



PHUSE/FDA Innovation Challenge

Pillar 2

Sarah Ferko,
Abhishek Pandey
IBM
28FEB2024



phuse.global



Contributors

Presenters



Sarah Ferko
Project Manager & Senior
Managing Consultant



Abhishek Pandey
Healthcare Strategy &
Managing Consultant



Matt Deady
Senior Data Scientist



Brian Goodness
Data Scientist

Pillar 2 Presentation

1

- Risk-based monitoring, throughout all stages of the clinical investigation, managing both risks to participants (e.g., a safety problem) and to data integrity (e.g., incomplete and/or inaccurate data)

2

- Development across clinical trial activities, such as data collection, data comparison, and/or data analysis practices



Agenda

- Introduction
- EHR Data Quality Study
 - Methods
 - Known EHR Data Quality Challenges
- AI Patient Summarizer Application

Introduction



Use of EHR data for clinical trials

Pharmaceutical companies can **utilize EHR data to enhance their data sources** within a clinical trial. For example, RWD can provide:

A comprehensive view of health history of patients from multiple sources

Study design considerations

A historical control arm or data for comparative effectiveness research

Data for long-term safety monitoring of products to support post-marketing commitments

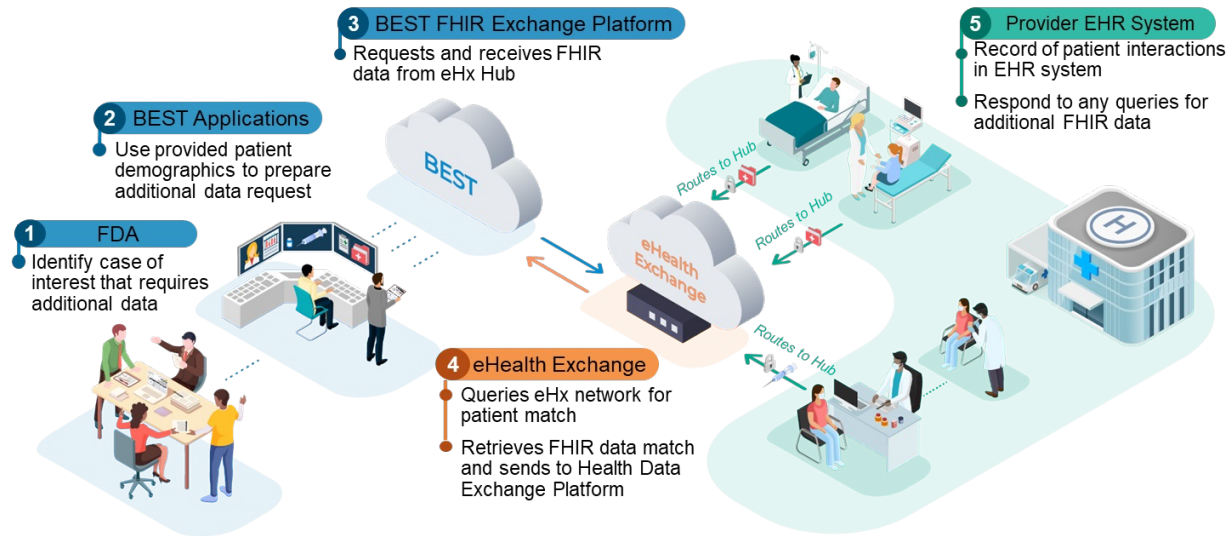
A diverse patient population

Identification of biomarkers, other prognostic indicators, and patient baseline characteristics



