

# **NLP2FHIR: A Scalable FHIR-based Clinical Data Normalization Pipeline and Its Research Applications**

**Guoqian Jiang, MD, PhD**

**Division of Digital Health Sciences**

**Department of Health Sciences Research**

**Mayo Clinic**

**October 15, 2020**

# Motivation

- Heterogeneity of electronic health record (EHR) systems and clinical natural language processing (NLP) tools impedes second use of EHR data for clinical and translation research.
- Standards-based modeling of EHR data holds great promise for system interoperability and large-scale data integration and analytics.
- HL7 Fast Healthcare Interoperability Resources (FHIR)
  - A next generation standards framework for exchanging electronic healthcare data
  - Detailed clinical information models as NLP target and output
  - Powerful modeling mechanism (e.g., profiles and extensions)


[Home](#)   [Getting Started](#)   [Documentation](#)   [Resources](#)   [Profiles](#)   [Extensions](#)   [Operations](#)   [Terminologies](#)
 [Table of Contents](#) > [Resources](#)

# >100 Resources

This page is part of the FHIR Specification (v4.0.1: R4 - Mixed [Normative](#) and [STU](#)). This is the current published version. For a full list of available versions, see the [Directory of published versions](#).

## 1.2 Resource Index

FHIR Infrastructure  Work Group	Maturity Level: N/A	Standards Status: Informative
---------------------------------	---------------------	-------------------------------

This page is provided to help find resources quickly. There is also a more [detailed classification, ontology, and description](#). For background to the layout on the layers in this page, see the [Architect's Overview](#). See also the abstract Base Resources [Resource](#) and [DomainResource](#).

Categorized	Alphabetical	R2 Layout	By Maturity	Security Category	By Standards Status	By Committee	
<b>Clinical</b>  <b>Summary</b> <ul style="list-style-type: none"> <li>AllergyIntolerance 3</li> <li>AdverseEvent 0</li> <li>Condition (Problem) 3</li> <li><u>Procedure</u> 3</li> <li>FamilyMemberHistory 2</li> <li>ClinicalImpression 0</li> <li>DetectedIssue 1</li> </ul>	<b>Diagnostics</b> <ul style="list-style-type: none"> <li><u>Observation</u> </li> <li>Media 1</li> <li>DiagnosticReport 3</li> <li>Specimen 2</li> <li>BodyStructure 1</li> <li>ImagingStudy 3</li> <li>QuestionnaireResponse 3</li> <li>MolecularSequence 1</li> </ul>	<b>Medications</b> <ul style="list-style-type: none"> <li>MedicationRequest 3</li> <li>MedicationAdministration 2</li> <li>MedicationDispense 2</li> <li>MedicationStatement 3</li> <li>Medication 3</li> <li>MedicationKnowledge 0</li> <li>Immunization 3</li> <li>ImmunizationEvaluation 0</li> <li>ImmunizationRecommendation 1</li> </ul>	<b>Care Provision</b> <ul style="list-style-type: none"> <li>CarePlan 2</li> <li>CareTeam 2</li> <li>Goal 2</li> <li>ServiceRequest 2</li> <li>NutritionOrder 2</li> <li>VisionPrescription 2</li> <li>RiskAssessment 1</li> <li>RequestGroup 2</li> </ul>	<b>Request &amp; Response</b> <ul style="list-style-type: none"> <li>Communication 2</li> <li>CommunicationRequest 2</li> <li>DeviceRequest 1</li> <li>DeviceUseStatement 0</li> <li>GuidanceResponse 2</li> <li>SupplyRequest 1</li> <li>SupplyDelivery 1</li> </ul>			
<b>Financial</b>  <b>Support</b> <ul style="list-style-type: none"> <li>Coverage 2</li> <li>CoverageEligibilityRequest 2</li> <li>CoverageEligibilityResponse 2</li> <li>EnrollmentRequest 0</li> <li>EnrollmentResponse 0</li> </ul>	<b>Billing</b> <ul style="list-style-type: none"> <li>Claim 2</li> <li>ClaimResponse 2</li> <li>Invoice 0</li> </ul>	<b>Payment</b> <ul style="list-style-type: none"> <li>PaymentNotice 2</li> <li>PaymentReconciliation 2</li> </ul>	<b>General</b> <ul style="list-style-type: none"> <li>Account 2</li> <li>ChargeItem 0</li> <li>ChargeItemDefinition 0</li> <li>Contract 1</li> <li>ExplanationOfBenefit 2</li> <li>InsurancePlan 0</li> </ul>				

