# Introduction to EMERSE: EMERSE 101

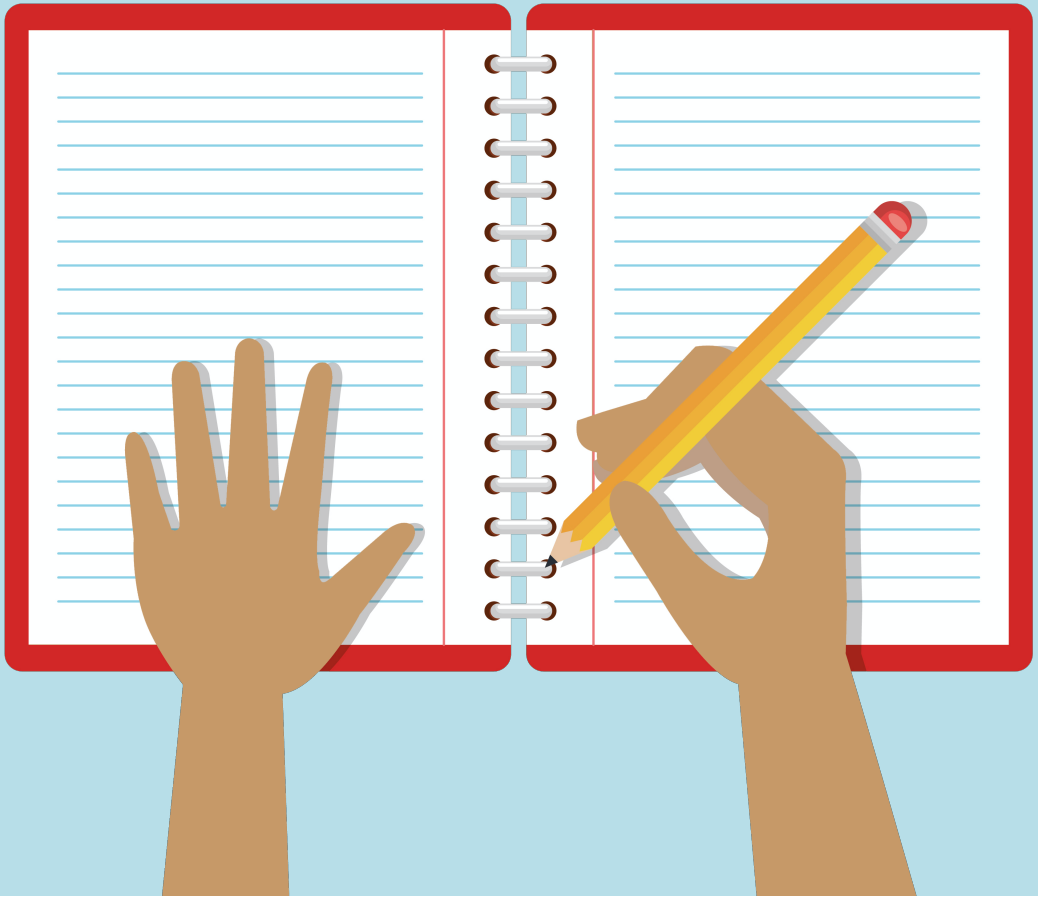
David Hanauer, MD, MS
University of Michigan
hanauer@umich.edu