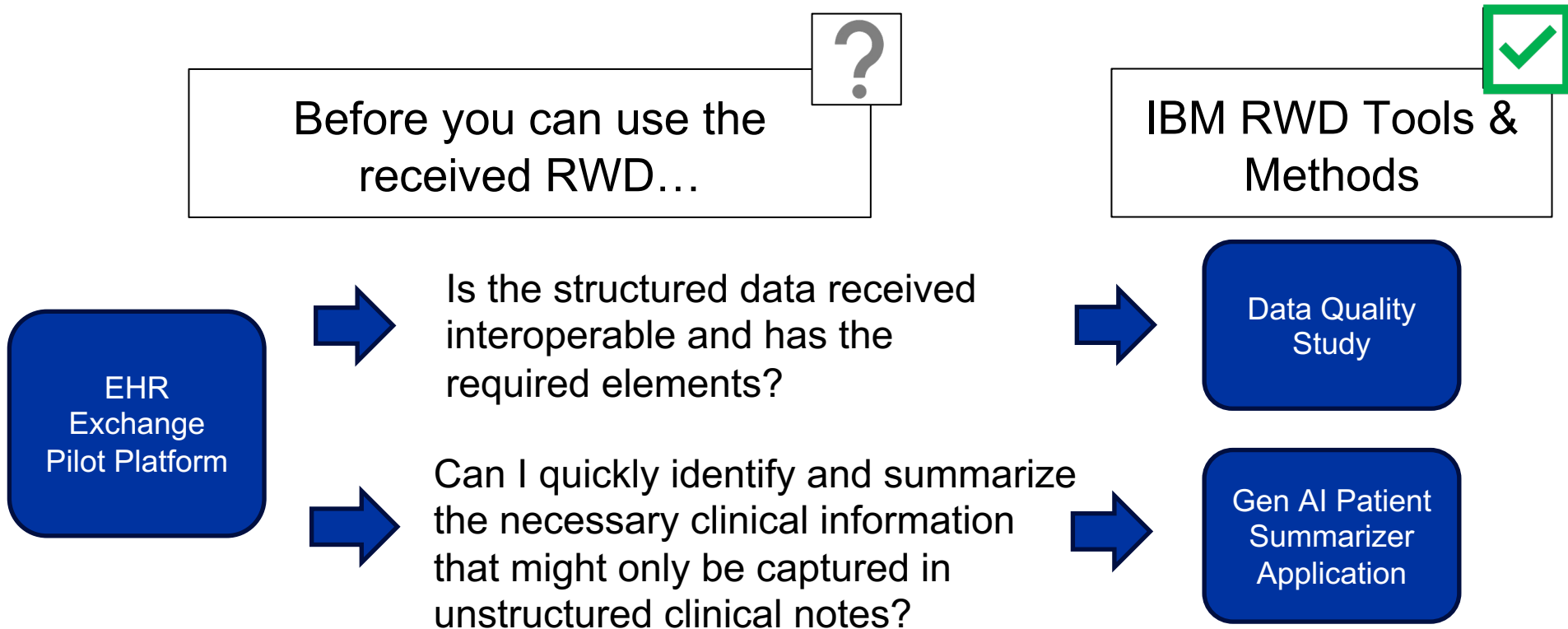
RWD Experience through FDA EHR Exchange Pilot Platform

As part of its work for the U.S. Food and Drug Administration (FDA)'s Biologics Effectiveness and Safety (BEST) initiative, IBM built a pilot platform that allows the FDA to receive real world data (RWD) from Electronic Health Records (EHRs) to improve their post-market surveillance capabilities.





Lessons Learned: AI Tools and Methods Can Help Prepare RWD For Use





EHR RWD Exchanged

- All RWD received through the platform is sent by standardized FHIR APIs made available by the health provider.
- For any health provider that hopes to be certified by ONC (Office of the National Coordinator for Health Information Technology), the ONC requires a set of data to be made available to request through API. This set of data is described by:
 - **USCDI:** A standardized set of data elements required for sharing information between health IT systems
 - **U.S. Core:** Provide detailed technical specifications and implementation guidance to support the adoption of a version of the USCDI.

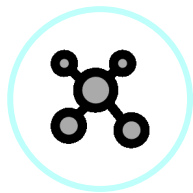
USCDI	US Core
V1	3.1.1*
V1	4.0.0
V2	5.0.1
V3	6.1.0

*The Minimum US Core Version is currently 3.1.

EHR Data Quality Study Methods



Data Quality Study



Primary Goal

Understand Data Quality of Provider FHIR Data

Data quality analysis to understand if the HL7 FHIR data received, via an Exchange network, can improve clinical trials

- An **automated data quality analysis** could be developed to review (and re-review) patient HL7 FHIR data received from healthcare provider's FHIR API endpoints to assess suitability for clinical trial submissions
- An **industry standard data quality framework**, such as the one suggested by Kahn, et. Al., is valuable for evaluating the quality of clinical trial data

Data Quality Study: Methods

Kahn, et. al. Data Quality Framework
(Implemented by OHDSI DQD)



Conformance

Do data values adhere to specified standards and formats?

Example: Are lab test resources coded using required interoperable code systems?



Completeness

Are variables present and do they contain all recorded values?

Example: Are diagnosis codes and abatement dates captured for all encounters?



Plausibility

Are data values believable?

Example: Do patients have realistic birth dates? Is the medication start date after the birth date?