## Structure

Name	Flags	Card.	Type	Description & Constraints								
Condition	I TU		DomainResource	<p>Detailed information about conditions, problems or diagnoses</p> <p>+ Guideline: Condition.clinicalStatus SHALL be present if verificationStatus is not entered in error and category is problem-list-item</p> <p>+ Rule: If condition is problem then clinicalStatus is required</p> <p>+ Rule: Condition.category is problem-list-item</p> <p>Elements defined in this resource are subject to the following constraints:</p> <p><b>modifierExtension</b></p> <p>External IDs for this condition</p> <p><b>active   recurrence   Condition Clinical Status</b></p> <p><b>unconfirmed   provisional   ConditionVerification</b></p> <p><b>problem-list-item   e</b></p> <p><b>Condition Category Codes (Extensible)</b></p> <p>Subjective severity of condition</p> <p><b>Condition/Diagnosis Severity (Preferred)</b></p> <ul style="list-style-type: none"> <li>Include these codes as defined in <a href="http://snomed.info/sct">http://snomed.info/sct</a></li> </ul> <table border="1"> <thead> <tr> <th>Code</th> <th>Display</th> </tr> </thead> <tbody> <tr> <td>24484000</td> <td>Severe</td> </tr> <tr> <td>6736007</td> <td>Moderate</td> </tr> <tr> <td>255604002</td> <td>Mild</td> </tr> </tbody> </table>	Code	Display	24484000	Severe	6736007	Moderate	255604002	Mild
Code	Display											
24484000	Severe											
6736007	Moderate											
255604002	Mild											
identifier	$\Sigma$	0..*	Identifier									
clinicalStatus	?! $\Sigma$ I	0..1	CodeableConcept									
verificationStatus	?! $\Sigma$ I	0..1	CodeableConcept									
category		0..*	CodeableConcept									
severity		0..1	CodeableConcept									
code	$\Sigma$	0..1	CodeableConcept	Identification of the condition, problem or diagnosis								
bodySite	$\Sigma$	0..*	CodeableConcept	Anatomical location, if relevant								
subject	$\Sigma$	1..1	Reference(Patient   Group)	<b>Condition/Problem/Diagnosis Codes (Example)</b>								
encounter	$\Sigma$	0..1	Reference(Encounter)	<b>SNOMED CT Body Structures (Example)</b>								
onset[x]	$\Sigma$	0..1		Who has the condition?								
onsetDateTime			dateTime	Estimated or actual date, date-time, or age								
onsetAge			Age									
onsetPeriod			Period									
onsetRange			Range									
onsetString			string									
abatement[x]	I	0..1		When in resolution/remission								
abatementDateTime			dateTime									
abatementAge			Age									
abatementPeriod			Period									
abatementRange			Range									
abatementString			string									
recordedDate	$\Sigma$	0..1	dateTime	Date record was first recorded								

# Datatype

# Valueset



This page is part of the FHIR Specification (v4.0.1: R4 - Mixed Normative and STU). This is the current published version. For a full list of available versions, see the [Directory of published versions](#) ↗

[Content](#) [Examples](#) [Detailed Descriptions](#) [Mappings](#) [Profiles & Extensions](#) [R3 Conversions](#)

## 9.2.5 Resource Condition - Examples

Patient Care ↗	Work Group	Maturity Level: N/A	Standards Status: Informative	Security Category: Patient	Compartments: Encounter, Patient, Practitioner, RelatedPerson
----------------	------------	---------------------	-------------------------------	----------------------------	---------------------------------------------------------------

Example Name	id	Format
General Condition Example	example	XML JSON Turtle
2nd Example	example2	XML JSON Turtle
Real-word condition example (fever)	f201	XML JSON Turtle
Real-word condition example (malignancy)	f202	XML JSON Turtle
Real-word condition example (sepsis)	f203	XML JSON Turtle
Real-word condition example (renal insufficiency)	f204	XML JSON Turtle
Real-word condition example (Bacterial Infection)	f205	XML JSON Turtle
Real-word condition example (heart)	f001	XML JSON Turtle
Real-word condition example (lung)	f002	XML JSON Turtle
Real-word condition example (abscess)	f003	XML JSON Turtle
Stroke example (for prognosis)	stroke	XML JSON Turtle
Family history concern	family-history	XML JSON Turtle



# Exchange Format: XML, JSON, RDF

## Condition Model Definition

Name	Flags	Card.	Type
Condition	I TU		DomainResource
Identifier		Σ 0..*	Identifier
clinicalStatus		?! Σ I 0..1	CodeableConcept
verificationStatus		?! Σ I 0..1	CodeableConcept
category		0..*	CodeableConcept
severity		0..1	CodeableConcept
code		Σ 0..1	CodeableConcept
bodySite		Σ 0..*	CodeableConcept
subject		Σ 1..1	Reference(Patient   Group)
encounter		Σ 0..1	Reference(Encounter)
onset[x]		Σ 0..1	
onsetDateTime			dateTime
onsetAge			Age
onsetPeriod			Period
onsetRange			Range
onsetString			string
abatement[x]	I	0..1	
abatementDateTime			dateTime
abatementAge			Age
abatementPeriod			Period
abatementRange			Range
abatementString			string
recordedDate	Σ	0..1	dateTime

## Condition Instance Data in JSON

```
{
  "resourceType": "Condition",
  "id": "example",
  "text": {
    "status": "generated",
    "div": "<div xmlns='http://www.w3.org/1999/xhtml'>Severe burn of left ear (Date: 24-May 2012)</div>"
  },
  "clinicalStatus": {
    "coding": [
      {
        "system": "http://terminology.hl7.org/CodeSystem/condition-clinical",
        "code": "active"
      }
    ]
  },
  "verificationStatus": {
    "coding": [
      {
        "system": "http://terminology.hl7.org/CodeSystem/condition-ver-status",
        "code": "confirmed"
      }
    ]
  },
  "category": [
    {
      "coding": [
        {
          "system": "http://terminology.hl7.org/CodeSystem/condition-category",
          "code": "encounter-diagnosis",
          "display": "Encounter Diagnosis"
        },
        {
          "system": "http://snomed.info/sct",
          "code": "439401001",
          "display": "Diagnosis"
        }
      ]
    }
  ],
  "severity": {
    "coding": [
      {
        "system": "http://snomed.info/sct",
        "code": "2448400",
        "display": "Severe"
      }
    ]
  }
}
```