Twitter: @informaticsGeek
@projectEMERSE

August 18, 2021

http://project-emerse.org/presentations.html

# If you're thinking of taking notes



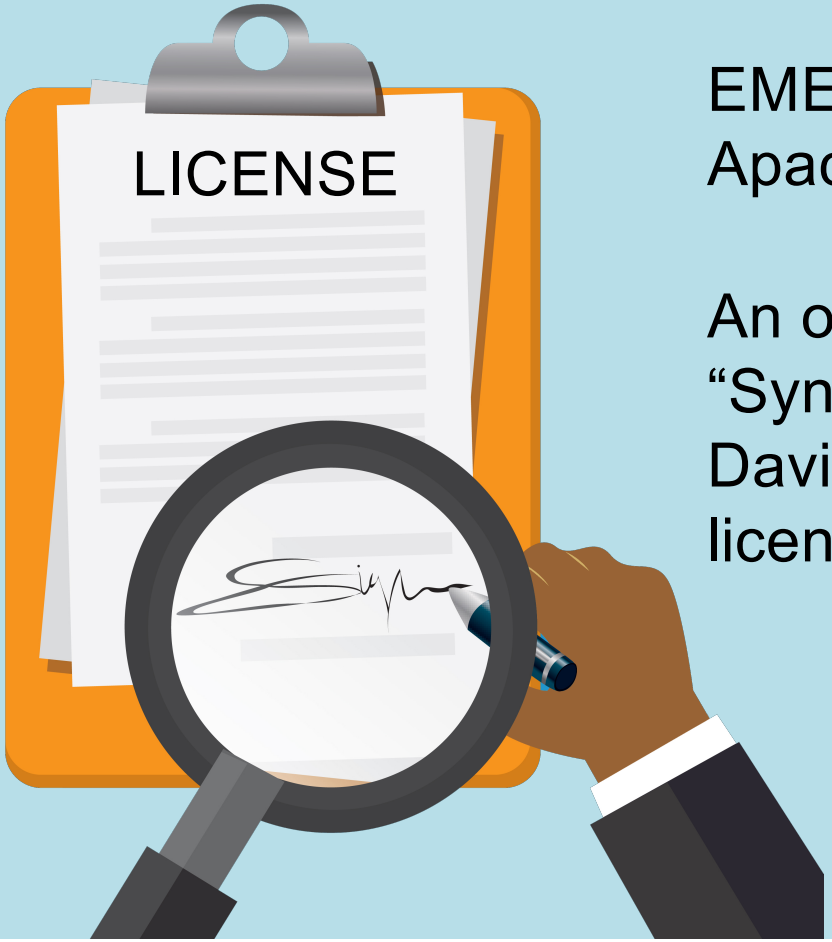
these slides can be found at:

link will be on
most slides

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Disclosure




EMERSE is available at no cost under the Apache 2.0 open source license

An optional addition to EMERSE is the “Synonyms Collection”, for which David Hanauer is entitled to a portion of the license fees

