requirements (see figure 2 below).

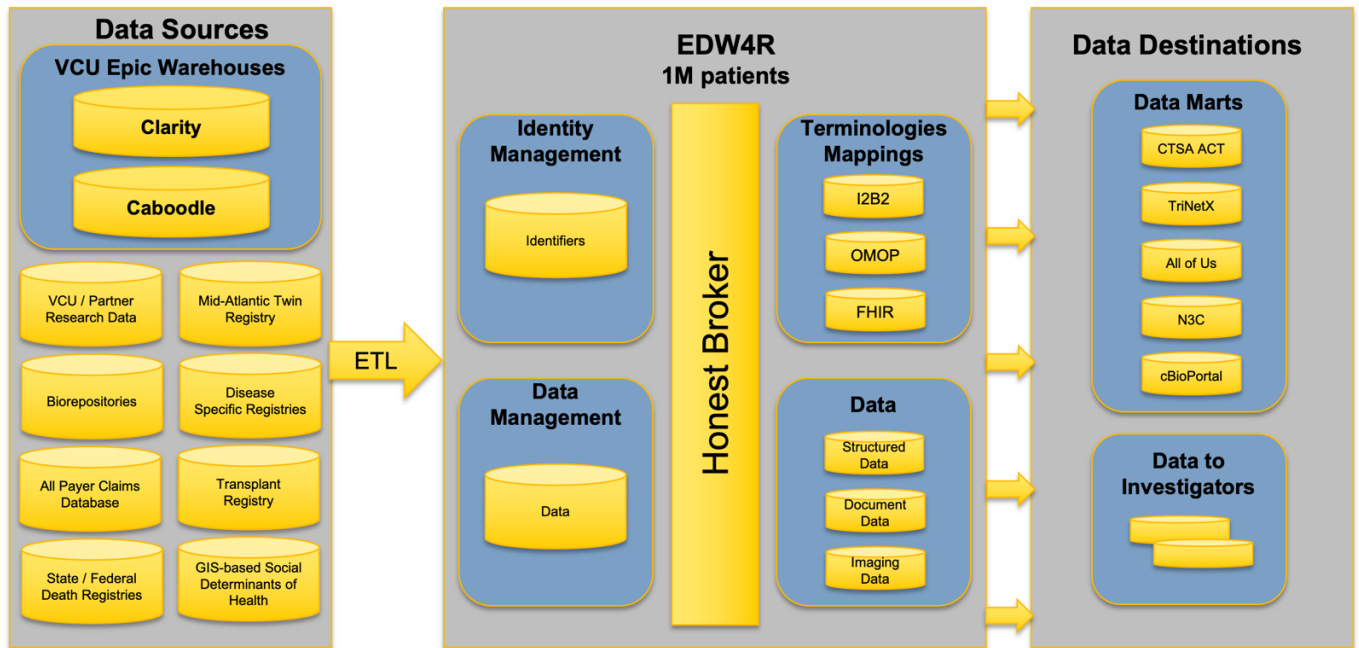


Figure 2 EDW4R Structure

3.1 General Schema of the Repository Design

The EDW4R is a collaborative effort between VCU and VCUHS to provide long-term storage and access to clinical, and administrative data to support scholarly work. This will be accomplished by use of a cloud-based, secure data platform that enables access to biomedical and health data as well as advanced analytical and statistical tools.

3.1.1 Total Number & Description of Study Sites/Total Number of Subjects Projected

Data will be collected for the repository from the VCUHS EHR, which includes data from all facilities and associated satellite locations owned and operated by VCUHS and its owned and controlled affiliates.

Data stored in the repository will include the entire medical record of patients registered at VCUHS facilities and will include patient-level clinical (such as but not exclusively cancer data, COVID-19 data, overdose data) and administrative data.

The repository will be maintained within the secure VCUHS IT environment with access restricted to Honest Brokers as approved by VCUHS. The repository will adhere to all VCUHS policies and procedures governing data and information

technology. Investigators requesting data from the repository must be qualified and trained per VCU policies CR-AD-105 (Investigator Responsibilities) and CR-AD-110 (Participating Investigator Qualification) and all requests for data from the repository will require PROC/PRMC/IRB approval prior provisioning of data.

3.1.2 Overview of Data Collection

Data is extracted from the VCUHS EHR (Epic) into the EDW4R on regular basis. In addition, other data sources are included in the EDW4R such as imaging, administrative data (such as billing and scheduling). The up-to-date list of data sources feeding data into the EDW4R will be available in the EDW4R Data Architecture SOP. This SOP also describes the processes of adding and removing data sources.

