# Integrated Knowledge Management (IKM) Volume 7

## Version 1 - Last Updated 05/16/2024

# Table of Contents

I. Use Cases ..... 1
1. Summary of Findings - Current State of Safety Systems ..... 3
1.1. Motivation ..... 3
1.2. Background ..... 3
1.3. Objectives and Approach ..... 4
1.4. Current Draft Model of the Laboratory Ecosystem Control Structure ..... 4
1.5. Requirements ..... 5
1.6. Discussion ..... 8
1.7. References ..... 8
2. Summary of Findings – RWE Search and Query Methods ..... 9
2.1. Real World Evidence (RWE) Introduction ..... 9
2.2. Long COVID Use Case: Loosely Defined and/or Poorly Understood Scenarios ..... 9
2.3. References ..... 11
3. Use Cases ..... 12
3.1. Notice ..... 12
3.2. Use Case: Komet Application ..... 12
3.2.1. Komet Use Case – Rapid Knowledge Management ..... 12
3.2.2. Komet Use Case – Integration of Terminology Standards ..... 12
3.2.3. Komet Use Case – Detecting Equivalent Concepts ..... 13
3.3. Use Case: Considerations for Real World Data Interoperability and Data Agility in Patient Journey to Diagnosis for Emerging Acute and Chronic Infectious Diseases ..... 13
3.3.1. Introduction to the Emerging Acute and Chronic Infectious Diseases Use Case ..... 13
3.3.2. Aims for the Emerging Acute and Chronic Infectious Diseases Use Case ..... 14
3.3.3. Emerging Acute and Chronic Infectious Diseases Use Case Background ..... 14
3.3.4. Methods to Develop the Emerging Acute and Chronic Infectious Diseases Use Case ..... 15
3.3.5. Results from the Emerging Acute and Chronic Infectious Diseases Use Case ..... 19
3.3.6. Emerging Acute and Chronic Infectious Diseases Use Case Discussion ..... 22
3.3.7. Future Considerations for the Emerging Acute and Chronic Infectious Diseases Use Case ..... 23
3.4. References ..... 23

# List of Figures

1.1. Detailed Laboratory Ecosystem Control Structure ..... 5
1.2. Draft Inset Model of the Encoding for Data Flow in the Laboratory Ecosystem Control Structure ..... 5
3.1. Generalized Current State Long COVID Workflow ..... 16
3.2. Current State Outpatient Long COVID Workflow ..... 17
3.3. Future State Outpatient Long COVID Workflow ..... 18
3.4. Idealized Future State Data Long COVID Workflow ..... 19

# List of Tables

3.1. Future State Data Collection Points ..... 21

# Part I. Use Cases

# Table of Contents

1. Summary of Findings - Current State of Safety Systems ..... 3
1.1. Motivation ..... 3
1.2. Background ..... 3
1.3. Objectives and Approach ..... 4
1.4. Current Draft Model of the Laboratory Ecosystem Control Structure ..... 4
1.5. Requirements ..... 5
1.6. Discussion ..... 8
1.7. References ..... 8
2. Summary of Findings – RWE Search and Query Methods ..... 9
2.1. Real World Evidence (RWE) Introduction ..... 9
2.2. Long COVID Use Case: Loosely Defined and/or Poorly Understood Scenarios ..... 9
2.3. References ..... 11
3. Use Cases ..... 12
3.1. Notice ..... 12
3.2. Use Case: Komet Application ..... 12
3.2.1. Komet Use Case – Rapid Knowledge Management ..... 12
3.2.2. Komet Use Case – Integration of Terminology Standards ..... 12
3.2.3. Komet Use Case – Detecting Equivalent Concepts ..... 13
3.3. Use Case: Considerations for Real World Data Interoperability and Data Agility in Patient Journey to Diagnosis for Emerging Acute and Chronic Infectious Diseases ..... 13
3.3.1. Introduction to the Emerging Acute and Chronic Infectious Diseases Use Case ..... 13
3.3.2. Aims for the Emerging Acute and Chronic Infectious Diseases Use Case ..... 14
3.3.3. Emerging Acute and Chronic Infectious Diseases Use Case Background ..... 14
3.3.4. Methods to Develop the Emerging Acute and Chronic Infectious Diseases Use Case ..... 15
3.3.5. Results from the Emerging Acute and Chronic Infectious Diseases Use Case ..... 19
3.3.6. Emerging Acute and Chronic Infectious Diseases Use Case Discussion ..... 22
3.3.7. Future Considerations for the Emerging Acute and Chronic Infectious Diseases Use Case ..... 23
3.4. References ..... 23

# 1. Summary of Findings - Current State of Safety Systems

## 1.1. Motivation

Clinical laboratories are a key element in the overall healthcare ecosystem and play a role in most healthcare decisions today. Clinical laboratories are responsible for generating highly reliable laboratory data (orders, results, and interpretations) to drive effective care delivery. Clinical data must transverse many connected systems while maintaining context and precise semantic meaning. While each sending and receiving system of clinical data is configured individually to understand data inputs, the loss of meaning between systems is difficult to avoid, impacts patient safety, and hinders data science opportunities like machine learning and artificial intelligence. Plausible efforts to control for the nuances of laboratory data exchange have existed for decades. The advancement of technology increasingly emphasizes the need for standardization across the ecosystem as data science methods and capabilities mature. Safety and quality controls exist in every layer of the laboratory ecosystem, starting at the top of the ecosystem with Federal policymakers down to the individual patients receiving care. While the use of idiosyncratic (i.e., local institution–specific) identifiers for laboratory tests is recommended by Standards Development Organizations (SDO), such as Systematized Nomenclature of Medicine International (SNOMED International), entities in the laboratory data ecosystem must implement precise data mappings to a universal terminology. Idiosyncratic local identifiers alone hinder seamless sharing of laboratory observations between disparate health systems.

## 1.2. Background

According to the Health Information Management (HIM) Body of Knowledge™ by the American Health Information Management Association (AHIMA), Health Data Standards (HDS) are documented agreements on representations, formats, and definitions of common data. [1] To achieve interoperability, health data entities must conform to industry standards and specifications which act as a source of truth for methods of codifying information captured and exchanged. Data standards are developed and maintained by SDOs who are member-supported organizations which act like a legislative body with detailed internal processes to ensure consistency and fairness among the entities subject to them.

Two of the major Health Information Technology (IT) SDOs in our analysis are Regenstrief Institute, Inc. and Systematized Nomenclature of Medicine International (SNOMED International). Regenstrief Institute, Inc. organized LOINC® in 1994 to standardize a common terminology for laboratory and clinical observations as trends in electronic clinical data exchange were taking form. [2] Today, LOINC® encodings are most often exchanged via Health Level 7 (HL7) International Version 2 transactions between health systems. SNOMED Clinical Terms® (SNOMED CT®) is an international standard for several purposes including problem list and public health reporting and is required by many countries' certification criteria for EHRs. It is a computer-processable collection of medical terms, codes, synonyms, and definitions used in clinical documentation and reporting. Both SNOMED CT® and LOINC® can be used as a common terminology to represent clinical information consistently and comprehensively in the electronic exchange of health data. However, two or more common terminologies which serve near-identical purposes is oxymoronic and leaves room for improvement, consolidation, and harmonization. For lab data exchange, SNOMED CT® is typically used to encode test results and observations and LOINC® is often used to encode tests (however, LOINC® does also represent some test results) and these relationships, overlaps, and contradictions present implementers with the need to use these terminologies in more integrated ways.

# 1.3. Objectives and Approach

This analysis aims to provide foundational safety research and development support to reduce diagnostic errors and assist with proficiency testing and compliance by integrating systems theory and safety engineering methods used by other high-risk industries, such as aviation, military special operations, and nuclear power into laboratory testing and management processes and systems. This research will apply a System Safety approach to assess, measure, document, and analyze the safety and quality of the whole laboratory data ecosystem.