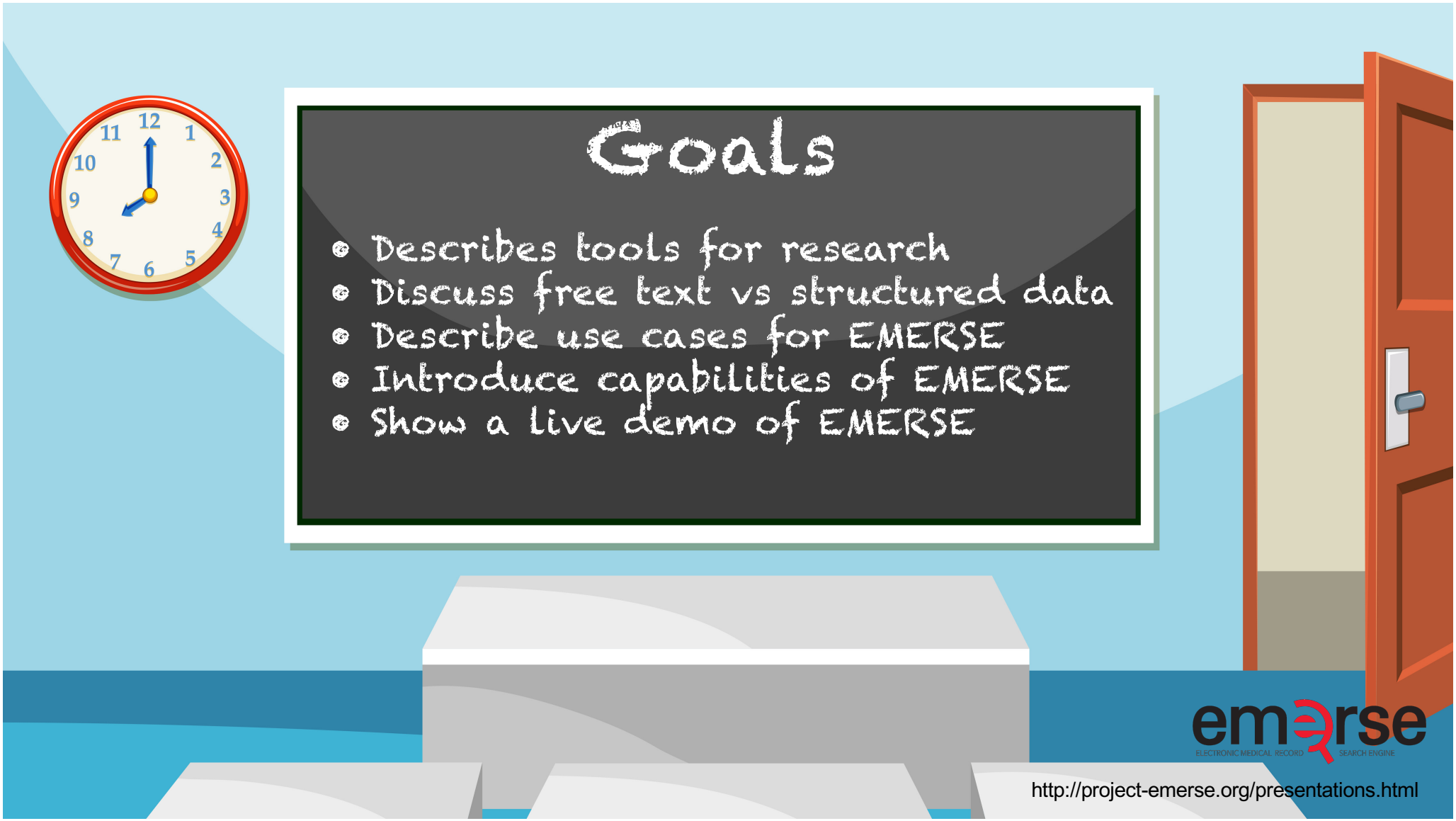
emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

*roadmap for today's presentation*

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html


Goals
Describes tools for research
Discuss free text vs structured data
Describe use cases for EMERSE
Introduce capabilities of EMERSE
Show a live demo of EMERSE
emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE
http://project-emerse.org/presentations.html

# What is EMERSE?

A search engine, kind of like Google, for free text medical record documents

(technically, it's an *information retrieval* system)

Simple to use

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# What is EMERSE?

A research tool, available at no cost

- requires local implementation
- picking the right tool for your research is important;  multiple tools are often needed

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Unstructured vs Structured Data

| EMERSE is for this… | …not this |
|---|---|
| *Unstructured Data (free text)* | *Structured Data* |
| Mrs. Jones is a 56 year old female with a history of HTN, hypercholesterolemia, and T2DM who comes to the clinic today with a 3 day h/o dizziness and severe headache on the left side. | WBC: 5.6<br>Total cholesterol: 182<br>Weight: 67.4<br>AST: 30<br>ALT: 52 |

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# **80%** of EHR data are in unstructured free text