The data is collected and stored in an identified manner to enable research requiring identifiers and data linkage across time. Data will be provided to investigators based on the IRB approval for their individual research protocols.

Repository Population

This repository is designed to foster future research, and, thus includes comprehensive patient and encounter-level data from the VCUHS data sources.

3.1.3 Inclusion Criteria

Registered as a patient at a VCUHS facility. Patients may request to have their data removed from the EDW4R as discussed in VCUHS notice of privacy practices and authorization for healthcare services form that all patients are required to sign which states: "Teaching and Research Hospital: (i) qualified individuals, as part of their education and training at VCUHS, may observe or participate in the delivery of healthcare services to me; (ii) my patient records may be reviewed by students, residents, trainees, employees, and/or faculty of VCUHS, Virginia Commonwealth University ("VCU") and other approved institutions for teaching and research purposes; and (iii) my records may be reviewed for research eligibility and I may be approached about my interest in taking part in a research study."

3.1.4 Exclusion Criteria

Not registered as a patient at a VCUHS facility.

4.0 REPOSITORY PROCEDURES

4.1 Data Collection

Data will be obtained from multiple sources and stored in the research data repository. Those sources include, but are not limited to: EHR, administrative, and imaging data.

The data will be electronically transitioned from the Epic production system to the research data repository on regular basis. Other data may be updated at different frequencies based on the source of the data.

4.2 Medical Record / PHI Elements Collected

The purpose of the EDW4R is to provide a basis for scholarly discovery through data, which includes the necessity of enabling the linkage of clinical data from VCUHS with other data sources using patient identifiers. Thus, PHI will be stored in the EDW4R. Storage of PHI is necessary to provide Honest Broker services by linking disparate data sources based on patient identifiers, as well as to identify patients for clinical research purposes.

4.3 Data From Existing Repositories That Include VCUHS Data

The EDW4R also will consist of pre-existing data from VCUHS and VCU sources. Data collected under protocol HM15380 (“Honest Broker System and Processes for De-identification of Clinical or Other PHI Data”) and HM20023298 (“VCU Clinical Data Repository for Research”) will be transitioned to the VCU/VCUHS Research Data Repository and housed within a single platform to ensure compliance, security, and efficiencies. This IRB protocol will replace these previous protocols, which will be closed. Data includes, but is not limited to:

- VCUHS, MCVH, and MCVP administrative data, including billing and appointments from the Cerner EHR;
- VCUHS cancer registry data;
- Clinical data from Cerner and Epic (the VCUHS electronic medical record) including discrete fields (e.g., lab results) and text fields (e.g., radiology reports);
- Clinical and administrative data from the bone marrow transplant database;
- Imaging data from the PACs system
- Clinical and administrative data from the radiation oncology systems;
- Clinical genomics data from the Clinical Genomics lab at VCU, as well as outside laboratories. These data include any data from the data processing pipelines from raw data (fastq, SAM, BAM, VCF) to annotated clinical findings; and
- Biospecimen information from the Massey Cancer Center Tissue and Data Acquisition and Analysis Core. The data includes patient and specimen identifiers as well as pathology, clinical and genomics data describing the specimens.

A complete list of data sources will be stored in the EDW4R SOPs.

5 REPOSITORY ADMINISTRATION

To maximize capabilities for future research and protect patient confidentiality, policies and procedures are in place to ensure the security of the repository data and SOPs are available to control provisioning of data, the mechanisms to release data, and the proper pathway to identify/de-identify data when released.

5.1 Study Organization

Access to the EDW4R will be strictly limited to Honest Brokers who have met all training and access requirements. Honest Brokers will be approved by VCUHS and will undergo extensive training prior to initiating Honest Broker responsibilities. The training will include general human subjects research training and specific honest broker training as described in detail in the Honest Broker SOPs.

Data from the repository will be distributed by the Honest Brokers via one of the three conditions for release of data:

- (1) without identifiers (IRB exempt determination, and no HIPAA authorization or waiver required; recipient will be required to execute a Data Use Agreement);
- (2) with only a limited data set (as defined by HIPAA) (usually IRB exempt and no HIPAA authorization or waiver required; recipient will be required to execute a Data Use Agreement);
- (3) with identifiers (requires IRB approval or IRB determination of exemption and may require waiver of consent and HIPAA authorization, or individual consent and HIPAA authorization).