One of the initial tasks is to assess and evaluate the safety of the current laboratory data ecosystem using a System Safety Engineering Approach to identify safety hazards, map the design, and model control structures of the current system. This research aim uses the System Theoretic Process Analysis (STPA) safety assessment process to identify hazards affecting data reliability, interoperability, data integrity, and data quality across the ecosystem, as well as providing an understanding of the clinical laboratory ecosystem controls, constraints, actuators, processes, and feedback loops for safe and effective operation. [3] By modeling a control structure, we can assess every relationship therein to measure the effectiveness of control actions and identify those which are unsafe to the system operation, human operators, and ultimately, patients.

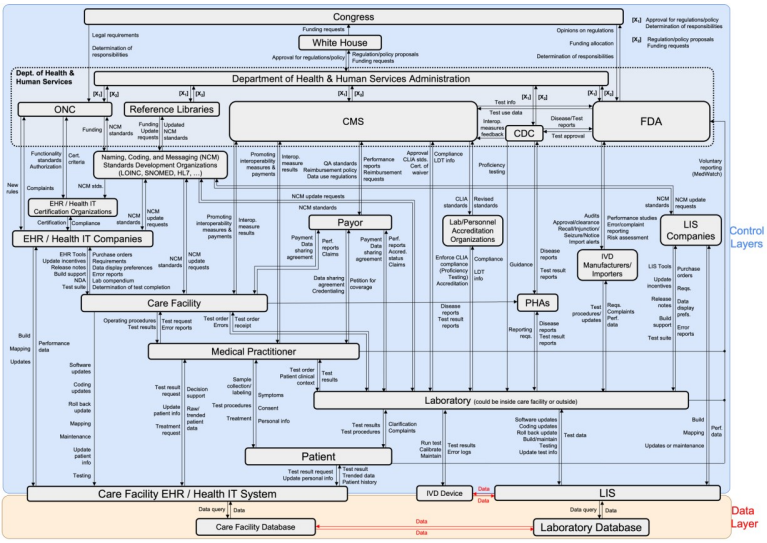
This document analyzes an initial control structure to identify and evaluate safety and quality controls and loss scenarios within the laboratory data ecosystem. Specifically, we focus on describing entities, behaviors, and business requirements to complement the current model of the control structure that supports interoperability, use of terminology systems, and industry coding standards such as LOINC® and SNOMED CT®. Within the control structure diagrams that have been developed, we are enumerating requirements for the "Data Flow for Encoding" to describe how data is encoded from labs and Laboratory Information System (LIS) systems to standards, how data is encoded between manufacturers and standards, and how new test names and codes are distributed by SDOs. These requirements will help to clarify and articulate control action scenarios. They further pinpoint specific scenarios in the safety assessment analysis by identifying unsafe control actions and loss scenarios related to the standards development, implementation, and maintenance processes in laboratories, devices, and standards.

# 1.4. Current Draft Model of the Laboratory Ecosystem Control Structure

The initial modeling and analyses of the laboratory ecosystem control structure visualize the influences between ecosystem entities such as policy set by the federal government and the relationship that hospitals and laboratories form to design and perform the care delivery in collaboration with tech companies. The current analysis emphasizes audits performed by a public entity on in vitro diagnostic (IVD) manufacturers and laboratories. While the initial feedback with our stakeholders is positive, it could be made more comprehensive with the inclusion of SDOs' process such as the interactions and influences between the various players with LOINC® and SNOMED International. Our documentation describes the context and business requirements to supplement these diagrams.

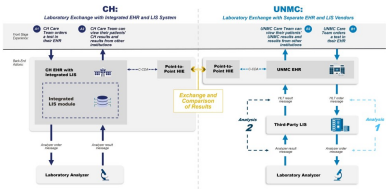
Figure 1.1, "Detailed Laboratory Ecosystem Control Structure" below illustrates the current draft model of the laboratory ecosystem control structure and Figure 1.2, "Draft Inset Model of the Encoding for Data Flow in the Laboratory Ecosystem Control Structure " includes a specific draft depiction of the data flow for encoding.

**Figure 1.1. Detailed Laboratory Ecosystem Control Structure**




In Figure 1.2, “Draft Inset Model of the Encoding for Data Flow in the Laboratory Ecosystem Control Structure ”, “middleware” is currently in the model as an interface between devices and LIS/LIMS; however middleware can and does exist in between several other systems in this ecosystem. Additionally, “LIMS” are typically the source for reporting to public health agencies.

**Figure 1.2. Draft Inset Model of the Encoding for Data Flow in the Laboratory Ecosystem Control Structure**




# 1.5. Requirements

This section details business requirements related to the data flow for encoding data in laboratory information for LISs, IVD vendors, and SDOs. These requirements are intended to help articulate control action scenarios and pinpoint specific scenarios in the safety assessment analysis in identifying unsafe control actions and loss scenarios related to the standards development, implementation, and maintenance processes in laboratories, devices, and standards.

**Laboratories must document a comprehensive list of tests that they perform.** Laboratories maintain a comprehensive test menu (also referred to ask a test catalog or laboratory test compendiums). These test menus contain the list of all tests performed by a#particular#lab and tests that are offered to consumers. The test menus are often delineated to include a list of tests that a particular analytical instrument is capable of performing. There are many details about tests that are important for laboratorians to understand and represent in test menus. These details are not limited to the following: name of the test, the component (or analyte) being measured, the specimen type, the property being measured, the timing of the measurement, the system (usually sample type for laboratory measurements), the scale of measurement, the method used to produce the observation, rejection and acceptance criteria, type of instrument used, specific make and model of instrument or kit used, testing priority, performing site, interpretation criteria, the individual responsible for interpretation, related diagnosis and other additional contextual information.

Of these data elements, the name of each test is of particular importance because the name of the test is typically displayed in user interfaces and electronic messages and reports. Consequently, many of the contextual details listed above are reduced to and obfuscated by only being represented by the name of each test alone.

The name of every test can be stored in laboratory information systems using local, institution-specific naming conventions. Ideally, laboratories can use a textual description with an unlimited number of characters to accurately describe each laboratory order and result. However, this can result in challenges due to field space restrictions and character limits of databases in the information system. Alternatively, “short names” may be pragmatically implemented to limit the characters allowed in textual names/descriptions of tests but can be complex with abbreviations that result in descriptions that are very difficult to interpret (e.g., albumin ser-plr fld-MCDiff, which refers to the mass concentration difference in grams per liter of albumin in serum and pleural fluid). [2] Rather than only having textual descriptions, laboratory information systems also use unique local identifiers (i.e., codes) to identify and store tests and results.

**Laboratories are expected and, in many cases, required by the government to share data with public health institutions, medical centers, and disease registries to support population-based care, precision medicine,** and research.

Laboratories must be able to receive information about orders for tests and ensure that they are performing the equivalent test. Additionally, laboratories must be able to send information about tests performed and associated results and ensure they are sending results for the correct test that was ordered. However, use of local institution-specific names and local codes in LISs results in difficulty with interpreting the data and identifying equivalence between tests and results without a precise mapping to a common, universal terminology.

Therefore, industry coding standards have been developed and required for use to standardize the codes and terms embedded in data exchange messages. Federal agencies, such as Center for Medicaid and Medicare Services (CMS) and Office of the National Coordinator for Health Information Technology (ONC), have worked in close coordination to define requirements in a joint effort to promote interoperability. The Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 grants ONC the authority to establish programs to promote the adoption of health technologies such as electronic health records and secure data exchange. Additional laws, like the ONC 21st Century Cures Act and CMS Interoperability and Patient Access Rule, and agreements, like ONC's version 2.0 of the Trusted Exchange Framework and Common Agreement (TEFCA), identify specific standards (as shown in HL7’s Fast Healthcare Interoperability Resources (FHIR) Release 4.0.1) to be adopted by industry. Many of these regulations are required for use in hospitals, but non-hospital laboratories are not as strictly required to adhere to the use of industry coding standards.

The adoption of standards like LOINC® aid in distinguishing laboratory data among disparate health systems. However, reliance on standard laboratory test names alone does not go far enough in realizing the benefit of interoperability or minimizing loss scenarios.