The diagram illustrates the mapping of fields from the Condition Model Definition table to the JSON instance. Red arrows point from the table columns to the JSON object properties. The mappings are:

- Identifier → clinicalStatus (multiple arrows)
- clinicalStatus → clinicalStatus
- verificationStatus → verificationStatus
- category → category
- severity → severity
- code → clinicalStatus (multiple arrows)
- bodySite → clinicalStatus (multiple arrows)
- subject → category
- onset[x] → onset (multiple arrows)
  - onsetDateTime → onset.dateTime
  - onsetAge → onset.Age
  - onsetPeriod → onset.Period
  - onsetRange → onset.Range
  - onsetString → onset.string
- abatement[x] → abatement (multiple arrows)
  - abatementDateTime → abatement.dateTime
  - abatementAge → abatement.Age
  - abatementPeriod → abatement.Period
  - abatementRange → abatement.Range
  - abatementString → abatement.string
- recordedDate → recordedDate

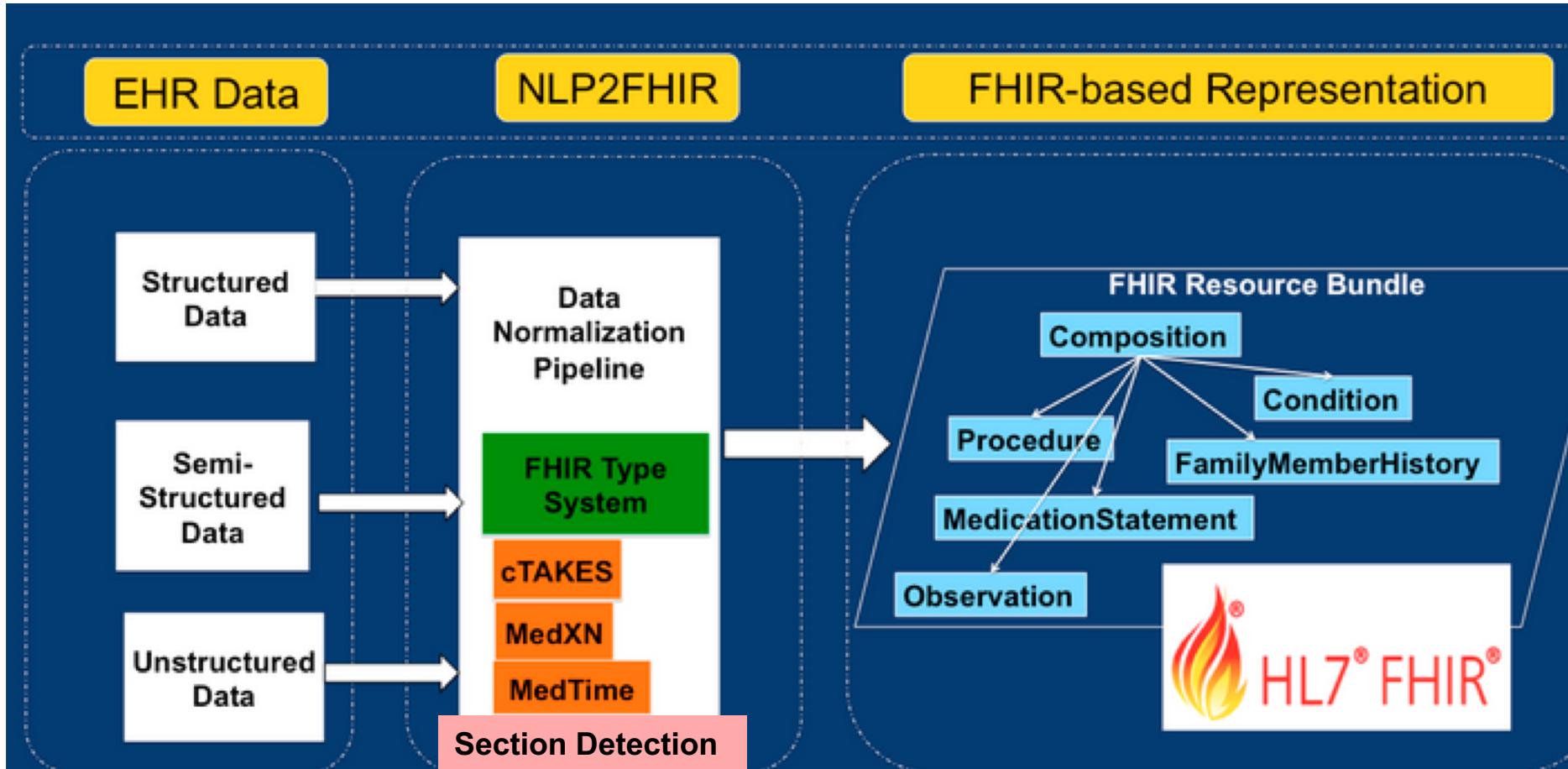
# Use of FHIR in Research

- Major EHR vendors and healthcare providers have been involved in the development and adoption of FHIR through HL7 Argonaut Project.
- NIH Notice on Use of FHIR in Research
  - To encourage NIH researchers to explore the use of the FHIR standard to capture, integrate, and exchange clinical data for research purposes and to enhance capabilities to share research data
- National collaborations for clinical research
  - The Sync for Science (S4S) initiative - help patient share EHR data with researchers and empower individuals to participate in health research.
  - NCATS CD2H –coordinate and contribute to many FHIR initiatives, with the goal of making its adoption for translational research useful, easy, and a best practice.