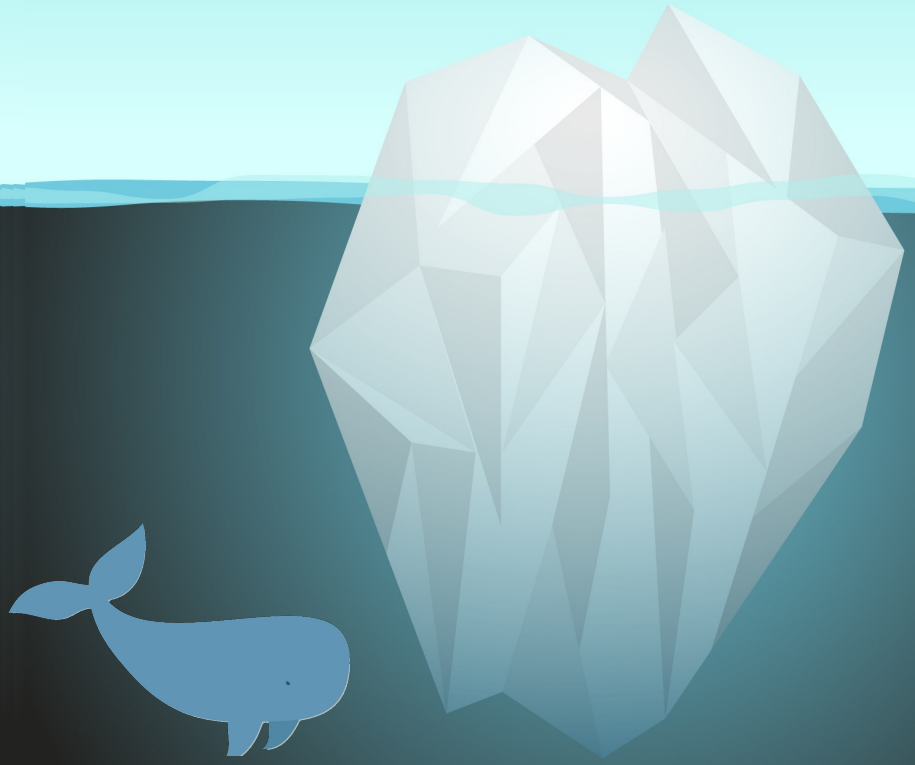
structured data

unstructured data/free text

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# **80%** of EHR data are in unstructured free text



"potential cases may be missed or wrongly dated when coded data alone are used..” (PMID: 23964710)

"structured data is best used for first pass filtering of EHR data…prior to a more nuanced second pass using unstructured data" (PCMID: PMC4333685)

"...urinary incontinence, an important patient-centered outcome following prostate cancer treatment, was reported almost exclusively in the free text of EHRs and was rarely coded as an ICD 9 diagnosis code.” (PMCID: PMC4899050)

"Claims and structured EHR data give an incomplete picture of burden related to geriatric syndromes. Geriatric syndromes are likely to be missed if unstructured data are not analyzed." (PMID: 29972595)

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# What isn't EMERSE?

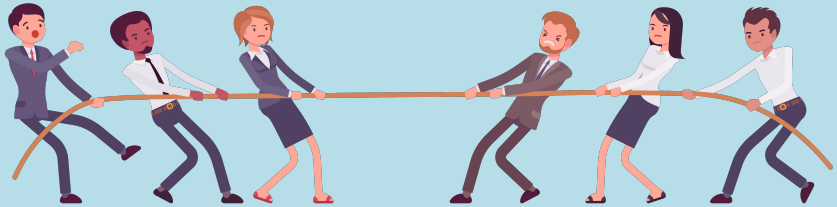
It is not: natural language processing (NLP)

- but full NLP may not always be needed
- and NLP can be hard to use or customize

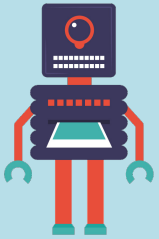
emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# EMERSE vs NLP