**Laboratories should reproducibly encode their test data using industry coding standards and assignment of standard codes should be subjected to a defined process across the code lifecycle.**

Laboratories assign LOINC® codes to the local laboratory tests contained in their test menu. The LOINC® codes need to be mapped in the test definition dictionaries in the LIS. Typically, each laboratory makes its own decisions regarding assignment of LOINC® codes. If different teams are defining LOINC® codes for the LIS and the EHR, or using and storing codes outside of the EHR, there may be discrepancies even within the same institution.#[2]

**Laboratories need to understand and be aware of the subtleties and pitfalls of test code selection that greatly impact the ability to map codes accurately to local test code compendiums.**

For the best practice, a laboratory professional well-versed in LOINC® from the laboratory that is performing the testing should select the optimal LOINC® test order and result codes, as this person best

knows the nuances of how the testing is performed. Downstream or upstream individuals often do not have the complete test details available to them, particularly with regards to testing performed at reference laboratories. In addition, nonlaboratory physicians and technical team members typically do not have adequate information to choose the optimal LOINC® codes correctly (unless the codes are provided by the performing laboratory). From a safety systems perspective, this is potentially a weak control that solely depends on a human controller to prevent loss scenarios and potential harms.

Laboratory personnel responsible for mapping a laboratory's local test compendium to LOINC® codes often do not possess adequate understanding or expertise for LOINC®, available full encoding options, and nuances between similar codes. For this reason, among others, the task of mapping a local test compendium to LOINC® is frequently delegated to information technology staff who, too, are inadequately trained in laboratory data standards and the implementation. [2] Electronic Health Record vendors and their staff interface engineers to build and test HL7 interfaces alongside lab personnel. However, the level of coordination between lab personnel and interface engineers varies from case to case. During interface testing for laboratory interfaces, the most common system errors result from incorrect LOINC® encodings within the OBR and OBX segments of an HL7 Version 2 transaction. The errors are most commonly addressed by information technology staff who in some cases prioritize time and effort over accurate, precise data mappings which can lead to poor due diligence and downstream loss scenarios. Similar process-shortcomings can apply to additional laboratory data elements which demand a certain level of domain expertise, like units of measure and specimen type.

Despite having mapped their unique laboratory tests to LOINC®, laboratories may still need to rely on creating their own internal names and codes for test results. For example, laboratories may need to use a test code to distinguish glucose measurements performed in a core laboratory from those in their satellite facilities. [2] However, all these individual test codes will still need to be mapped to the appropriate, but identical, LOINC® code. Changes to LOINC® codes create challenges for maintenance in LISs, EHRs, and other systems, particularly when the LOINC® code changes for a test that has already been mapped.

**Laboratories need to maintain code sets and mappings over time, which is an error-prone and manual, laborious process.**

Requests for new codes can be made to Regenstrief Institute, the curator of the LOINC® database, through a simple process, and updates to LOINC® are made biannually.

The LIS laboratory test definitions and dictionaries are dynamic and frequently changing, thus requiring a formal process for ensuring uniformity of coding among different instruments or methods and, if applicable, the capacity to accommodate different codes that are correctly applied depending on the specific methodology used for that particular result. There should also be a formal audit process and maintenance to keep up with the changes that occur with each LOINC® release to ensure the accuracy and appropriateness of previous code selection as the database continues to mature. [2]

LOINC's® limitations are poorly understood outside of the laboratory domain, reducing the potential utility for it. Critics of LOINC® will point out the laborious process for assigning the most correct code to each laboratory test offered, and that in practice, LOINC® codes are not uniformly assigned across laboratories. The complexities require a certain level of domain-expertise to understand two tests that may have the same correctly assigned LOINC® code may not necessarily have equivalence to allow for interoperability of their result data.

Ideally, scripting and other automated methods must be developed to facilitate the task of assigning standard codes to local tests.

**IVD Manufacturers**

The production of IVDs requires the assignment of the appropriate LOINC® code(s) for the specific tests which a given device can perform. LOINC® codes are typically in a manufacturer's package insert, as

well as their database. If an appropriate LOINC® code does not exist for a lab test, a manufacturer may submit a request to Regenstrief Institute to create a new LOINC® code.

Many laboratory instrument and reagent vendors that are aware of the immense challenges associated with interoperability have begun to include LOINC® definitions in their package inserts. Such package inserts have become a logical source to provide information that can help laboratories map tests with an appropriate LOINC® code because these vendors are well equipped to identify the appropriate LOINC® code for their specific laboratory test. Moreover, for the introduction of novel laboratory test technology, IVD vendors are the logical respondent to new LOINC® code requests.

The LOINC® to Vendor IVD specification provides LOINC® maps to IVD test results and includes the following important data attributes: the manufacturer, instrument model, unique device identifier, vendor transmission code, vendor specimen description, vendor result description, and test name. Optional data elements are also included to be helpful for laboratory professionals' mapping to LOINC®.

# 1.6. Discussion

The absence of laboratory semantic interoperability for IVD data has been cited as a significant impediment to safe and effective public healthcare. The erosion of accuracy for IVD test data due to interoperability failures can have patient safety consequences and impede timely access to and analysis of lab data on a nationwide or global scale. Differences in encoding laboratory data may have substantive differences in the level of granular detail they convey and contribute to challenges when operationalizing these data into patient care, data exchange, interoperability messages, and data aggregation. The aggregation and compounding of these types of interoperability failures can lead to erroneous conclusions and the potential for patient harm even when the correct LOINC® code is selected. A paper by Stram et al, highlights two laboratories that are both performing urinalysis and selected the same LOINC® code for the test, analyte, method, and instrument. [2] However, when the data was transferred from the LIS to EHR, the quantitative result of count/microliter was being translated to different units in the EHR and would have led to erroneous clinical treatment decisions based on the incorrect interpretation of the threshold for treatment. The use of accurate codes is not sufficient to guarantee that information exchange is accurate across the different levels in the laboratory data ecosystem. A research and development program that applies a System Safety approach to assess, measure, document, and analyze the data quality, integrity, agility, and reliability of the whole lab data ecosystem will be beneficial in promoting safe and effective laboratory medicine. Furthermore, a Knowledge Management Platform that is based on High Reliability Organization (HRO) principles and that offers an integrated and harmonized ecosystem for working across disparate standards will help with managing complexity and change management across standards.

# 1.7. References

1. Wise, L., Miller, N., Warner, D. (n.d.). The role of CDI in Semantic Interoperability. The Role of CDI in Semantic Interoperability / AHIMA, American Health Information Management Association. Retrieved December 14, 2022, from https://bok.ahima.org/doc?oid=302470#.Y5on41HMKUk

2. Michelle Stram, Tony Gigliotti, Douglas Hartman, Andrea Pitkus, Stanley M. Huff, Michael Riben, Walter H. Henricks, Navid Farahani, Liron Pantanowitz; Logical Observation Identifiers Names and Codes for Laboratorians:#Potential Solutions and Challenges for Interoperability.#Arch Pathol Lab Med#1 February 2020; 144 (2): 229–239. doi:#https://doi.org/10.5858/arpa.2018-0477-RA

3. Leveson N, Thomas J, Harrington, Polly, Rodrigo, Rose, Stephen P, Keller, Alana. System Safety within Laboratory Data Exchanges Report [Internet]. Massachusetts Institute of Technology Partnership for Systems Approaches to Safety and Security and Synensys, LLC; 2023 Sep. Available from: https://synensysglobal.com/wp-content/uploads/2023/12/System-Safety-within-Laboratory-Data-Exchanges-Report.pdf

# 2. Summary of Findings – RWE Search and Query Methods

## 2.1. Real World Evidence (RWE) Introduction

Search and query methods are used for information retrieval to access unstructured data. Most health information is stored in an electronic health record (EHR), and most laboratory data are stored in a laboratory information system (LIS). It is important to understand how data is stored and retrieved from these sources, and how to aggregate this data from different sources. Information systems record and manage clinical statements using a variety of standard or ad-hoc models. However, reliable querying and information retrieval requires consistency not only at the format level (e.g., Clinical Document Architecture (CDA), FHIR, HL7 V2) but also the content model (i.e., the information model such as HL7 CIMI model, or Observational Medical Outcomes Partnership (OMOP)), and, finally, the semantic and terminology model. There is not only a potential for a lack of consistency with representing disparate health data with current data models efforts but also further variation in how the data are entered into information systems by end-users. These differences pose challenges for how the data are modeled and stored, thereby generating implications on data retrieval, data analysis, and accuracy of clinical analysis results. For these reasons, the ability to test for equivalence across multiple data points is important. A well-defined query that yields accurate results in one health system is not guaranteed to be successful in another system because of a difference in the underlying data models. If normalization can occur at the data level, time and resources can be saved through sharing queries across different systems to achieve data fluidity and the ability to build layers of data from different sources.