Data Quality Resource Tests

- For this presentation, we filtered on a subset of data quality tests that matched or were similar to CDISC variables or data conformance rules (i.e., the published SDTM and SDTMIG Conformance Rules v2.0) that are needed for clinical trial submission to the FDA.

Data Quality Tests for Clinical Trial Submission

FHIR Resource	Maps to CDISC Data	Count of Tests		
		Completeness	Conformance	Plausibility
Condition / Document Reference	Adverse Event Details	4	2	8
MedicationRequest	Drug or concurrent medications	5	1	4
Observation	Lab test results	4	2	4
Patient	Patient Demographics	6	5	3



Known EHR Data Quality Challenges



Missing Medication Compliance Data

- **Medication data is only available for prescription orders**, which gives the prescribed dosage timing and amount but does not confirm adherence
- **Future versions** of U.S. Core IG **plan to require medication adherence**
- Currently, information on **medication adherence** may be found in **unstructured clinical note data**



Data Element Not Required by USCDI

- **Some CDISC data elements are not required** to be available for exchange because they are **not part of USCDI's required core data set**
- Fortunately, **EHR vendors often support data elements beyond the bare minimum**



Use of Locally Defined, Interoperable Code Systems / Value Sets

- **Semantic interoperability**, when data are not only exchanged between two systems but also able to be understood by both, is **still a challenge for some elements**
- The U.S. Core IG **allows for codes or values that are locally defined by an EHR site** instead of **standard values** set by the specification which helps to **ease the burden of implementation**, but **decreases the semantic interoperability**



Data Quality Study: Further Work

- To apply the EHR data quality checks to a clinical trial submitted in support of an NDA/BLA, **additional work is needed so that the checks for conformance, completeness, and plausibility are expanded beyond the original domain areas selected and tested for this pilot***.
 - This includes **additional plausibility checks** relevant to each domain area.
- Checks for **variables unique to clinical trial data** would need to be **added**
 - **Example:** Study administrative variables that are assigned to each subject such as planned arm and actual arm

** Demographics, Adverse Events, Laboratory Results, and Exposure/Concomitant Medications*

Generative AI Patient Summarizer



Generative AI Patient Summarizer

- **High data volume per patient case** may lead to issues in clinical analysis of the received EHR data
 - Especially analysis of **unstructured text of clinical notes**
 - Unstructured text of clinical notes which are estimated to be **~80%** of all clinical data
- **Generative AI Use Case:** We use AI to extract pertinent evidence and summarize all the data (structured and unstructured) in a patient case



Generative AI Patient Summarizer

- Investigate through implementation of the **Retrieval Augmented Generation (RAG) AI design pattern** which allows it to deliver insights from large corpuses of unstructured text using semantic search and generative AI.
- **Research Addresses Key Gen AI Challenges**
 - How to leverage GenAI **securely**?
 - How to trust generated output through references and **explainability**?
 - How to apply GenAI to an **agency knowledge base and sensitive/proprietary data**?

