MITRE is a not-for-profit organization that operates federally funded research and development centers, sponsored by the federal government.

FFRDCs assist the United States government with:

- Scientific research and analysis
- Development and acquisition
- Systems engineering and integration

We also have an independent research program that explores new and expanded uses of technologies to solve our sponsors' problems.
THE U.S HEALTHCARE SYSTEM: WORST IN THE DEVELOPED WORLD

U.S. lags behind in healthcare innovation

True interoperability ‘not even close’
High Demand for EHR Datasets

Non-clinical or secondary uses including: software development, testing, clinical training where realistic data is required
Restrictions

*Real patient records carry privacy, confidentiality, consent, policy, and legal restrictions*
Privacy Risks

Deidentified and anonymized records have been successfully reidentified
EHR datasets are difficult to obtain. Anonymized records are being bought and/or sold by federal and state health departments, hospitals, health insurers, pharmacists, government lobby groups, marketers, and data brokers.
Synthetic Patient Generation

Realistic Health Data

No Cost, No Restrictions
<table>
<thead>
<tr>
<th>Reasons Patients Visit PCP</th>
<th>Causes of Premature Death (Years of Life Lost)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Routine Infant/Child Health Check</td>
<td>1 Ischemic Heart Disease</td>
</tr>
<tr>
<td>Essential Hypertension</td>
<td>2 Lung Cancer</td>
</tr>
<tr>
<td>Diabetes Mellitus</td>
<td>3 Alzheimer’s Disease</td>
</tr>
<tr>
<td>Normal Pregnancy</td>
<td>4 COPD</td>
</tr>
<tr>
<td>Respiratory Infections (Pharyngitis, Bronchitis, Sinusitis)</td>
<td>5 Cerebrovascular Disease</td>
</tr>
<tr>
<td>General Adult Medical Examination</td>
<td>6 Road Injuries</td>
</tr>
<tr>
<td>Disorders of Lipoid Metabolism</td>
<td>7 Self-Harm</td>
</tr>
<tr>
<td>Ear Infections (Otitis Media)</td>
<td>8 Diabetes Mellitus</td>
</tr>
<tr>
<td>Asthma</td>
<td>9 Colorectal Cancer</td>
</tr>
<tr>
<td>Urinary Tract Infections</td>
<td>10 Drug Use Disorders</td>
</tr>
</tbody>
</table>
As of August 2018: Veteran-Focused Modules,
In partnership with VHA

<table>
<thead>
<tr>
<th>“Top 5” Conditions Affecting Veterans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperlipidemia</td>
</tr>
<tr>
<td>Hypertension</td>
</tr>
<tr>
<td>Osteoarthritis</td>
</tr>
<tr>
<td>Type 2 Diabetes</td>
</tr>
<tr>
<td>Major Depressive Disorder</td>
</tr>
<tr>
<td>PTSD</td>
</tr>
<tr>
<td>Cancers:</td>
</tr>
<tr>
<td>- Colorectal</td>
</tr>
<tr>
<td>- Lung</td>
</tr>
<tr>
<td>- Prostate</td>
</tr>
</tbody>
</table>
Publicly available data sources

a.k.a. care pathways, critical pathways, integrated care pathways, care maps, clinical practice algorithms
Modules are written in JSON
Synthea Module Builder UI
https://synthetichealth.github.io/module-builder/

**Examplitis**
State Type: *ConditionOnset*

**Enter value**

Target Encounter: *Wellness_Encounter*

**Add Assign to Attribute**

Codes
- remove
  - System: SNOMED-CT
  - Code: 123
  - Display: Examplitis

Transition Type: *Direct*

Transition To: *Wellness_Encounter*
Control States control the flow through the state machine

- **Age_Guard**: Guard to allow if age $\geq 18$ years
  - **Initialize_Chemo_Counter**: SetAttribute, set 'cr_chemo_count' = '0'
  - **Count_A_Chemo_Treatment**: Counter, increment value of attribute 'cr_chemo_count' by 1
Clinical States drive the progression of disease and care