Researchers must adhere to the following requirements prior to accessing data provisioned by the Honest Brokers:

- VCU required Human Subjects Research training documentation
- Signed data use agreement (DUA) to acknowledge rules and regulations associated with use of EDW4R.

5.2 Data Collection and Management

5.2.1 Computer Systems

Data will be stored in the Snowflake Data Cloud, which is a secure platform that unifies multiple sources of data. Snowflake has built-in cloud data security measures such as always-on encryption of data in transit and at rest and complies with government and industry regulations. Snowflake is HIPAA-compliant and government security certified and has been approved by VCU and VCUHS data security. Access to the Snowflake EDW4R is only provided to Wright Center informatics Honest Brokers.

5.2.2 Confidentiality of Subjects:

The EDW4R will be protected during storage and provisioning by maintaining strict confidentiality and security procedures. Access will be limited to Honest Brokers, Data will be disclosed to researchers only as consistent with IRB approval, legal requirements and VCUHS policies.

Regulations:

45 CFR 164.512(i) (HIPAA research rules); 45 CFR 164.514(b) (requirements for de-identification of data and creation of limited data sets)

5.3 Data Provisioning

Upon researcher completion of all training requirements and confirmation of all executed agreements, data will be provisioned to researchers by the Honest Broker. De-identification of data will adhere to expert determination and/or Safe Harbor, as determined by the IRB and VCUHS policies.

5.4 Providing Results to Subjects

This is a repository of data and no results will be reported back to subjects.

5.5 Regulatory and Ethical Considerations

5.5.1 Risk Assessment

Risks associated with the EDW4R are no greater than minimal with the primary risk being a breach of privacy and confidentiality for patients whose data are included in the repository. While there is always a risk that patient data may be improperly released or accessed, this risk is not higher than what a person might reasonably experience in being a patient at VCUHS. Risk will be minimized through the following: policies and procedures are in place to protect against unauthorized access or release of patient data, storage of data will be within the secure VCUHS IT environment, only Honest Brokers will have access in accordance with the Honest Broker SOP, and all federal and state law and VCU and VCUHS policies will be followed related to safeguards for privacy and security of the data.

5.5.2 Potential Benefits of Participation

This repository is designed to foster future research and, therefore, there are no direct benefits for patients whose data are included in the repository.

5.6 Recruitment Strategy

Data for all patients registered at a VCUHS facility will be included in the repository.

5.7 Informed Consent/Assent and HIPAA Authorization

A waiver of consent is requested as the research involves no more than minimal risk to subjects; The research could not be carried out practicably without the waiver or alteration; The waiver or alteration will not adversely affect the rights and welfare of the subjects; Where appropriate, the subjects will be provided with additional information about their participation; and this research could not be carried out practicably without using the information/specimen in an identifiable form.

5.8 Payment to Subjects/Families

There will be no payment to patients whose data are stored in the repository.

6 SAFETY MANAGEMENT

6.1 Clinical Adverse Events

Unanticipated problems involving risks to subjects and others will be monitored throughout the study.

6.2 Adverse Event Reporting

Since the study procedures are not greater than minimal risk and are limited to existing data and specimens, SAEs are not expected. If any unanticipated problems related to the research involving risks to subjects or others happen during the course of this study these will be reported to the IRB in accordance with VCU IRB policy. AEs that do not meet prompt reporting requirements will be summarized in narrative or other format and submitted to the IRB at the time of continuing review or will be tracked and documented internally by the study team but not submitted to the IRB.

7 PUBLICATION

Publications based on the EDW4R will be based on results of the individual IRB approved studies that access the EDW4R through the Honest Brokers. For all publications only aggregate data without individually identifiable information will be published, as explicitly stated in the data use agreements.

References

1. Obeid JS, Tarczy-Hornoch P, Harris PA, et al. Sustainability considerations for clinical and translational research informatics infrastructure. *J Clin Trans Sci* 2018; 2 (5): 267–75.
2. MacKenzie SL, Wyatt MC, Schuff R, Tenenbaum JD, Anderson N. Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey. *J Am Med Inform Assoc* 2012; 19 (e1): e119–24.
3. Obeid JS, Beskow LM, Rape M, et al. A survey of practices for the use of electronic health records to support research recruitment. *J Clin Trans Sci* 2017; 1 (4): 246–52.
4. Payne PRO, Johnson SB, Starren JB, Tilson HH, Dowdy D. Breaking the translational barriers: the value of integrating biomedical informatics and translational research. *J Investig Med* 2005; 53 (4): 192–200.
5. Pulley JM, Shirey-Rice JK, Lavieri RR, et al. Accelerating precision drug development and drug repurposing by leveraging human genetics. *Assay Drug Dev Technol* 2017; 15 (3): 113–9.