EMERSE: *augmented* intelligence

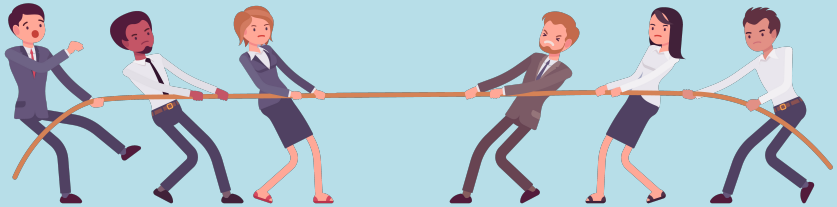


NLP: *artificial* intelligence

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# EMERSE vs NLP



NLP

EMERSE

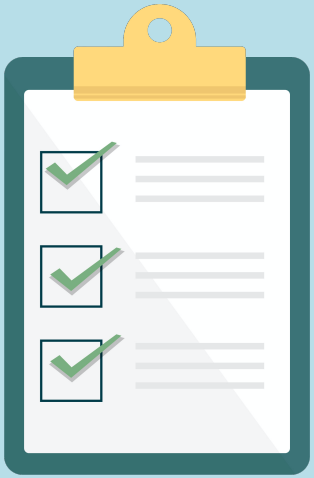
emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Find cohorts

EMERSE allows you to find cohorts based on things mentioned in the notes

- diseases
- drugs
- symptoms
- anything*



*if it is mentioned

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Find cohorts

It’s perfect for finding rare things…

…like rare cancers

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Find cohorts

It's perfect for finding diagnoses that don't have a specific ICD-10 code… like leiomyosarcoma…

**ICD-10-CM Diagnosis Code C49.6** [convert to ICD-9-CM]

Malignant neoplasm of connective and soft tissue of trunk, unspecified

Malignant neoplasm of conn and soft tissue of trunk, unsp; Angiosarcoma, trunk; Cancer of the soft tissue, trunk; Leiomyosarcoma, trunk; Liposarcoma, trunk; Malignant fibrous histiocytoma, trunk; Primary angiosarcoma of trunk; Primary leiomyosarcoma of trunk; Primary liposarcoma of trunk; Primary malignant fibrous histiocytoma of trunk; Primary malignant neoplasm of soft tissues of trunk; Primary rhabdomyosarcoma of trunk; Primary sarcoma of trunk; Rhabdomyosarcoma, trunk; Sarcoma, trunk; Malignant neoplasm of back NOS

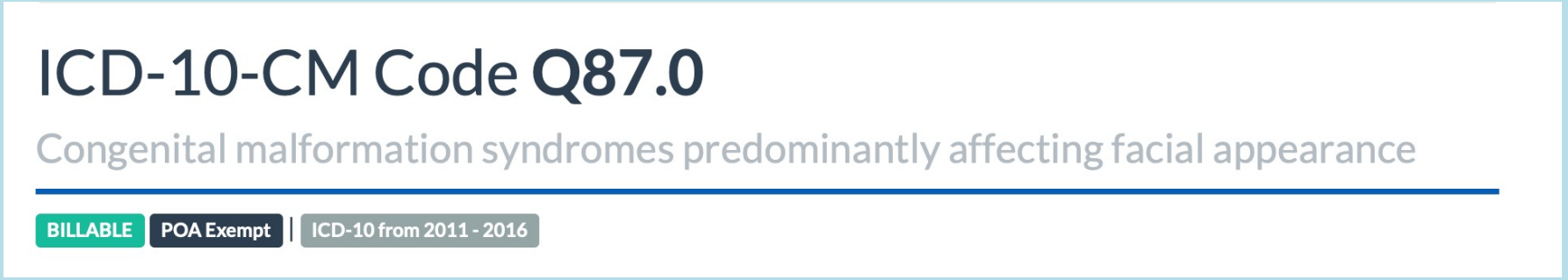
https://www.icd10data.com/search?s=leiomyosarcoma

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Find cohorts

…or Bainbridge-Ropers Syndrome




…or hundreds of other disorders with no specific code

https://www.icd10data.com/search?s=leiomyosarcoma

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Working with cohorts

find a cohort from within EMERSE

upload a list of patients

import list from another system

EMERSE patient lists
(from 1 to 100,000 patients per list)

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Highlight documents for chart review

Thoracocentesis confirmed the recurrence of mantle cell lymphoma. Disease restaging work-up revealed multicompartment lymphadenopathy in the neck, mediastinal, retrocrural, retroperitoneal and pelvic regions. Bone marrow was also involved. The patient was treated with a total of six cycles of rituximab, cyclophosphamide, vincristine, doxorubicin and dexamethasone (R-HyperCVAD) completed in January 2007. That treatment led to complete remission that lasted until October 2008, when the disease was found to have recurred in the left pleural space and retroperitoneum without bone marrow involvement.