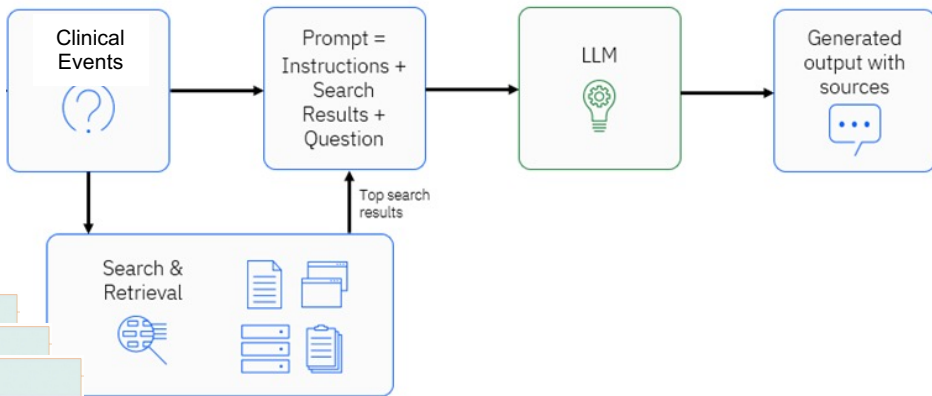




Our Retrieval Augmented Generation App

“Clinical Events” include allergies, medical conditions and medications

```
<patient> <url=http://hl7.org/fhir/>  
<id> <id="100"/>  
<meta>  
  <timestamp> <value="2017-01-04T11:40"/>  
</meta>  
<web>  
  <patient> <url=http://hl7.org/fhir/>  
  <id> <id="100"/>  
  <meta>  
    <timestamp> <value="2017-01-04T11:40"/>  
  </meta>  
  <test> <url=http://hl7.org/fhir/>  
  <id> <id="100"/>  
  <meta>  
    <timestamp> <value="2017-01-04T11:40"/>  
  </meta>  
</web>  
<test>  
  <id> <id="100"/>  
  <meta>  
    <timestamp> <value="2017-01-04T11:40"/>  
  </meta>  
</test>  
</patient>  
<patient> <url=http://hl7.org/fhir/>  
<id> <id="100"/>  
<meta>  
  <timestamp> <value="2017-01-04T11:40"/>  
</meta>  
<web>  
  <patient> <url=http://hl7.org/fhir/>  
  <id> <id="100"/>  
  <meta>  
    <timestamp> <value="2017-01-04T11:40"/>  
  </meta>  
  <test> <url=http://hl7.org/fhir/>  
  <id> <id="100"/>  
  <meta>  
    <timestamp> <value="2017-01-04T11:40"/>  
  </meta>  
</web>  
<test>  
  <id> <id="100"/>  
  <meta>  
    <timestamp> <value="2017-01-04T11:40"/>  
  </meta>  
</test>  
</patient>  
<patient> <url=http://hl7.org/fhir/>  
<id> <id="100"/>  
<meta>  
  <timestamp> <value="2017-01-04T11:40"/>  
</meta>  
<web>  
  <patient> <url=http://hl7.org/fhir/>  
  <id> <id="100"/>  
  <meta>  
    <timestamp> <value="2017-01-04T11:40"/>  
  </meta>  
  <test> <url=http://hl7.org/fhir/>  
  <id> <id="100"/>  
  <meta>  
    <timestamp> <value="2017-01-04T11:40"/>  
  </meta>  
</web>  
<test>  
  <id> <id="100"/>  
  <meta>  
    <timestamp> <value="2017-01-04T11:40"/>  
  </meta>  
</test>  
</patient>
```



Synthetic Patients' Clinical Note FHIR Data Files

Patient Summarizer Application: Custom application built to populate answers to whether evidence of clinical events exist in a patients note based on similarity scores and then LLM based retrieval.



Generative AI Patient Summarizer

Patient Summarizer

PATIENT LIST

 **Abdul218**
urn:uuid:b0a06ead-cc42-aa48-dad6-841d4aa679fa
Last Updated: 2002-02-08T00:00:00

 **Abe604**
urn:uuid:ccfc4db2-2026-7adb-3db0-33f3828140bb
Last Updated: 2021-09-26T00:00:00

 **Adelaida985**
urn:uuid:31a2e8ec-69fc-8a71-3ab6-36cbdd508713
Last Updated: 2017-02-21T00:00:00

 **Adriana394**
urn:uuid:71a8b156-760b-df6b-859e-eefc7932a526
Last Updated: 2021-11-06T00:00:00

 **Agueda283**
urn:uuid:76b289fd-e825-734c-8446-316f59643593
Last Updated: 2021-10-03T00:00:00

 **Ahmed109**
urn:uuid:92fb7efc-5cfd-f8d3-927b-42f8ee099531
Last Updated: 2021-06-

 Patient Record

 Summary

 **ABE604**
urn:uuid:ccfc4db2-2026-7adb-3db0-33f3828140bb

Found 38 Records

Encounter ID	urn:uuid:1ee1a688-7b07-470f-81b9-561b14c458fa
Practitioner Name	Dr. Mabel261 O'Kon634
Date	1994-07-31
CHIEF COMPLAINT	No complaints.
HISTORY OF PRESENT ILLNESS	Abe604 is a 18 year-old non-hispanic white male.
SOCIAL HISTORY	Patient is an active smoker and is an alcoholic. Patient identifies as heterosexual. Patient comes from a middle socioeconomic background. Patient has completed some college courses. Patient currently has Humana.
ALLERGIES	No Known Allergies.
MEDICATIONS	No Active Medications.