# Gaps on how to represent unstructured information in clinical narratives using FHIR

- Current HL7 Argonaut project has not yet provided a solution to standardize unstructured data
  - Using NLP to gain access to the narrative content in EHRs via FHIR will be of great value to data analytics, quality improvement, and advanced decision support.
- Few studies have been done on
  - (1) the tool development for generating the FHIR resource instances from clinical narratives leveraging the NLP technology; and
  - (2) assessing the discrepancies between FHIR data models and NLP type systems.
- An FHIR standard-based common type system would better improve semantic interoperability between heterogeneous clinical data sources, given the rapid adoption of FHIR as an international standard in different EHR systems.

# NLP2FHIR: A FHIR-based Clinical Data Normalization Pipeline



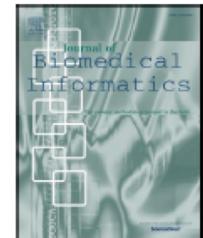
## Problem List – Reported (LOINC: 11450-4)

```
{
  "resourceType": "Bundle",
  "id": "999c5b02-3070-4785-b43b-8dcb92fd92b8",
  "type": "document",
  "entry": [
    {
      "fullUrl": "Composition/ON03FP00001D00235.txt",
      "resource": {
        "resourceType": "Composition",
        "id": "ON03FP00001D00235.txt",
        "identifier": {
          "system": "urn:ietf:rfc:3986",
          "value": "urn:uuid:f3fe1c0a-39cf-3c8a-a9ed-cd053d814b76"
        },
        "section": [
          {
            "code": {
              "coding": [
                {
                  "system": "http://hl7.org/fhir/ValueSet/doc-section-codes",
                  "code": "11450-4"
                }
              ]
            },
            "text": {
              "status": "additional",
              "div": "<div xmlns=\"http://www.w3.org/1999/xhtml>Problem list: 1. Breast cancer, stage I-c -CT chest (4/15/64) right apical thickening, repeat CT 6mos. -Abnormal screening mammogram (04/28/64): 1.3cm right LIQ mass with adjacent 6mm mass -R core needle bx (5/04/64): invasive ductal CA and DCIS -Lumpectomy & SNbx (5/24/64): 1.4cm mass, grade 2/3, 5mm focus of DCIS extending to within 1 mm of final lateral margin, deep margin of the 14mm mass positive for intraductal Ca, focal angiolymphatic invasion present. Tumor is ER+3, PR+3, her2neu not overexpressed, Ki67<10%. 0/4 SN negative. Reexcision planned 6/15/64</div>"
            },
            "entry": [
              {
                "reference": "Condition/edac5dae-423f-334b-b806-a3c5acdfde9f"
              },
              {
                "reference": "Procedure/43f61928-5307-3a19-a824-8ce446ea77d6"
              },
            ]
          }
        ]
      }
    }
  ]
}
```

↑

```
{
  "fullUrl": "Condition/edac5dae-423f-334b-b806-a3c5acdfde9f",
  "resource": {
    "resourceType": "Condition",
    "id": "edac5dae-423f-334b-b806-a3c5acdfde9f",
    "identifier": [
      {
        "system": "urn:ietf:rfc:3986",
        "value": "urn:uuid:edac5dae-423f-334b-b806-a3c5acdfde9f"
      }
    ],
    "code": {
      "coding": [
        {
          "system": "http://snomed.info/sct",
          "code": "190121004"
        }
      ],
      "text": "Breast cancer"
    },
    "bodySite": [
      {
        "coding": [
          {
            "system": "http://snomed.info/sct",
            "code": "361079003"
          }
        ],
        "text": "Breast"
      }
    ],
    "abatementString": "positive"
  }
},
```

↑



## Developing a FHIR-based EHR phenotyping framework: A case study for identification of patients with obesity and multiple comorbidities from discharge summaries



Na Hong<sup>a</sup>, Andrew Wen<sup>a</sup>, Daniel J. Stone<sup>a</sup>, Shintaro Tsuji<sup>a</sup>, Paul R. Kingsbury<sup>a</sup>, Luke V. Rasmussen<sup>b</sup>, Jennifer A. Pacheco<sup>b</sup>, Prakash Adekkattu<sup>c</sup>, Fei Wang<sup>c</sup>, Yuan Luo<sup>b</sup>, Jyotishman Pathak<sup>c</sup>, Hongfang Liu<sup>a</sup>, Guoqian Jiang<sup>a,\*</sup>

<sup>a</sup> Mayo Clinic, Rochester, MN, USA

<sup>b</sup> Northwestern University Feinberg School of Medicine, Chicago, IL, USA

<sup>c</sup> Weill Cornell Medicine, New York City, NY, USA

### ARTICLE INFO

**Keywords:**

Clinical phenotyping  
HL7 Fast Healthcare Interoperability Resources (FHIR)  
Electronic Health Records (EHRs)  
Natural language processing  
Algorithm portability