Organizations like the National COVID Cohort Collaborative (N3C), Sentinel, Observational Health Data Sciences and Informatics (OHDSI) and others prove it is possible to develop, support and use real-world data (RWD) at “scale” across the U.S. and internationally in research, public health, and clinical quality improvement. However, the highlighted limitations provide the opportunity to address gaps commonly observed across efforts such as data quality and interoperability and access to timely data, while replicating the successes.

A specific, yet generalizable, use case demonstrates the need for real-world evidence (RWE) analytics, identify gaps, and form the basis of requirements necessary to establish a health data ecosystem capable for widespread collective efforts, public health surveillance, research, and care delivery. The long COVID use case describes the need to broadly survey and analyze large data sets. Doing so helps identify and define poorly understood health conditions for which new and/or existing regulated technologies or substances can be evaluated for safety and efficacy. RWD needs to be collected, queried, and evaluated to test and validate the “hypothesis” of the regulatory exercise.

## 2.2. Long COVID Use Case: Loosely Defined and/or Poorly Understood Scenarios

Long COVID is a poorly understood sequela to a COVID infection. In general, it is characterized by a wide range of ongoing health symptoms lasting weeks or months following a COVID infection, including fatigue, post-exertional malaise, fever, respiratory symptoms, neurological effects, and more. Even the CDC’s definition notes that the symptom list is not comprehensive and that post-COVID conditions may vary in symptom type, effect, and timespan, demonstrating how incomplete our medical understanding remains. [1] The NIH has stood up the Researching COVID to Enhance Recovery (RECOVER) Initiative, allocating $1.15 billion to better understand the development and symptoms of long COVID. [2] Similarly, the U.K. has invested over £50 million into long COVID research initiatives. [3] Countless other countries

and research institutions are doing the same as evidence mounts demonstrating the significant public health risk of this largely undefined condition.

Our incomplete understanding of long COVID is akin to the Human Immunodeficiency Virus (HIV) / Acquired Immunodeficiency Syndrome (AIDS) epidemic of the 1980s and 1990s where the differentiation between HIV and AIDS was not yet known. Only after substantial research brought the proper discovery and understanding of HIV infection and AIDS, effective treatments and secondary preventative treatments developed. [4] Likewise, Alzheimer's disease and related dementias (RD) were once considered syndromic; that is, poorly understood and only clinically characterized. After further research, we now recognize Alzheimer's as a biologically diagnosed condition - a definitive diagnosis that is determined postmortem only. [5]

The use of RWD resources has shown promise for hard-to-diagnose conditions. For example, Rheumatoid Arthritis related interstitial lung disease (RA-ILD) is a difficult diagnosis often only diagnosed reliably by rheumatologists in conjunction with substantial diagnostic work-up. However, studies at the VA and UNMC PCORnet and VA datasets demonstrate that commonly discrete data elements in the EHR can be used to identify RA-ILD patients based on their data phenotype. With the addition of low-level, basic Natural Language Processing (NLP) for radiology report evaluations, high levels of Positive Predictive Value (PPV) identification of RA-ILD patient by data phenotype is possible.

To address the long COVID use case, the FDA must have access to substantial amounts of data for patients positively diagnosed for COVID, for patients who exhibit identified long COVID symptoms (with and without a positive COVID diagnosis), and patients who have had no known COVID infection(s) and do not exhibit long COVID symptoms. These data need to be complete and have all appropriate data elements coded with standard values to enable meaning to be determined across facilities. Complete, standardized patient data across time and facilities will allow researchers to access symptoms, existing morbidities, demographics (including weight and age), treatments taken (for COVID or other conditions), vaccinations received, screening and diagnostic test data, and potentially other information like drug use, alcohol use, and smoking. This will enable the use of statistical methods and advanced analytics to quickly determine the relevant data elements and values necessary to more accurately establish a clinical definition for diagnosing an emerging condition like long COVID, along with identifying risk factors for contracting it, and provide evidence to develop tests and drugs for confirming and treating the condition.

In addition to "validated" data, such as existing laboratory and other diagnostic test data, provider-level patient reports and physician interpretation data must be collected. Furthermore, patient-reported, non-physician translated data should be considered, collected, and incorporated with the more clinically oriented data. Existing RWD ecosystems could add value to long COVID research. For example, the Sentinel system, which collects large quantities of claims records, could assist in identifying patient cohorts or determining secondary variables that may indicate long COVID symptoms. However, the known limitations of systems like Sentinel must be considered, such as potential incompleteness, poor association to actual patient outcomes, or diagnosis oversimplification within billing applications.

There are several conditions that must be considered to effectively address misunderstood conditions like long COVID with RWD. Though it may be incomplete, there needs to be a baseline description of the condition to identify variables and data sources of interest. This definition may be treated as fluid and updated as research progresses, so there must be adequate agility to evaluate, collect, and integrate new data elements inclusive of existing health care as well as novel observations. As with all varied data collections, an emphasis on standardization and normalization is key to achieve interoperability. In the capture process, non-discrete data will need to be harvested and converted to discrete data. Effective analysis will rely on the ability to query the data in multiple ways with multiple parameters, both within and across care sites.

As long COVID continues to grow as a public health priority, a proactive RWD research approach could provide valuable insights into the condition that may be difficult to discern through other research methods. To effectively understand and treat long COVID and other poorly understood diseases, researchers must be able to differentiate case and control patients with high reliability to then test the efficacy and safety of

new diagnostic and treatment options. RWD may have considerable potential in clearly identifying patient cohorts to advance understanding of the disease and develop effective treatments.

# 2.3. References


1. CDC. Post-COVID Conditions [Internet]. Centers for Disease Control and Prevention. 2022. Available from: https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html#:~:text=People%20with%20post%2DCOVID%20conditions%20(or%20long%20COVID)%20may

2. NIH launches new initiative to study “Long COVID” [Internet]. National Institutes of Health (NIH). 2021. Available from: https://www.nih.gov/about-nih/who-we-are/nih-director/statements/nih-launches-new-initiative-study-long COVID

3. Researching long COVID | NIHR [Internet]. www.nihr.ac.uk. Available from: https://www.nihr.ac.uk/covid-19/researching-the-long-term-impact.htm

4. Deeks SG, Overbaugh J, Phillips A, Buchbinder S. HIV infection. Nature Reviews Disease Primers. 2015 Oct 1;1(1).

5. Karantzoulis S, Galvin JE. Distinguishing Alzheimer’s disease from other major forms of dementia. Expert review of neurotherapeutics [Internet]. 2011;11(11):1579–91. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22014137

# 3. Use Cases

## 3.1. Notice

The Use Case chapter currently consists of a draft outline. In the coming year, our team will develop and refine content that can be incorporated.

## 3.2. Use Case: Komet Application

This section outlines three key components of the Komet application, Tinkar knowledge architecure, and open-source contribtuions.

### 3.2.1. Komet Use Case – Rapid Knowledge Management

A primary benefit of Komet and the Tinkar architecture used to manage integrated knowledge is the ability to rapidly update integrated knowledge in accordance with emerging or needed concepts and to address irregularities or errors.

Komet is capable of not only identifying errors but also resolving them. For example, if you searched for “chronic lung disease” and opened the “Chronic lung disease (disorder)” that takes an user to the concept navigator. The user may notice that this concept only displays two children concepts, which seems incomplete as there should be many more children concepts for such a broad disorder. To explore and understand why this potential error is occurring, the user then could click on both the ‘Properties’ and ‘Timeline’ buttons to open additional information about the “Chronic lung disease (disorder)” concept. In the ‘Properties’ frame, the user could select ‘Hierarchy’ and turn the ‘Range’ option on in the Timeline frame. Doing so will display how the concept has changed over time, additions to concepts, and allows users to compare select versions of a concept.