Generative AI Patient Summarizer

Patient Summarizer

PATIENT LIST

Abdul218
urn:uuid:b0a06ead-cc42-aa48-dad6-841d4aa679fa
Last Updated: 2002-02-08T00:00:00

Abe604
urn:uuid:c0fc4db2-2026-7adb-3db0-33f3828140bb
Last Updated: 2021-09-26T00:00:00

Adelaida985
urn:uuid:31a2e8ec-69fc-8a71-3ab6-36cbdd508713
Last Updated: 2017-02-21T00:00:00

Adriana394
urn:uuid:71a8b156-760b-df6b-859e-eefc7932a526
Last Updated: 2021-11-06T00:00:00

Agueda283
urn:uuid:76b289fd-e825-734c-8446-316f59643593
Last Updated: 2021-10-03T00:00:00

Ahmed109
urn:uuid:92fb7efc-5cfd-f8d3-927b-42f8ee099531
Last Updated: 2021-06-09T00:00:00

Patient Record

Summary

SUMMARY

The patient's name is Abe604 and he is a 45-year-old non-hispanic white male. He has a history of part-time employment, limited social contact, not being in the labor force, full-time employment, acute bronchitis, acute viral pharyngitis, first degree burn, social isolation, and stress. He is married, an active smoker, and an alcoholic. He identifies as heterosexual and comes from a middle socioeconomic background. The patient has completed some college courses and currently has UnitedHealthcare as his health insurance. He has been prescribed medications such as hydrochlorothiazide, fentanyl, and acetaminophen. The patient has also received immunizations for influenza, seasonal, and the COVID-19 vaccine. Procedures conducted include assessments of health and social care needs, depression screening, screening for substance abuse, and burn care.

ALLERGIES	No Known Allergies.
MEDICAL CONDITIONS	full-time employment limited social contact social isolation stress acute viral pharyngitis first degree burn acute bronchitis 1994-07-31 hypertension 1994-07-31 full-time employment 2014-06-23 acute viral pharyngitis 2014-11-04 Hyperlinks to the source data clinical note full-time employment limited social contact social isolation stress acute viral pharyngitis first degree burn acute bronchitis 1994-07-31 hypertension 1994-07-31 full-time employment 2014-06-23 acute viral pharyngitis 2014-11-04 Onset dates of conditions extracted when available



AI Patient Summarizer: Future Functionality

Future functionality can be added for Gen AI to answer questions about patients, with links to the clinical note where it finds its answer.

Example Application:
ONC guidance chatbot
links to summary
meeting minutes where
it finds the answer to
questions on FDA
regulatory guidance

The screenshot displays the official website of the U.S. Food & Drug Administration (FDA), specifically the Office of New Drugs (OND). The header features the FDA logo and navigation links. The main content area is titled "Office of New Drugs" and includes a "What We Do" section with a list of responsibilities. A sidebar on the left contains links to various CDER resources. A chatbot interface, labeled "FDA OND CHAT", is overlaid on the right side, showing a user query about the agenda of a meeting and the chatbot's response with a link to the summary minutes.

U.S. FOOD & DRUG ADMINISTRATION

← Home / About FDA / FDA Organization / Center for Drug Evaluation and Research | CDER / Office of New Drugs

Office of New Drugs

Share Tweet LinkedIn Email Print

Center for Drug Evaluation and Research | CDER

CDER Offices and Divisions

Meeting Presentations | Drugs

CDER Exclusivity Board

Accelerating Rare disease Cures (ARC) Program

CDER Contact Information

CDER Manual of Policies & Procedures | MAPP

Frequently Asked Questions about CDER

Reports & Budgets | CDER

The Office of New Drugs (OND) ensures that safe and effective drugs and biologics are available to the American people. OND includes eight review offices with 27 review divisions. OND's role is twofold: review applications and make an approval decision; set guidance and policy to ensure an efficient review process. OND reviews drug applications, interacts with the pharmaceutical industry and ultimately decides whether the benefits of a drug outweigh the known risks.

What We Do:

- Provide regulatory oversight for investigational studies during drug development.
- Make decisions regarding marketing approval for new (innovator or non-generic) drugs, including decisions related to changes to already marketed products.
- Provide guidance to regulated industry on a wide variety of clinical, scientific, and regulatory matters.
- Communicate with the pharmaceutical industry to achieve the common goal of more efficient drug development.
- Oversee the 21st Century Review initiative, which is a set of performance standards CDER follows when conducting drug reviews. The goal of this initiative is to make the drug review process organized and integrated, and ensure all decision makers are heard.

FDA OND CHAT

What was the agenda of the Gastrointestinal Drugs Advisory Committee Meeting May 19, 2023?

The agenda of the Gastrointestinal Drugs Advisory Committee Meeting on May 19, 2023, was to discuss the new drug application (NDA) 212833, obeticholic acid (OCA) 25 mg oral tablets, submitted by Intercept Pharmaceuticals, Inc., for the treatment of precirrhotic liver fibrosis due to nonalcoholic steatohepatitis.