- **Emphysema Symptom 1**: Shortness of Breath: 70 - 100

- **Medication Order**: RxNorm[308192]: Amoxicillin 500 MG Oral Tablet
  Assign to Attribute: 'antibiotic_prescription'

- **Chest CT Scan**
  - Procedure: SNOMED-CT[418891003]: Computed tomography of chest and abdomen
  - Perform at Diagnosis Encounter II
  - Reason: Suspected Lung Cancer
  - Duration: 20 - 60 minutes

- **Record_HA1C**
  - Observation: Record value from Vital Sign 'Blood Glucose' in %
    LOINC[4548-4]: Hemoglobin A1c/Hemoglobin total in Blood Category: laboratory

- **Diabetic CarePlan**
  - SNOMED-CT[69830004]: Diabetes self management plan
    Reason: diabetes_stage
  - Activities:
    - SNOMED-CT[16067007]: Diabetic diet
    - SNOMED-CT[22906009]: Exercise therapy
  - Goals:
    - Observation 'LOINC [4548-4] Hemoglobin A1c total in Blood' < 7.0
    - Maintain blood pressure below 140/90 mmHg
    - Improve and maintenance of optimal food health
    - Address patient knowledge deficit on diabetic self-care
Model Validation

Conditions and comorbidities:
Population - Incidence and prevalence
Individual - Frequencies and progression
Example Module – Appendicitis
Smith292, John949

Born: March 25, 1947
Gender: Male

Encounters:

Conditions:

Procedures:
Smith292, John949

Born: March 25, 1947
Gender: Male

Encounters:

Conditions:

Procedures:

Current Date of Simulation:
March 25, 1947
Smith292, John949

Born: March 25, 1947
Gender: Male

Encounters:

Conditions:

Procedures:

Current Date of Simulation: March 25, 1947
Smith292, John949
Born: March 25, 1947
Gender: Male

Encounters:

Conditions:

Procedures:

Current Date of Simulation:
March 25, 1947
Current Date of Simulation:
March 25, 1947

Smith292, John949
Born: March 25, 1947
Gender: Male

Encounters:

Conditions:

Procedures:
Current Date of Simulation: March 25, 1947

Smith292, John949

Born: March 25, 1947
Gender: Male

Encounters:

Conditions:

Procedures:
Smith292, John949
Born: March 25, 1947
Gender: Male

Encounters:

Conditions:

Procedures:

Current Date of Simulation:
March 25, 1947
Smith292, John949

Born: March 25, 1947
Gender: Male

Encounters:

Conditions:

Procedures:

Current Date of Simulation:
March 25, 1947
Smith292, John949

Born: March 25, 1947
Gender: Male

Encounters:

Conditions:

Procedures:

Current Date of Simulation:
July 23, 2013
Smith292, John949

Born: March 25, 1947
Gender: Male

Encounters:

Conditions:

Procedures:

Current Date of Simulation:
July 23, 2013
Smith292, John949
Born: March 25, 1947
Gender: Male

Encounters:

Conditions:

Procedures:

Current Date of Simulation:
July 23, 2013
Smith292, John949
Born: March 25, 1947
Gender: Male

Encounters:
Emergency Room Admission,
July 23, 2013

Conditions:
Appendicitis, Diagnosed July 23, 2013
Rupture of appendix, Diagnosed July 23, 2013

Procedures:

Current Date of Simulation:
July 23, 2013
Smith 292, John 949
Born: March 25, 1947
Gender: Male

Encounters:
Emergency Room Admission, July 23, 2013

Conditions:
Appendicitis, Diagnosed July 23, 2013
Rupture of appendix, Diagnosed July 23, 2013

Procedures:
Smith292, John949
Born: March 25, 1947
Gender: Male

Encounters:
Emergency Room Admission,
July 23, 2013

Conditions:
Appendicitis, Diagnosed July 23, 2013
Rupture of appendix, Diagnosed July 23, 2013