### ABSTRACT

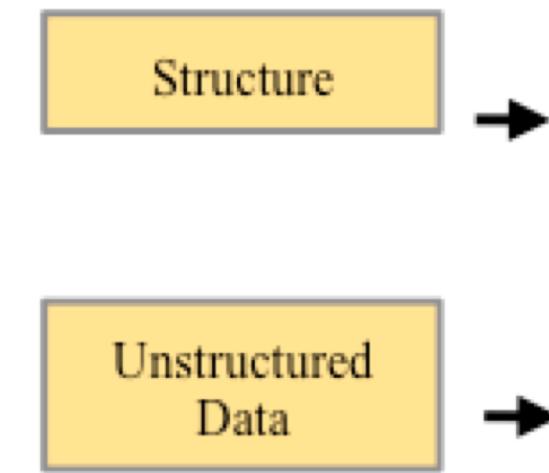
**Background:** Standards-based clinical data normalization has become a key component of effective data integration and accurate phenotyping for secondary use of electronic healthcare records (EHR) data. HL7 Fast Healthcare Interoperability Resources (FHIR) is an emerging clinical data standard for exchanging electronic healthcare data and has been used in modeling and integrating both structured and unstructured EHR data for a variety of clinical research applications. The overall objective of this study is to develop and evaluate a FHIR-based EHR phenotyping framework for identification of patients with obesity and its multiple comorbidities from semi-structured discharge summaries leveraging a FHIR-based clinical data normalization pipeline (known as NLP2FHIR).

**Methods:** We implemented a multi-class and multi-label classification system based on the i2b2 Obesity Challenge task to evaluate the FHIR-based EHR phenotyping framework. Two core parts of the framework are: (a) the conversion of discharge summaries into corresponding FHIR resources – Composition, Condition, MedicationStatement, Procedure and FamilyMemberHistory using the NLP2FHIR pipeline, and (b) the implementation of four machine learning algorithms (logistic regression, support vector machine, decision tree, and random forest) to train classifiers to predict disease state of obesity and 15 comorbidities using features

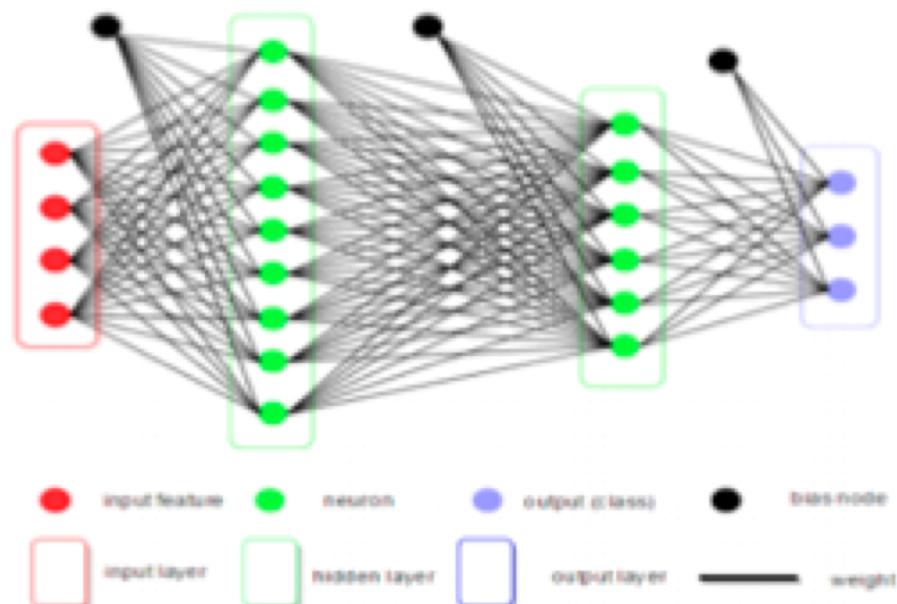
# The goal is to Improve Portability and Interpretability of Phenotype Algorithms