6. Denny JC, Bastarache L, Roden DM. Phenome-Wide Association Studies as a Tool to Advance Precision Medicine. *Annu Rev Genom Hum Genet* 2016; 17 (1): 353–73.
7. Kohane IS, Churchill SE, Murphy SN. A translational engine at the national scale: informatics for integrating biology and the bedside. *J Am Med Inform Assoc* 2012; 19 (2): 181–5.
8. Hripcsak G, Ryan PB, Duke JD, et al. Characterizing treatment pathways at scale using the OHDSI network. *Proc Natl Acad Sci U S A* 2016; 113 (27): 7329–36.
9. Claerhout B, Kalra D, Mueller C, et al. Federated electronic health records research technology to support clinical trial protocol optimization: evidence from EHR4CR and the InSite platform. *J Biomed Inform* 2019; 90: 103090.
10. Champion TR, Sholle ET, Davila MA. Generalizable middleware to support use of redcap dynamic data pull for integrating clinical and research data. *AMIA Jt Summits Transl Sci Proc* 2017; 2017: 76–81.
11. Longhurst CA, Harrington RA, Shah NH. A “green button” for using aggregate patient data at the point of care. *Health Aff (Millwood)* 2014; 33 (7): 1229–35.
12. Baghal A, Zozus M, Baghal A, Al-Shukri S, Prior F. Factors associated with increased adoption of a research data warehouse. *Stud Health Technol Inform* 2019; 257: 31–5.
13. Danciu I, Cowan JD, Basford M, et al. Secondary use of clinical data: the Vanderbilt approach. *J Biomed Inform* 2014; 52: 28–35.
14. Chute CG, Beck SA, Fisk TB, Mohr DN. The enterprise data trust at Mayo Clinic: a semantically integrated warehouse of biomedical data. *J Am Med Inform Assoc* 2010; 17 (2): 131–5.
15. Sholle ET, Kabariti J, Johnson SB, et al. Secondary use of patients’ electronic records (SUPER): an approach for meeting specific data needs of clinical and translational researchers. *AMIA Annu Symp Proc* 2017; 2017: 1581–8.
16. Kamal J, Liu J, Ostrander M, et al. Information warehouse—a comprehensive informatics platform for business, clinical, and research applications. *AMIA Annu Symp Proc* 2010; 2010: 452–6.
17. Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE—an integrated standards-based translational research informatics platform. *AMIA Annu Symp Proc* 2009; 2009: 391–5.
18. Wade TD, Hum RC, Murphy JR. A Dimensional Bus model for integrating clinical and research data. *J Am Med Inform Assoc* 2011; 18 (Suppl 1): i96–102.
19. Starren JB, Winter AQ, Lloyd-Jones DM. Enabling a learning health system through a unified enterprise data warehouse: the experience of the Northwestern

University Clinical and Translational Sciences (NUCATS) Institute. Clin Transl Sci 2015; 8 (4): 269–71.

20. Mosa ASM, Yoo I, Apathy NC, Ko KJ, Parker JC. Secondary use of clinical data to enable data-driven translational science with trustworthy access management. Mo Med 2015; 112 (6): 443–8.

21. Waitman LR, Warren JJ, Manos EL, Connolly DW. Expressing observations from electronic medical record flowsheets in an i2b2 based clinical data repository to support research and quality improvement. AMIA Annu Symp Proc 2011; 2011: 1454–63.

22. Champion TR, Craven CK, Dorr DA, Knosp BM. Understanding enterprise data warehouses to support clinical and translational research. J Am Med Inform Assoc. 2020 Jul 1;27(9):1352-1358.

23. Visweswaran S, McLay B, Cappella N, Morris, M, Milnes JT, Reis SE, Silverstein JC, Becich MJ. An atomic approach to the design and implementation of a research data warehouse. J Am Med Inform Assoc. 2021 1-8

APPENDIX

Attach a listing of data elements from each data source, if not listed in Sections 4 and 5.

Data elements include all data stored in the VCUHS EHR. As such a full listing of data elements would not be feasible.