Procedures:
Blood Test, Performed July 23, 2013
Smith292, John949

Born: March 25, 1947
Gender: Male

Encounters:
Inpatient Encounter, July 23, 2013 →

Conditions:
Appendicitis, Diagnosed July 23, 2013
Rupture of appendix, Diagnosed July 23, 2013

Procedures:
Blood Test, Performed July 23, 2013
Smith292, John949

Born: March 25, 1947
Gender: Male

Encounters:
Emergency Room Admission,
Inpatient Encounter,
July 23, 2013 →

Conditions:
Appendicitis, Diagnosed July 23, 2013
Rupture of appendix, Diagnosed July 23, 2013

Procedures:
Blood Test, Performed July 23, 2013
Appendectomy, Performed July 23, 2013

Current Date of Simulation: July 23, 2013
Smith292, John949

Born: March 25, 1947
Gender: Male

Encounters:
Emergency Room Admission,
Inpatient Encounter,
July 23, 2013 →

Conditions:
Appendicitis, Diagnosed July 23, 2013
Rupture of appendix, Diagnosed July 23, 2013
History of appendectomy, Diagnosed July 23 2013

Procedures:
Blood Test, Performed July 23, 2013
Appendectomy, Performed July 23, 2013
Smith292, John949
Born: March 25, 1947
Gender: Male

Encounters:
Emergency Room Admission,
Inpatient Encounter,
July 23, 2013 →

Conditions:
Appendicitis, Diagnosed July 23, 2013
Rupture of appendix, Diagnosed July 23, 2013
History of appendectomy, Diagnosed July 23 2013

Procedures:
Blood Test, Performed July 23, 2013
Appendectomy, Performed July 23, 2013

Current Date of Simulation: July 23, 2013
Smith292, John949
Born: March 25, 1947
Gender: Male

Encounters:
Emergency Room Admission,
Inpatient Encounter,
July 23, 2013 →

Conditions:
Appendicitis, Diagnosed July 23, 2013
Rupture of appendix, Diagnosed July 23, 2013
History of appendectomy, Diagnosed July 23, 2013

Procedures:
Appendectomy, Performed July 23, 2013

Current Date of Simulation:
July 23, 2013
Smith292, John949
Born: March 25, 1947
Gender: Male

Encounters:
Emergency Room Admission,
Inpatient Encounter,

Conditions:
Appendicitis, Diagnosed July 23, 2013
Rupture of appendix, Diagnosed July 23, 2013
History of appendectomy, Diagnosed July 23, 2013

Procedures:
Appendectomy, Performed July 23, 2013

Current Date of Simulation: July 27, 2013
Smith292, John949

Born: March 25, 1947
Gender: Male

Encounters:

Conditions:
Appendicitis, Diagnosed July 23, 2013
Rupture of appendix, Diagnosed July 23, 2013
History of appendectomy, Diagnosed July 23 2013

Procedures:
Appendectomy, Performed July 23, 2013

Current Date of Simulation: July 27, 2013
Export Formats

- HL7® C-CDA®
- HL7® FHIR®
- Human Readable HTML or Plain Text
- CSV
- You tell us!
Easy to Set Up – Under an Hour
No Cost, No Restrictions
SyntheticMass

1,000,000 synthetic patients available for download today free of cost, free of restrictions

https://syntheticmass.mitre.org
Population

Number of Residents

Region Type: Cities and Towns
Data Set: Synthetic data generated from Synthea
Total Population: 1,009,150
Mean: 2,875
Max: Boston Cities and Towns: 99,401
Min: Gosnold Cities and Towns: 21
How are people using Synthea?
Software Development & Integration

HL7 FHIR Connectathons since May 2017
PULSE@MassChallenge
IHIC Datathon, Athens October 2017
FHIR DevDays, Boston June 2018
Education

Risk-free data for students and researchers in clinical informatics
Data Analysis