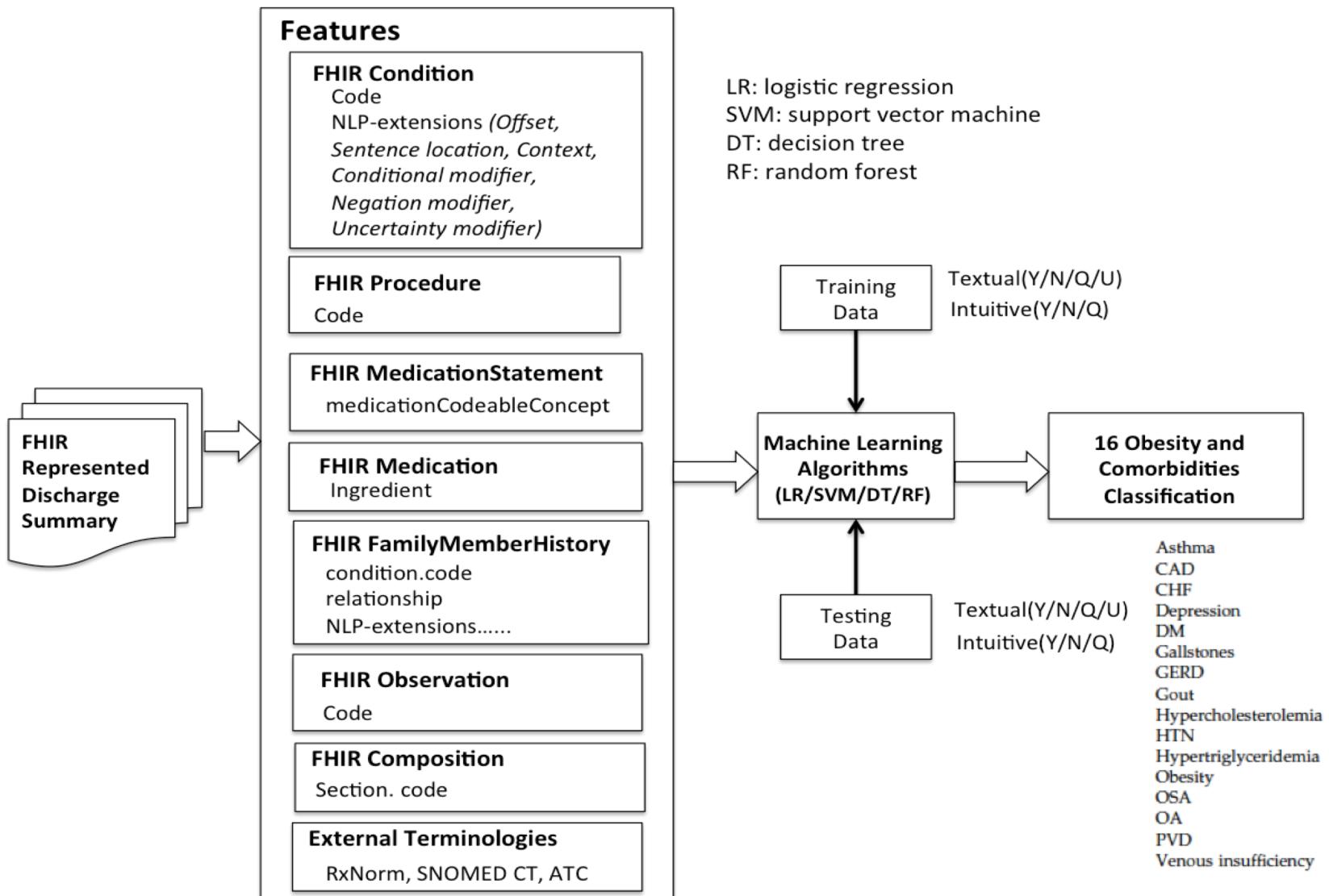
## EHR-Driven Deep Phenotyping



A 3-layers fully connected neural network (DNN)



# Classification of Obesity and Its Comorbidities



## CORD-19-on-FHIR -- Semantics for COVID-19 Discovery

---

CORD-19-on-FHIR is a superset of the COVID-19 Open Research Dataset ([CORD-19](#)) data, provided by the [Allen Institute](#) to support research on COVID-19 / SARS-CoV-2 / Novel Coronavirus. It is represented in [FHIR RDF](#) and was produced by data mining the CORD-19 dataset and adding semantic annotations. The purpose is to facilitate linkage with other biomedical datasets and enable answering research questions.

### Wiki

---

<https://github.com/fhircat/CORD-19-on-FHIR/wiki>

### Semantic annotations

---

So far, CORD-19-on-FHIR adds the following semantic annotations to the CORD-19 dataset, based only on parsing the titles and abstracts:

- Conditions - 103,968 instances
- Medications - 16,406 instances
- Procedures - 54,720 instances

We plan to parse the full text articles soon, for those that we can access.

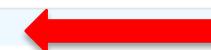
If you can generate other semantic linkages, please let us know. Collaboration is invited. Please issue a pull request or contact Guoqian Jiang [jiang.guoqian@mayo.edu](mailto:jiang.guoqian@mayo.edu) and Harold Solbrig [solbrig@jhu.edu](mailto:solbrig@jhu.edu).

JSON Raw Data Headers

```

resourceType: "Composition"
id: "00623bf2715e25d3acacb3f210d6888ed840e3cb.json"
identifier:
  system: "urn:ietf:rfc:3986"
  value: "urn:uuid:33a112c3-0131-3617-9cb1-765b5474a049"
status: "final"
confidentiality: "N"
section:
  0:
    code:
      coding:
        0:
          system: "http://hl7.org/fhir/ValueSet/doc-section-codes"
          code: "61150-9"
    text:
      status: "additional"
      div: "<div xmlns=\"http://www.w3.org/1999/xhtml\">TITLE: Transmissible gastroenteritis virus infection decreases arginine uptake by downregulating CAT-1 expression </div>"
    mode: "working"
entry:
  0:
    reference: "Condition/d207290f-d42b-35cd-a57c-db9feffc3f10"
  1:
    reference: "Condition/ce256190-76b2-3fad-ae10-6060dee4c3cf"
  2:
    reference: "Condition/dabd9625-c8d7-3df2-baf5-a1596fe36c8e"
  3:
    reference: "MedicationStatement/30e2e66d-99da-348c-b242-e56ac7b7e6a6"
  4:
    reference: "Medication/a56d2d61-ec6b-323e-b8c5-6cb11fdb735d"
  5:
    reference: "Procedure/0b2d9261-9015-3dbb-9d1f-54fa5f1f4c14"

```

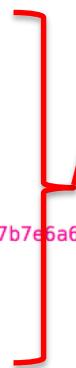


## Composition



## Section

## Semantic Annotations



```

resourceType: "Condition"
id: "00006534-eee4-3d38-bbe9-45374ee0f3c7"
identifier:
  0:
    system: "urn:ietf:rfc:3986"
    value: "urn:uuid:00006534-eee4-3d38-bbe9-45374ee0f3c7"
code:
  coding:
    0:
      system: "http://snomed.info/sct"
      code: "233604007"
      text: "pneumonia"
abatementString: "positive"

```



COVID-19 is an emerging, rapidly evolving situation.

Get the latest public health information from CDC: <https://www.coronavirus.gov>Get the latest research from NIH: <https://www.nih.gov/coronavirus>.Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/coronavirus/>

Journal List &gt; OUP Public Health Emergency Collection &gt; PMC7314034

## Oxford University Press Public Health Emergency Collection

Public Health Emergency COVID-19 Initiative

[J Am Med Inform Assoc.](#) 2020 May 27 : ocaa117.