Citations:

- Final Summary Minutes of the G ...
- Final Summary Minutes of the G ...
- Final Summary Minutes of the G ...

Type your question here ...

Questions?

Thank you!

Please reach out with questions

smferko@us.ibm.com

abhi.pandey@ibm.com

mjdeady@us.ibm.com

bgoodness@us.ibm.com



**The Global Healthcare
Data Science Community**

Contact Channels

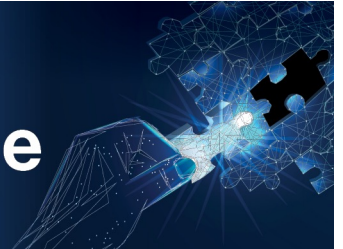
UK +44 1843 609600
office@phuse.global
phuse.global

Social Media

@phusetwitta
/phusebook
/phusetube
/company/phuse



**PHUSE/FDA
Innovation Challenge**



Appendix



EHR Exchange Pilot Platform: Background

USCDI: A standardized set of data elements required for sharing information between health IT systems, including health information such as patient demographics, clinical notes, medications, allergies, problems, procedures, laboratory results, and imaging reports.

United States Core Data for Interoperability (USCDI)

The United States Core Data for Interoperability (USCDI) is a standardized set of health data classes and constituent data elements for nationwide, interoperable health information exchange. Review the [USCDI Fact Sheet](#) to learn more.

A USCDI Data Class is an aggregation of Data Elements by a common theme or use case.

A USCDI Data Element is a piece of data defined in USCDI for access, exchange or use of electronic health information.

USCDI ONC New Data Element & Class (ONDEC) Submission System



Please reference the [USCDI version 1 document](#) to the left for applicable standards versions associated with USCDI v1.

Allergies and Intolerances

Harmful or undesired physiological responses associated with exposure to a substance.

Substance (Medication)
Substance (Drug Class)
Reaction

Immunizations

Record of vaccine administration.

Immunizations

Procedures

Activity performed for or on a patient as part of the provision of care.

Procedures

Laboratory

Analysis of clinical specimens to obtain information about the health of a patient.

Tests

Provenance

The metadata, or extra information about data, regarding who created the data and when it was created.

Assessment and Plan of Treatment

Health professional's conclusions and working assumptions that will guide treatment of the patient.

<https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi>



EHR Exchange Pilot Platform: Background

U.S. Core: Provide detailed technical specifications and implementation guidance to support the adoption of the USCDI. It includes subset of FHIR resources and data elements to specify the content, structure, and format of the data to be exchanged aligning systems and implementations.

HL7 FHIR® US Core Implementation Guide STU3 Release 3.1.1

Home Guidance FHIR Artifacts Security Examples Downloads

Table of Contents Home > Profiles and Extensions > StructureDefinition US Core Condition

This page is part of the US Core (v3.1.1: STU3) based on FHIR R4. The current version which supercedes this version is 6.1.0. For a full list of available versions, see the [Directory of published versions](#). Page versions: STU4 STU3 STU2 STU1

Narrative Content Detailed Descriptions XML JSON

3.26.1 StructureDefinition-us-core-condition

This profile sets minimum expectations for the [Condition](#) resource to record, search, and fetch a list of conditions associated with a patient. It identifies which core elements, extensions, vocabularies and value sets **SHALL** be present in the resource when using this profile.

Example Usage Scenarios:

The following are example usage scenarios for the US Core Condition profile:

- Query for a Patient's current or historical conditions
- [Record or update](#) a Patient's conditions

3.26.1.1 Mandatory and Must Support Data Elements

The following data-elements are mandatory (i.e data **MUST** be present) or must be supported if the data is present in the sending system ([Must Support](#) definition). They are presented below in a simple human-readable explanation. Profile specific guidance and examples are provided as well. The [Formal Profile Definition](#) below provides the formal summary, definitions, and terminology requirements.

Each Condition must have:

1. a status of the condition*
2. a category
3. a code that identifies the condition

<https://hl7.org/fhir/us/core/>



Data Quality Study: Results

- **Data Quality results** can be used to demonstrate **overall compliance** or **compare differences in EHR data between partners**
- **Example:** Synthetic results for a subset of tests relevant to CDISC patient demographics shown below, graphing **% of resources that passed for each partner** and the **average for all partners**

Data Quality Analysis for CDISC Demographic Data Elements: Synthetic Patients