While a user may see version history, there is still no information on why there are only two children concepts. The user can then go back to the concept navigator and edit the EL++ Axiom from ‘Necessary Set’ to ‘Sufficient Set’, select reasoner, and run the ‘ELK Reasoner’. This then shows the recent edit and update, which shows the increased number of children concepts to over 20.

While Komet, Tinkar, and our open-source approach allow the user to identify an error by navigating the site, understanding the concept versioning and history, making edits to resolve the issue in real-time, and committing the changes to the community at large, other Terminology Standards lack in coordinating this rapid approach. If an user identified this error and submitted it to SNOMED CT®, they might have to wait a full six months before SNOMED CT® pushed their bi-annual updates and might be forced to employ sub-optimal work arounds to address the issue until then. This real-time approach of managing integrated knowledge management also allows the creation of concepts, such as COVID-19, to meet emerging conditions and clinical needs.

### 3.2.2. Komet Use Case – Integration of Terminology Standards

Another fundamental component of Komet is the integration of various terminology standards. When searching “urine homocysteine measurement”, a user can see the SNOMED CT® concept ‘Urine Homocysteine Measurement (procedure)’ in the navigation panel, while simultaneously viewing LOINC® children concepts. Komet can seamlessly integrate concepts from different terminology standards and identify which concepts are associated by using a localized common language in the form of Analysis Normal Form (ANF).

### 3.2.3. Komet Use Case – Detecting Equivalent Concepts

Komet is also able to easily and clearly identify concepts that are equivalent and display any minor differences between them. If an user were to search for a concept to describe ankylosis in the left and right knees, they would be presented with ‘Observation of ankylosis of both knee joints (procedure)’ and ‘Ankylosis of bilateral knee joints (disorder)’. By running the Komet reasoner, selecting detect equivalencies, right clicking these two concepts, and selecting ‘Compare Concepts’, a user could determine that these concepts are equivalent. A user could then do a deep-dive into the concept axioms using a side-by-side comparison to understand if there were any minute differences between the two.

## 3.3. Use Case: Considerations for Real World Data Interoperability and Data Agility in Patient Journey to Diagnosis for Emerging Acute and Chronic Infectious Diseases

### 3.3.1. Introduction to the Emerging Acute and Chronic Infectious Diseases Use Case

Increased generation of real-world data (RWD) and real-world evidence (RWE) from personal health devices, electronic heath records (EHR), and other sources has fueled a new wave of advanced analytics and has proven to be critical when evaluating emerging health trends. [1] Digital health devices, over-the-counter (OTC) diagnostic tests, and other technologies are important sources of RWD and can be used as decision support tools. However, interoperability issues including challenges with data transmission, storage, retrieval, and aggregation must be addressed to support safe and effective use of the data. RWE can drive important insights into public health and predict emerging diseases, pandemics such as SARS-CoV-2 (COVID-19), and other population health trends. Timely access to interoperable data is critical when developing guidance and improving patient care and outcomes.

Clinical workflows are highly cited as rich sources for analysis of clinical decision-making support and process improvement. Clinical process modeling and workflows can be used to understand a clinician’s reasoning, understanding, and use of patient data. Workflows capture the thought process behind decision making, help visualize the process, and can be considered as a decision tree. [2] Certain workflows can be used to understand the clinical data life cycle. In Vankipuram et al., the authors use a clinical workflow analytics framework using radio-frequency identification (RFID) to identify three phases of data flow: transmission, analysis, and transformation. [3] In these studies, the authors highlight the benefits and potential use of analyzing workflows to improve patient care or enhance clinical decision support tools. Other workflow analyses incorporate predictive modeling to analyze and predict most efficient operating models that drive effective interventions. [4] Advanced analytics including artificial intelligence and other technologies are being explored in workflows. [5] However, advanced analytical outcomes are fully dependent on the data feeding into the system.

Business Process Modeling and Notation (BPMN) is a diagram technique used to represent a flowchart of activities and decision processes and is beneficial to healthcare organizations seeking to implement process improvements. BPMN is a tool organizations use to evaluate clinical decision support systems, model clinical treatment protocols, and easily relay complex processes. BPMN is often used to model organizational or task related processes, but its use can be broadened to understand systemic flows of information. [6-8] As Kassim et al. point out in their systematic review, most BPMN focuses on clinical decision support but stops short of modeling patient healthcare trajectories and, with that, health information. [9]

Existing approaches of clinical process modeling analyze the data but fail to explore data considerations such as data capture and transmission. The need for interoperable data is evident. In previous health emergencies, important sources of data were identified too late, and the use of those data was further delayed by incompatibilities in interoperability. Preparing federal agencies and public health authorities with RWE tools will facilitate early identification of potential pandemics and diseases and allow timely and well-informed decision making. To reach this level of preparedness, it is critical to understand where potential data are generated, stored, and transmitted in a clinical workflow.

## 3.3.2. Aims for the Emerging Acute and Chronic Infectious Diseases Use Case

In this body of work, we aim to demonstrate the importance of an interoperable data system and downstream analytical scenarios through a use case for Long COVID. We propose that the methods and findings are not solely applicable to a use case for Long COVID and are reproducible for any emerging acute and chronic infectious disease use case. This use case presents a Long COVID workflow highlighting a patient's journey from first onset of COVID-19 symptoms through a Long COVID diagnosis and describes how data are captured, transmitted, and aggregated in a way that can be generalized to other emerging acute and chronic infectious diseases. Through understanding of data generation, transmission, and analysis we can identify important information and trends and establish an ideal future state for the use of RWE.

*Aim 1: Document a Current State Long COVID Workflow Depicting Data Loss*

*Aim 2: Establish the Future State of Patient Care*

*Aim 3: Establish the Future State Data Workflow*

## 3.3.3. Emerging Acute and Chronic Infectious Diseases Use Case Background

The COVID-19 pandemic drastically impacted healthcare, laboratory operations, pharmaceuticals, and patient care delivery. While extensive resources have been allocated to studying COVID-19 in support of vaccine development, public health policies, and treatment, the long-term symptoms and impacts of COVID-19 are not well known. Long COVID is a poorly understood condition associated with lasting health problems that persist past the typical symptomatic period after an initial COVID-19 infection. While Long COVID is typically marked by longitudinal fatigue, respiratory issues, and brain fog, there is a growing pool of symptoms and complications. Health conditions, duration, and even test results can vary across Long COVID cases, leading to an inability to quickly diagnose Long COVID. Identifying Long COVID is further complicated by a variety of inaccurate, lost, or suboptimal data collection and reporting. Due to these difficulties, there is no diagnostic test available for Long COVID, and care providers often reach a Long COVID diagnosis by ruling out other diseases or disorders with similar symptoms. [10-14]

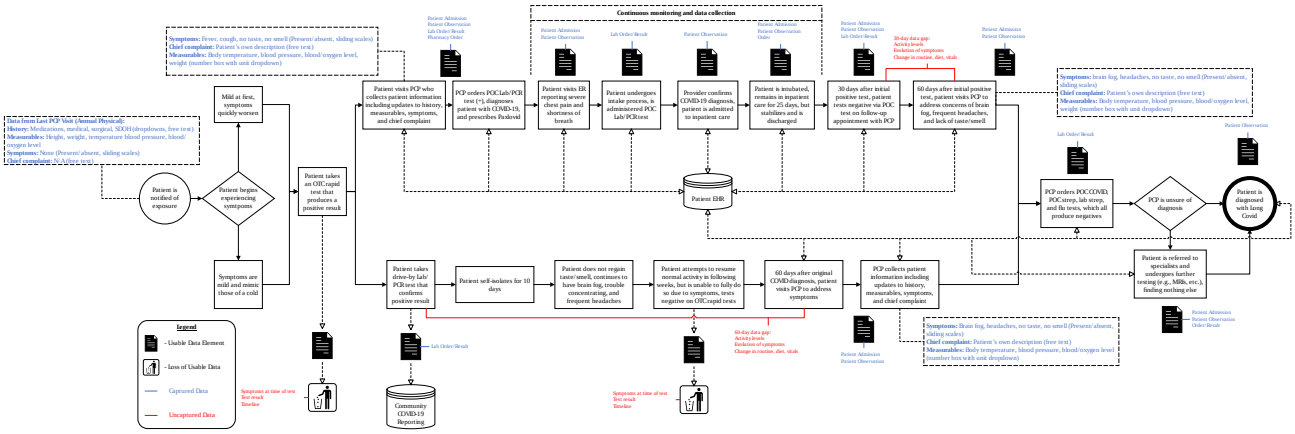
This Long COVID use case helps advance our understanding of the health data ecosystem and the necessary data capture and agility to ensure safe and effective patient care. This workflow broadly represents and can be applied to other emerging acute and chronic infectious diseases in that it highlights the data gaps and opportunities for use of RWE. We propose a future state workflow and proactive approach where data sets are analyzed and signals are detected before a condition is known, thereby yielding health information systems that are safe, effective, and interoperable.