PMCID: PMC7314034

Published online 2020 May 27. doi: [10.1093/jamia/ocaa117](https://doi.org/10.1093/jamia/ocaa117)PMID: [32458963](https://pubmed.ncbi.nlm.nih.gov/32458963/)

### Constructing Co-occurrence Network Embeddings to Assist Association Extraction for COVID-19 and Other Coronavirus Infectious Diseases

**David Oniani,<sup>1</sup> Guoqian Jiang, M.D., Ph.D,<sup>2</sup> Hongfang Liu, Ph.D,<sup>2</sup> and Feichen Shen, Ph.D<sup>2</sup>**[Author information](#) ▶ [Article notes](#) ▶ [Copyright and License information](#) [Disclaimer](#)

#### Abstract

#### Objective

As COVID-19 started its rapid emergence and gradually transformed into an unprecedented pandemic, the need for having a knowledge repository for the disease became crucial. To address this issue, a new COVID-19 machine readable dataset known as COVID-19 Open Research Dataset (CORD-19) has been released. Based on this, our objective was to build a computable co-occurrence network embeddings to assist association detection amongst COVID-19 related biomedical entities.

#### Materials and Methods



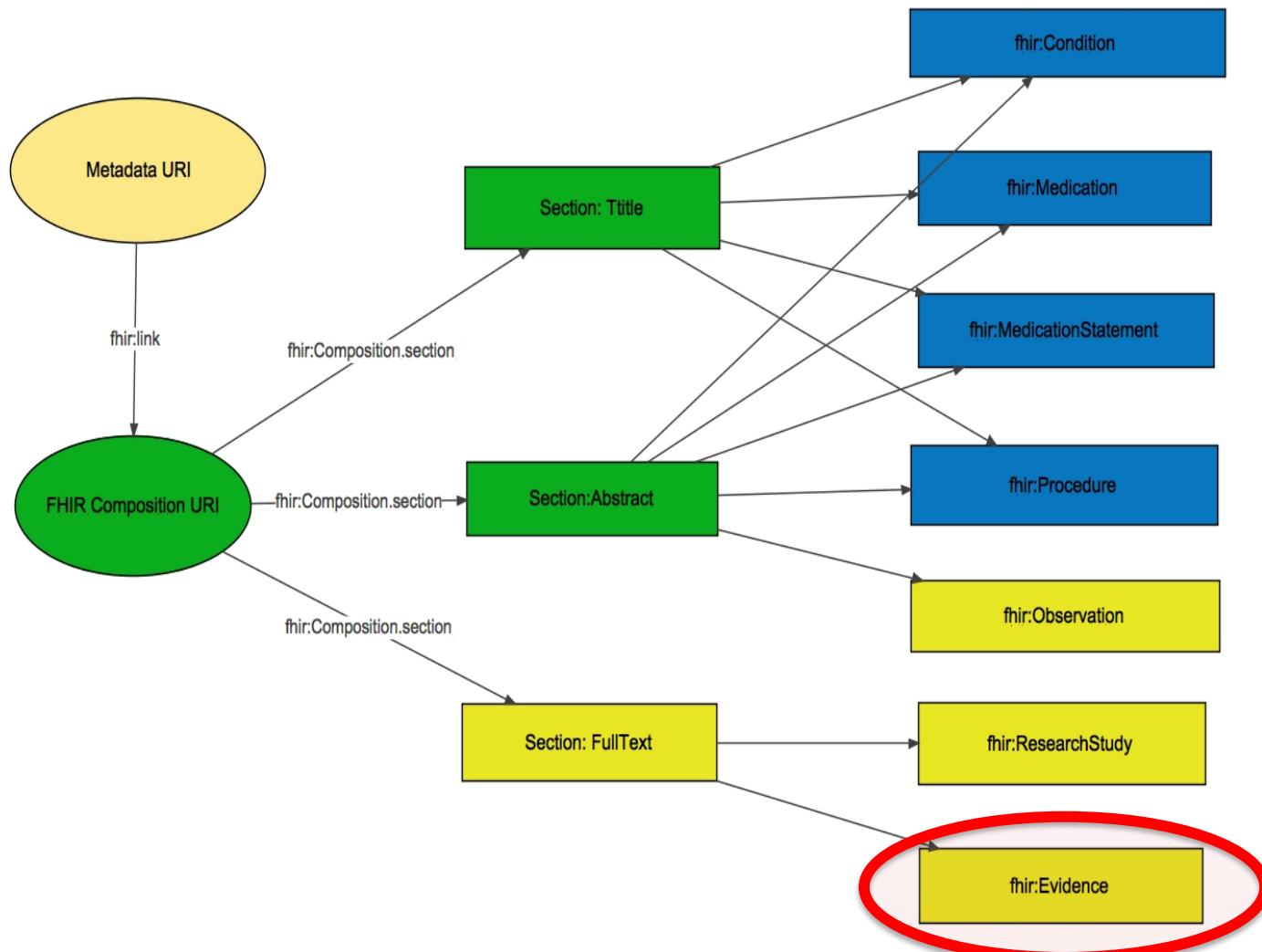
# Evidence Resource from EBMonFHIR

[Structure](#)[UML](#)[XML](#)[JSON](#)[Turtle](#)[R3 Diff](#)[All](#)