https://jmedicalcasereports.biomedcentral.com/articles/10.1186/1752-1947-4-329

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

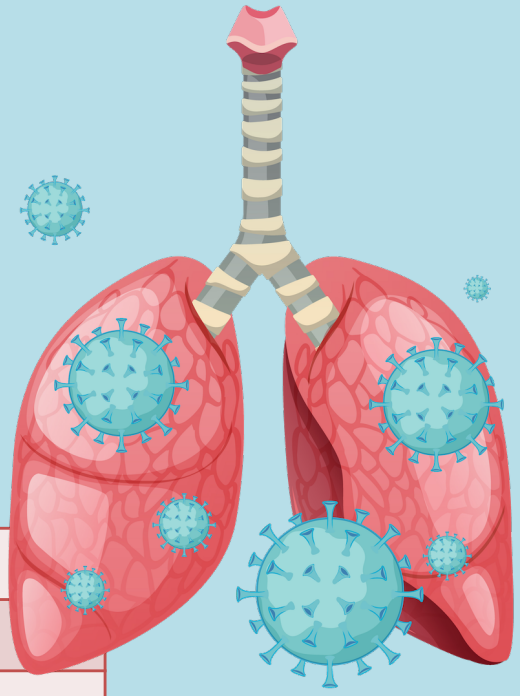
http://project-emerse.org/presentations.html

MICHIGAN MEDICINE
UNIVERSITY OF MICHIGAN

# Chart review

Research team at Michigan Medicine used EMERSE to identify presenting symptoms for COVID-19, including:

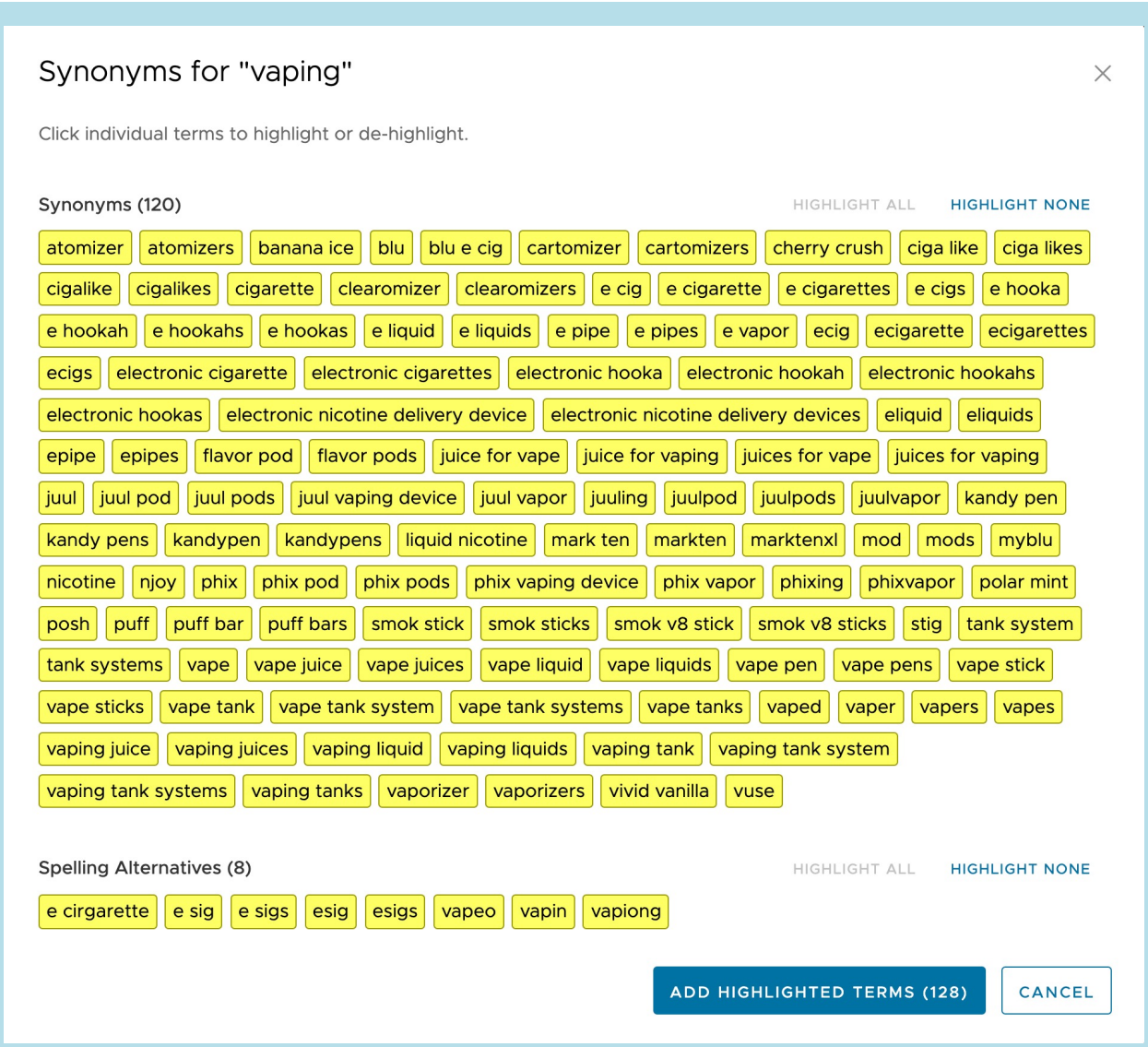
| | | |
|---|---|---|
| dyspnea | rhinorrhea or congestion | vomiting |
| wheezing | myalgia or arthralgia | abdominal pain |
| cough | fatigue or malaise | chest pain or tightness |
| sputum production | headache | loss of smell or taste |
| blood in sputum | loss of appetite | altered mental status |
| sore throat | diarrhea | weakness |
| fever or chills | nausea | lightheadedness or syncope |



emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

https://pubmed.ncbi.nlm.nih.gov/33046294/

http://project-emerse.org/presentations.html




# EMERSE query expansion

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# EMERSE query expansion

- Can use “EMERSE Synonyms Collection” or administrators can upload additional datasets (e.g., ICD-10, human phenotype ontology, MeSH terms)

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

# EMERSE Synonyms Collection

- Manually curated over past 14+ years
- Terms/phrases not found in standard vocabularies, taxonomies, ontologies (but found in clinical notes)
- Academic and non-academic licensing available
- > 1.2 million unique rows
- Details are at:

  http://project-emerse.org/synonyms.html
- To explore a subset, visit:

  http://www.davidhanauer.com/EMERSE_synonyms_T_Terms/

# EMERSE Synonyms

| Category | Example |
|---|---|
| Acronyms | CXR<br>chest x-ray |
| Abbreviations | chemotherapy<br>chemo |
| Professional and consumer terms | emesis<br>puking |
| Misspellings | diarriea<br>diahhrea |
| Trade and generic drug names | Tylenol<br>acetaminophen |
| Chemotherapy regimens | FOLFOX, folinic acid, fluorouracil, oxaliplatin |
| Word Stems | sleep, sleeping, sleepy |

| Category | Example |
|---|---|
| True synonyms | frigid<br>cold |
| Phrasing variations | lido w/ epi<br>epi-lidocaine |
| Malapropisms/ usage errors | beast mass<br>prostrate cancer |
| Root word variations | adenotonsillectomy<br>tonsilloadenoidectomy |
| Hyphenations | pseudosarcomatous<br>pseudo-sarcomatous |
| Idioms | under the knife<br>under the weather |
| Neologisms | trabecularity<br>(vs trabeculation) |