# 3.3.4. Methods to Develop the Emerging Acute and Chronic Infectious Diseases Use Case

## 3.3.4.1. Aim 1: Document a Current State Long COVID Workflow Depicting Data Loss

The team developed the generalized Long COVID workflow depicted in Figure 1 highlighting the steps a patient takes throughout their diagnostic journey with accompanying data generated at each step. The team gathered foundational Long COVID research to establish a comprehensive list of widely accepted and important data elements and data entry points for the workflow. [10-13] A template form was created to gather input from the team members on these data elements including actors, descriptions, triggers, pre and post conditions, assumptions, and data fields, and eight team members independently documented considerations in the workflow. Team members had varying professional backgrounds including MD, NP, RN, Informaticist, Terminologists, HL7 Interface Engineers, Public Health, and Systems Engineering. The team met to discuss the exercise and highlight differing approaches in the team's documentation and considerations. Common themes were extracted and used in Visio to develop an initial draft of a generalized Long COVID use case workflow schematic that was iterated upon based on subject matter expert and key stakeholder input. The workflow follows an inpatient and outpatient path and identifies common steps in a patient's diagnostic journey along with associated data to highlight opportunities for more accurate and comprehensive data collection. Several important data elements were identified including patient symptoms, medical history, laboratory tests and orders, vaccination history, demographics, and prescriptions, and it was noted whether those data are currently collected or lost.

**Figure 3.1. Generalized Current State Long COVID Workflow**




## 3.3.4.2. Aim 2: Establish the Future State of Patient Care

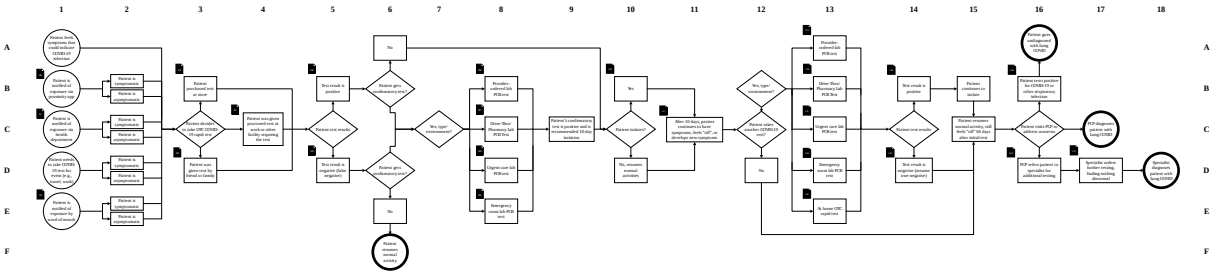
Once a standardized general Long COVID diagnostic workflow schematic was developed, the team honed in on the outpatient scenario to capture the patient journey in greater detail. The team chose to focus on an outpatient scenario as there is a larger propensity for data loss when the patient is not under the continuous monitoring of inpatient care. This new BPMN schematic closely examined the parallel paths that a patient in an outpatient setting could follow before arriving at a Long COVID diagnosis to support improved CDS, patient care, and data collection and usage. This resulted in an approximate current state patient workflow as well as an idealized future state patient workflow with a supplementary data flow schematic. Note that while these scenarios used Long COVID as an inspiration, the underlying principles can be applied to any similar emerging acute, chronic and infectious disease.

*Scenario 1: Current State Outpatient Long COVID Workflow*

Scenario 1 details an expanded view of the workflow as patients begin their journey to a Long COVID diagnosis through parallel paths. Step 1 demonstrates the different reasons that could prompt a patient to take a COVID-19 test such as notification of exposure from the health department, phone proximity

tracking, routine testing at work, prerequisite testing for travel, or feeling symptomatic. The parallel actions are listed for each step of the process in in the workflow schematic below in Figure 3.2, “Current State Outpatient Long COVID Workflow”. In this current state, there are several points at which data are produced but not collected and therefore lost.

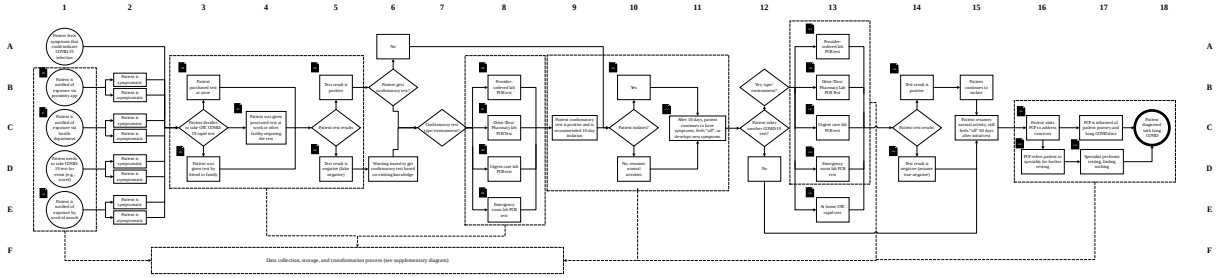
## Figure 3.2. Current State Outpatient Long COVID Workflow



*Scenario 2: Future State Outpatient Long COVID Workflow*

Scenario 2 explores an idealized future state for the outpatient workflow depicted in Figure 3.3, “Future State Outpatient Long COVID Workflow” that aims to minimize data loss. It should be noted that there are only subtle changes to the patient-centric workflow, but there are several points at which data are collected and stored, which, in turn, will enable more streamlined care and a more complete picture of health for the individual patient and the larger population. The collection, storage, and transmission of data are detailed in Aim 3.

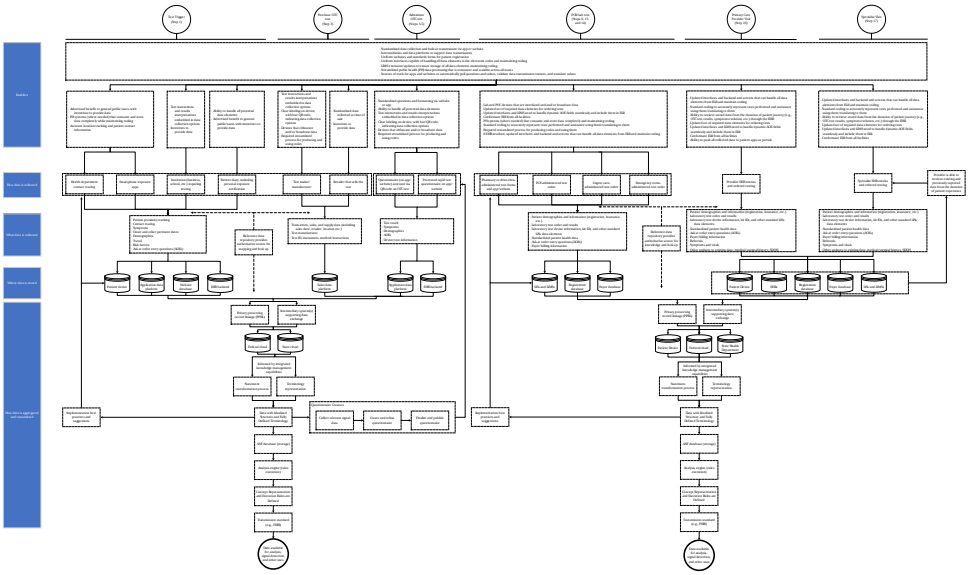
**Figure 3.3. Future State Outpatient Long COVID Workflow**




## 3.3.4.3. Aim 3: Establish the Future State Data Workflow

Building upon the Future State Outpatient Long COVID Workflow, the team worked with subject matter experts in public health reporting, laboratory data systems, and health informatics to identify points in the patient care workflow where data are captured, where data capture could be improved, and resources or technologies that could support the improved data capture. The team then developed a separate schematic (Figure 3.4, “Idealized Future State Data Long COVID Workflow”) to showcase an idealized flow of those data; the workflow is framed around an OTC COVID-19 test trigger and showcases how data should be captured, transmitted, stored, and prepared for downstream analytical and secondary use.