## Structure

Name	Flags	Card.	Type	Description & Constraints
Evidence	TU		MetadataResource	Single evidence bit + Warning: Name should be usable as an identifier for the module by machine processing applications such as code generation Elements defined in Ancestors: id, meta, implicitRules, language, text, contained, extension, modifierExtension, url, identifier, version, name, title, status, experimental, date, publisher, contact, description, useContext, jurisdiction, purpose, copyright, approvalDate, lastReviewDate, effectivePeriod
url	$\Sigma$	0..1	uri	Canonical identifier for this evidence, represented as a globally unique URI
identifier	$\Sigma$	0..*	Identifier	Additional identifier for the summary
version	$\Sigma$	0..1	string	Business version of this summary
title	$\Sigma$	0..1	string	Name for this summary (human friendly)
status	?! $\Sigma$	1..1	code	draft   active   retired   unknown <b>PublicationStatus (Required)</b>
date	$\Sigma$	0..1	dateTime	Date last changed
useContext	$\Sigma$	0..*	UsageContext	Use context
approvalDate		0..1	date	When the summary was approved by publisher
lastReviewDate		0..1	date	When the summary was last reviewed
contributor	$0..*$		Contributor	Contact detail and role for contributor to summary
relatedArtifact	$0..*$		RelatedArtifact	Link or citation to artifact associated with the summary
description		0..1	markdown	Description of the particular summary
assertion		0..1	markdown	Declarative description of the Evidence
note	$0..*$		Annotation	Footnotes and/or explanatory notes
variableDefinition		1..*	BackboneElement	Evidence variable
description		0..1	markdown	Description of the variable
note	$0..*$		Annotation	Footnotes and/or explanatory notes
variableRole	$\Sigma$	1..1	CodeableConcept	population   subpopulation   exposure   referenceExposure   measuredVariable   confounder <b>EvidenceVariableRole (Extensible)</b>

# CORD-19-on-FHIR Roadmap



Allen Institute Collaboration

W3C Biomedical Datasets in RDF

Translator Project Collaboration

OHDSI Collaboration

EBMonFHIR Collaboration

Mayo and JHU  
Clinical Research Collaboration



# CORD 19 Semantic Annotation Projects

David Booth edited this page 19 hours ago · 43 revisions

This page lists projects that are doing semantic annotation of the [CORD-19 dataset](#). If you know of a project that is not listed here, please add it **AND please contact [David Booth](#)**, who chairs a semi-weekly teleconference (11am Boston time) to coordinate and learn about each other's efforts.

Teleconferences are announced on the public [W3C Healthcare and Life Sciences mailing list](#).

## Table of Contents

- [2020-09-29 Houcemeddine Turki, University of Sfax, Tunisia: Wikidata and COVID-19, Creating a collaborative knowledge graph from CORD-19 scholarly publications](#)
- [2020-07-28 Marcin Joachimiak, Lawrence Berkeley Natl Lab: KG-COVID-19, A knowledge graph for COVID-19 response](#)
- [2020-07-21 Michael Liebman, IPQ Analytics: Modeling COVID-19, from the clinic back](#)
- [2020-06-23 Jin-Dong Kim Covid19-PubAnnotation](#)
- [2020-06-16 Victor Mireles, Semantic Web Company: COVID-19 Knowledge Graph.](#)
- [2020-06-16 Feichen Shen and David Oniani, Mayo Clinic: Constructing Co-occurrence Network Embeddings](#)
- [2020-06-02 Scott Malec, University of Pittsburgh: CORD-SEMANTICTRIPLES / Machine Reading for COVID-19 and Alzheimer's](#)
- [2020-06-02 Pedro Szekely, USC Information Sciences Institute: A Knowledge Graph Integrating Annotations On 20,000 COVID-19 Scientific Articles](#)
- [2020-05-26 Oliver Giles, SciBite: TERMite CORD19](#)
- [2020-05-19 Gaurav Vaidya: OmniCORD](#)
- [2020-05-19 Gollam Rabby, VSE University, Prague: Entity-Based-Document-Classification-on-the-CORD---19-Corpus](#)
- [2020-05-19 Marcin Joachimiak, Lawrence Berkeley National](#)
- [2020-05-19 Michael Liebman, IPQ Analytics LLC: Modeling COVID-19 From the Clinic Back](#)
- [2020-05-19 David Booth, Mayo Clinic \(consultant\): CORD-19-on-FHIR](#)
- [2020-05-12 Franck Michel, Université Côte d'Azur, CNRS, Inria: CORD-19 Named Entities Knowledge Graph \(CORD19-NEKG\)](#)
- [Project name: COVID-KG](#)
- [Project name: CORD-ReDrugS](#)

# Useful Links

- **NLP2FHIR GitHub:**
  - <https://github.com/BD2KOnFHIR/NLP2FHIR>
  - <https://github.com/BD2KOnFHIR/NLP2FHIR-Docker>
- **NLP2FHIR Demo at SMART Gallery**
  - <https://apps.smarthealthit.org/app/nlp2fhir-a-fhir-based-clinical-data-normalization-pipeline>

# Research Team

- **Standards & Interoperability**
  - **Guoqian Jiang (PI)**
  - **Nansu Zong (RA)**
  - **Andrew Wen (IS)**
  - **Daniel Stone (DSA)**
  - **Sijia Liu (IS)**
- **Clinical NLP**
  - **Hongfang Liu (PI)**
  - **Feichen Shen (RA)**
  - **Sunghwan Sohn (ACI)**

# NLP2FHIR Demo

**Daniel Stone**

**Andrew Wen**