Model validation – performing analytics to compare synthetic data to real data
Building tooling & models before moving to real data
Performance Testing

*Use Synthea to generate hundreds, thousands, or even millions of synthetic patient records, to test system performance at scale*
Citations of Synthea JAMIA Paper

https://doi.org/10.1093/jamia/ocx079

Use of Fast Healthcare Interoperability Resources (FHIR) in the Generation of Real World Evidence (RWE)
Regina Zoepf, FDA, Bethesda, MD, USA
Jeff Abolafia, Rho, Chapel Hill, NC, USA
Bhargava Reddy, UCB, Raleigh, NC USA

ABSTRACT
There is a lot of interest in the clinical trial community to understand what information is contained in electronic health records (EHRs) to support clinical trials. The use of FHIR has been endorsed by the Food and Drug Administration (FDA) and is widely being used by EHR vendors.

A pilot was conducted to assess whether data that are typically of interest to patients can be extracted from medical records through FHIR resources to a specifically, we assessed data in the SyntheaMed Synthea repository that diabetes diagnosis, medical history, concomitant medications, vital signs and personal health information.

Deliverables from this pilot include a sample annotated Case Report Form (CRF) Harmonization (CDASH), Study Data Tabulation Model (SDTM) and FHIR summary of our process followed and pilot experience, including challenges.

Influence of simulation on electronic health record use patterns among pediatric residents
Evander W Orenstein, Orr Irit Rasoool, Mark V Mai, Adam C Deizny, Wanczyno Phillips, Levon Utdjian, Anthony Luberis, Jill Posner, Rebecca Tenney-Soelino, Chris P Bonafide

Evaluating Text Analytic Frameworks for Mental Health Surveillance
Benjamin Mayer, Josh Arnold, Edmon Begoli, Everett Rush, Michael Drewry
Oak Ridge National Laboratory (ORNL)
Oak Ridge, Tennessee, USA

Kris Brown, Eduardo Ponce, Sudarshan Srinivasan
Electrical Engineering and Computer Science (EECS)
University of Tennessee, Knoxville
Circle Park Dr., Knoxville, TN 37996

Fast and simple comparison of semi-structured data, with emphasis on electronic health records
Max Robinson, Jennifer Hadlock, Jiayang Yu, Alireza Khatamian, Aleksandr Y Aravkin, Eric W Deutsch, Nathan D Price, Sui Huang, Gustavo Glusman

Teaching data science fundamentals through realistic synthetic clinical cardiovascular data
Ted Laderas, Nicole Vasilevsky, Bjorn Pederson, Melissa Haendel, Shannon McVeeney, David Dorr

doi: https://doi.org/10.1101/232611
OSEHRA Synthetic Patient Data Project Group

"This group will develop an open source toolset for generating clinically valid synthetic patient data and loading it into VistA (and potentially other healthcare IT products). Existing products and services such as Synthea and MiHIN will be surveyed, and the group will collaborate on requirements and specifications."

- Chair: Bo Dagnall, Perspecta

https://www.osehra.org/groups/synthetic-patient-data-project-group
Recent Additions

- Anemia Module from OSEHRA Working Group
- Patient Provider Selection Behavior
  - Nearest, Quality, Random, Network
- Weight Loss Module
- Split Records by Provider
- Run Modules in Isolation
- Vital Signs Generator
- Individual Clinicians
- FHIR R4 Support
- Numerous Bug Fixes
What's Next?
Join our open source community!

Submit a Pull Request
Report an Issue
Request a Feature
Let us know how you use Synthea!
Contact

• Dylan Hall, Software Engineer
dehall@mitre.org
• Jason Walonoski, Project Lead
jwalonoski@mitre.org

https://synthetichealth.github.io/synthea/
MITRE’s mission-driven teams are dedicated to solving problems for a safer world. Through our federally funded R&D centers and public-private partnerships, we work across government to tackle challenges to the safety, stability, and well-being of our nation.

Learn more [www.mitre.org](http://www.mitre.org)