**Figure 3.4. Idealized Future State Data Long COVID Workflow**



An important consideration when developing this workflow was the inclusion of enablers. Enablers serve to drive the best future state collection and use of the data that are generated in the future state patient care workflow starting with empowering or incentivizing the patient to report OTC test results and stretching to accurate reporting of point of care (POC) orders and results. The team expanded upon what data are collected and where the data are stored, leading to a detailed schematic of how data are aggregated, transmitted, and ultimately used.

Privacy Preserving Record Linkage (PPRL), a way to maintain patient privacy while collecting protected patient data and preserving data meaning across data sets, was included as an important step between data storage and aggregation and final use by the federal government. [16] Following PPRL, a loop was included in the data aggregation step “Data with Idealized Structure and Fully Defined Terminology” to include an opportunity for feedback and continual improvement of data collection practices and standards. This RWE and analytic-based feedback also informs questionnaire creators as they deploy updated questionnaires back to the original data collectors and serves as an opportunity to implement best practices and suggestions for further exploration. These loops are included to showcase how data insights can provide real-time feedback and identify opportunities for continual process improvements. The future state data workflow concludes with transformed, normalized, and de-identified data in repositories ready for secondary use across a variety of systems depending on content and purpose.

# 3.3.5. Results from the Emerging Acute and Chronic Infectious Diseases Use Case

## 3.3.5.1. Aim 1: Document a Current State Long COVID Workflow Depicting Data Loss

The team developed Figure 3.1, “Generalized Current State Long COVID Workflow”, a generalized workflow that outlines a patient’s journey to a Long COVID diagnosis by aggregating common themes, data elements, and components from the information that team members individually collected. In the current state, critical data associated with OTC testing, patient symptoms, symptom progression, and timelines are either lost or sub-optimally reported. Accurate and standardized data reporting at all stages of a patient’s

diagnosis is needed to support improved patient care, research, and public health and policy decision making.

*Outpatient Path*

A first instance of data loss occurs in the outpatient path when the patient takes the first at-home OTC test; test results, symptoms at time of testing, and timeline are uncaptured. However, lab orders and results are captured during patient testing at the drive-up testing site and results are reported to the "Community COVID-19 Reporting" database. Following the drive-up testing, there is a 60-day window of missed opportunity for critical capture of symptoms and at-home testing results, including timeline and progression data. During this 60-day period the patient experiences evolving symptoms and changes in activity levels, routines, and diet. Data are only reported again when the patient visits a primary care provider (PCP) for a follow-up visit, indicating the need for gathering outpatient data generated by the patient.

*Inpatient Path*

The inpatient path indicates a better environment for data capture. Similar to the outpatient workflow, initial OTC test results, symptoms, and timeline are not captured as the patient does not report these data. However, the patient enters an inpatient workflow environment where all data generated through the PCP, Lab, and Emergency Room (ER) are documented in the patient's EHR during a period of continuous monitoring. In this inpatient scenario there is only one point of data loss during a period of 30-days between testing negative and returning for subsequent testing after experiencing continued symptoms. During this period, data regarding activity levels, evolution of symptoms, and change in routine, diet, and vitals are lost.

The outpatient path eventually leads into the inpatient workflow as both patients' PCP consider a Long COVID diagnosis. This general workflow highlights data loss in OTC testing or outpatient scenarios, and considerable loss of uncaptured symptom and test results data.

## 3.3.5.2. Aim 2: Establish the Future State of Patient Care

Once the generalized Long COVID use case was developed, the team explored a more specific scenario that begins with a prompt for a patient to take an OTC COVID-19 test. Current and standardized data points were identified in the representative patient flow and categorized as captured or uncaptured data. The workflow highlighted the major steps over the course of a patient's Long COVID diagnostic journey, including potential variations in results or outcomes at certain steps.

The workflow begins with one of five steps that lead to the patient pursuing an at-home, OTC COVID-19 test and exploring home treatment/quarantine decisions without clinical input or EHR data capture. Uncaptured data generated during these steps in the diagnostic journey include exposure methods and related data, symptom diary options, patient background, OTC POC test sales and supply data, results from the OTC POC test, and more.

Step eight, when the patient looks to confirm the results of their OTC test, is when the process reaches the first point of clinical input as the patient takes one of four potential Lab PCR tests. In the current state, data preceding step eight would be lost or unutilized. Data generated at this step include registration, intake information, continuation of symptoms, lab orders, and results. Standardized data include patient demographics, insurance information, updated lab order information (device and test kit IDs), and EHR standard data elements, depending on the location of the test (e.g., pharmacy, urgent care, PCP, ER).

The workflow continues with data collection from both POC and OTC sources, aiming to collect as much as possible along the way. It should also be noted that there will be bidirectional data flow within this ideal future state. For example, a physician will have access to patient data collected throughout their journey to support their clinical decision making and, ultimately, diagnosis. Aim 3 explains how data will be transformed and utilized for specific purposes.

## 3.3.5.3. Aim 3: Establish the Future State Data Workflow

Using the outpatient Long COVID patient care workflow, the specific steps where data are or should be captured were identified and collated into a table. Subject matter experts then identified opportunities for improved data capture through various resources, technologies, and methods. Table 1 is an example of four data capture points from the larger table that identifies specific steps in the outpatient Long COVID workflow, if and how data are currently captured at those steps, new data that need to be collected, and methods for collecting those data. The specific data collection points are denoted with a data icon in the outpatient Long COVID patient care workflow and similar data collection steps are grouped by a box with a dashed line. This process was repeated throughout the workflow for each data capture point.

### Table 3.1. Future State Data Collection Points

| Reference | Current Data | Captured? (Y/N) | Collection Methods | Needs for Collection | Future State Standard Data |
|---|---|---|---|---|---|
| 5B<br>5D | • Date and time of test<br>• Symptom data (maybe)<br>• Test result | N/A (unless proctored) | N/A | • Data capture options with built in transmission paths<br>• Intermediary for delivery of data to appropriate entities<br>• Clear labeling to support data options<br>• Instructions to perform test or interpret result embedded on data capture options<br>• Incentives to provide data<br>• Devices that broadcast data<br>• Devices that obfuscate results | • Lab data standards (manufacturer, test, device ID, etc.)<br>• Test result<br>• Symptom data<br>• Exposure source (if known) |
| 10B | None | N/A | N/A | • Symptom diary options | • Symptoms<br>• Dates |

| | | | | | |
|---|---|---|---|---|---|
| | | | | • Contact recording | • Contacts |
| 11C | None | N/A | N/A | • Symptom diary options<br>• Contact recording | • Symptoms<br>• Dates<br>• Contacts |

The data collection groups were then transposed and used as the starting points for the idealized future state data workflow depicted in table above. This workflow outlines the idealized flow of data through the five key data dimensions listed in chronological order:

Enablers: Enablers are key processes, steps, or components that must be achieved to improve the quality and standardization of data collection steps. The future state data workflow begins with overarching enablers, such as intermediaries and data platforms to support data transmissions, and then dives into more specific requirements for each of the data collection methods like implementing functional barcode/QR code systems on OTC test boxes.

How Data are Collected: The workflow then highlights the various data collection tools and resources that can be used to collect critical patient health data.

What Data are Collected: The workflow then explores the type of data that are collected. Depending on the associated step in the future state outpatient workflow, the various data collection methods could capture retail data, test data, patient symptoms, ask at order entry (AOE) questions, and much more.

Where Data are Stored: Once the type of data that are collected is identified, the workflow outlines the storage systems and databases that originally house the data.

How Data are Aggregated and Transmitted: After data flow out of the initial data store and through a PPRL, the data will undergo a statement transformation process and terminology representation process that is informed by knowledge management reference implementation and produces data with idealized structure and terminology. These data then continue through Analysis Normal Form (ANF) storage and rules execution and applies transmission standards such as FHIR. Data are then finally ready to be transmitted to other healthcare systems such as research repositories or EHRs. Data with idealized structure and terminology can also follow an alternate path and take part in feedback loops that inform questionnaire development for data acquisition from patients and providers or can be used as real-world evidence to inform improvements to the initial data collection methods.

While the data workflow uses a COVID-19 test trigger, the purchase of an OTC COVID-19 test, the administration and use of an OTC COVID-19 test, and administration of a confirmatory laboratory PCR test, this process can be expanded to more data collection steps along a Long COVID diagnosis or even expanded to other emerging acute and chronic infectious diseases. This data workflow shows how data will need to move within and between EHRs, laboratory information systems (LIS), federal databases, website and application data stores, and other healthcare related systems and how RWE can be used as a feedback mechanism to support continual process and data collection improvements.

## 3.3.6. Emerging Acute and Chronic Infectious Diseases Use Case Discussion

In today’s healthcare ecosystem and expanding digital and personal health markets, large amounts of patient data are generated but uncaptured. This body of work aims to highlight some of the gaps in data collection and proposes potential opportunities to use existing technologies and resources to improve the quality and interoperability of data that are collected from previously uncaptured sources. A generalized

Long COVID and more granular workflows were developed as a use case to highlight specific instances of data loss and opportunities to improve collection of data across a variety of emerging acute and chronic infectious diseases that is critical to patient care, such as at home test results. Long COVID, a relatively new and poorly characterized condition, acts as a use case to demonstrate how capturing this additional data would allow providers, researchers, and many others involved in the healthcare system to make better informed patient care decisions and develop evidence-based indicators for a Long COVID diagnosis. Improving data interoperability will improve an organization's ability to ingest data in real-time and develop materials such as questionnaires or symptom trackers to deploy to the field. Targeted and dynamic data collection methods will give organizations better insight into the most pertinent questions that need to be answered for a variety of emerging conditions and diseases.

Our analysis of the Long COVID workflow uncovered several sources of RWD where extra effort should be made to gather and represent relevant data. RWD, including patient reported symptoms, are an opportunity to gather information that could lead to earlier identification of diseases. When considering the expanding market for OTC testing for many chronic and infectious diseases, it is evident the amount of critical data that are lost from the system and unavailable for further analyses.

## 3.3.7. Future Considerations for the Emerging Acute and Chronic Infectious Diseases Use Case

The Long COVID data workflow outlines idealized data stemming from a COVID-19 OTC test trigger, the purchase and use of a COVID-19 OTC test, and the administration of a confirmatory laboratory PCR test. These are all key data collection points in the Future State Long COVID patient workflow but do not comprehensively represent how data could flow in an idealized state. The team will continue to incorporate the remaining data collection steps and data groupings from the patient workflow to elucidate how various data storage systems house, transform, and transmit data as well as the enablers needed to improve and standardize the associated data collection. This continued work will be included in a future iteration of this work. While this work currently centers around a Long COVID diagnosis, the themes and insights will be used to inform data aggregation and secondary use for other emerging acute and chronic infectious diseases that will ultimately provide their own feedback and insights to further improve data collection.

Advanced analytical capability of monitoring and responding to data trends in real time goes beyond detecting emerging diseases and can extend to fast tracking device and diagnostic tool approvals. Liquid Biopsy (LBx) is a promising new diagnostic tool for screening and detecting cancers, however, traditional clinical trials require too many resources and cannot evaluate the breadth of emerging LBx assays. Aggregated patient data are needed along with timely data from EHRs, LIS, tumor registries, and other data repositories for appropriate evaluation. Many different RWD elements and classes are required to establish the sensitivity and specificity of a LBx, including data on a range of cancer negative/positive patients, lab results, pre/post cancer testing, past medical and family histories, medication history, environmental exposure, germline sequences, imaging studies, and demographics. Further understanding the sources of data can help inform future data flows similar to the Long COVID flow to create normalized, interoperable, and dynamic data repositories.

# 3.4. References

1. FDA. Real World Evidence [Internet]. U.S. Food and Drug Administration. 2022. Available from: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence

2. Cook RJ, Durning SJ. Clinical Process Modeling. Academic Medicine. 2019 Apr;1.

3. Vankipuram A, Traub S, Patel VL. A method for the analysis and visualization of clinical workflow in dynamic environments. Journal of Biomedical Informatics. 2018 Mar;79:20–31.

4. Crowley C, Guitron S, Son J, Pianykh OS. Modeling workflows: Identifying the most predictive features in healthcare operational processes. Vellido A, editor. PLOS ONE. 2020 Jun 11;15(6):e0233810.

5. Griewing S, Lingenfelder M, Wagner U, Gremke N. Use Case Evaluation and Digital Workflow of Breast Cancer Care by Artificial Intelligence and Blockchain Technology Application. Healthcare. 2022 Oct 20;10(10):2100.

6. de Bruin JS, Adlassnig K-P, Leitich H, Rappelsberger A. Separating Business Logic from Medical Knowledge in Digital Clinical Workflows Using Business Process Model and Notation and Arden Syntax. Studies in Health Technology and Informatics [Internet]. 2018 [cited 2023 Mar 15];248:17–24. Available from: https://pubmed.ncbi.nlm.nih.gov/29726414/

7. Natsiavas P, Stavropoulos TG, Pliatsios A, Karanikas H, Gavriilidis GI, Dimitriadis VK, et al. Using Business Process Management Notation to Model Therapeutic Prescription Protocols: The PrescIT Approach. Studies in Health Technology and Informatics. 2021 May 27;

8. Ruiz F, Garcia F, Calahorra L, Llorente C, Gonçalves L, Daniel C, et al. Business process modeling in healthcare. Studies in Health Technology and Informatics [Internet]. 2012 [cited 2023 Mar 15];179:75–87. Available from: https://pubmed.ncbi.nlm.nih.gov/22925789/

9. Kassim SA, Gartner J-B, Labbé L, Landa P, Paquet C, Bergeron F, et al. Benefits and limitations of business process model notation in modelling patient healthcare trajectory: a scoping review protocol. BMJ Open. 2022 May;12(5):e060357.

10. Hedin K, van der Velden AW, Hansen MP, Moberg AB, Balan A, Bruno P, et al. Initial symptoms and three months follow-up after acute COVID-19 in outpatients: An international prospective cohort study. The European Journal of General Practice [Internet]. 2023 Jan 19 [cited 2023 Mar 9];2154074. Available from: https://pubmed.ncbi.nlm.nih.gov/36655704/

11. Tran V-T, Porcher R, Pane I, Ravaud P. Course of post COVID-19 disease symptoms over time in the ComPaRe long COVID prospective e-cohort. Nature Communications. 2022 Apr 5;13(1).

12. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. Nature Reviews Microbiology [Internet]. 2023 Jan 13;1–14. Available from: https://www.nature.com/articles/s41579-022-00846-2

13. Malden DE, Tartof SY, Ackerson BK, Hong V, Skarbinski J, Yau V, et al. Natural Language Processing for Improved Characterization of COVID-19 Symptoms: Observational Study of 350,000 Patients in a Large Integrated Health Care System. JMIR Public Health and Surveillance [Internet]. 2022 Dec 30 [cited 2023 Mar 9];8(12):e41529. Available from:https://publichealth.jmir.org/2022/12/e41529

14. CDC. Long COVID or Post-COVID Conditions [Internet]. Centers for Disease Control and Prevention. 2021. Available from: https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html

15. Cholan RA, Pappas G, Rehwoldt G, Sills AK, Korte ED, Appleton IK, et al. Encoding laboratory testing data: case studies of the national implementation of HHS requirements and related standards in five laboratories. Journal of the American Medical Informatics Association [Internet]. 2022 May 25 [cited 2023 Mar 16];29(8):1372–80. Available from: https://academic.oup.com/jamia/article/29/8/1372/6592172#365931904

16. PPRL | N3C [Internet]. covid.cd2h.org. [cited 2023 Mar 9]. Available from: https://covid.cd2h.org/PPRL