http://project-emerse.org/presentations.html

# EMERSE Bundles

Saved sets of search terms

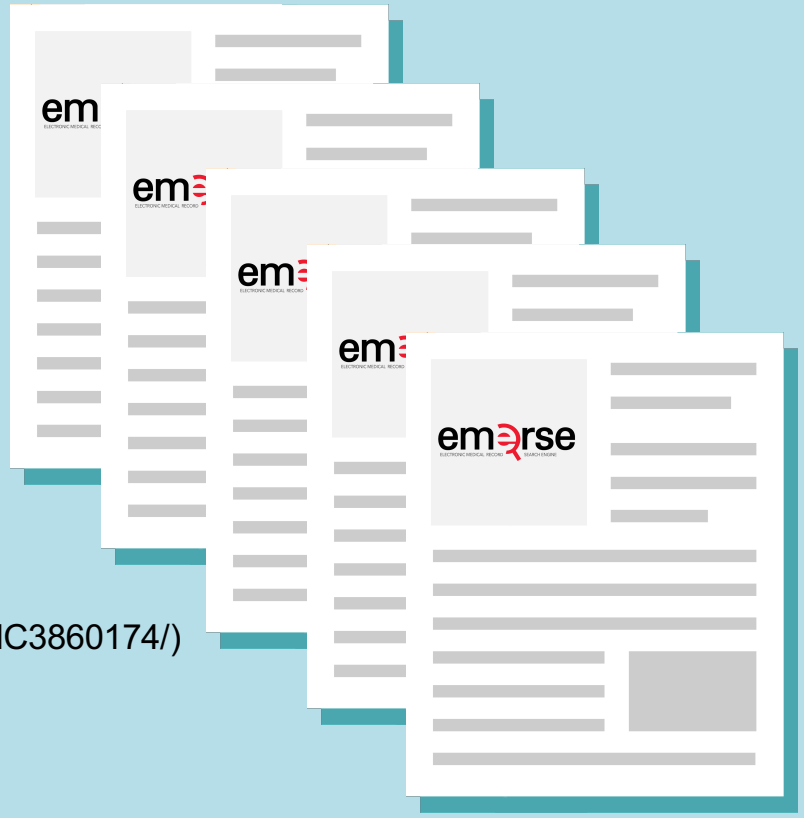
Shared among users

Helps standardize searches

Supports reproducibility

- Some teams publish Bundles in their publication's appendix

(example: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3860174/)




http://project-emerse.org/presentations.html

# EMERSE is *fast*

| Query to identify all patients with the following | Epic Clarity time (s) | EMERSE time (s) | EMERSE advantage |
|---|---|---|---|
| cavernous hemangioma | 14,652 | 2 | 7,320x |
| gray platelet syndrome | 14,940 | 2 | 7,470x |
| inferior lingular segment of the left upper lobe | 17,784 | 9 | 1,980x |

…enabling real-time querying

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# What's in EMERSE?

Free text notes*

- admission notes
- history and physical
- progress notes
- pathology reports
- radiology reports

*depends on local configuration

@ Michigan: 2.6+ million patients
210+ million notes
1995-present

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# What's not in EMERSE?

Structured data

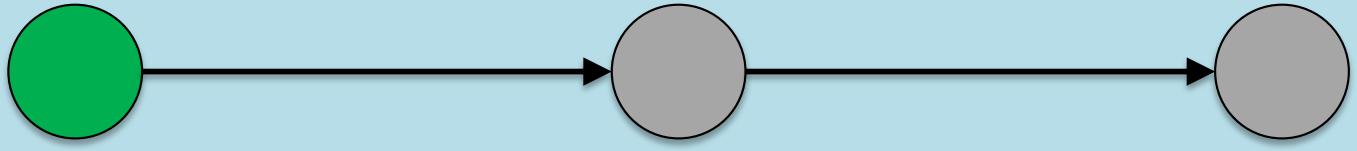
- discrete lab values
- vital signs
- billing codes

Other tools exist for structured data (DataDirect, i2b2, Leaf,  etc.)

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Typical workflow



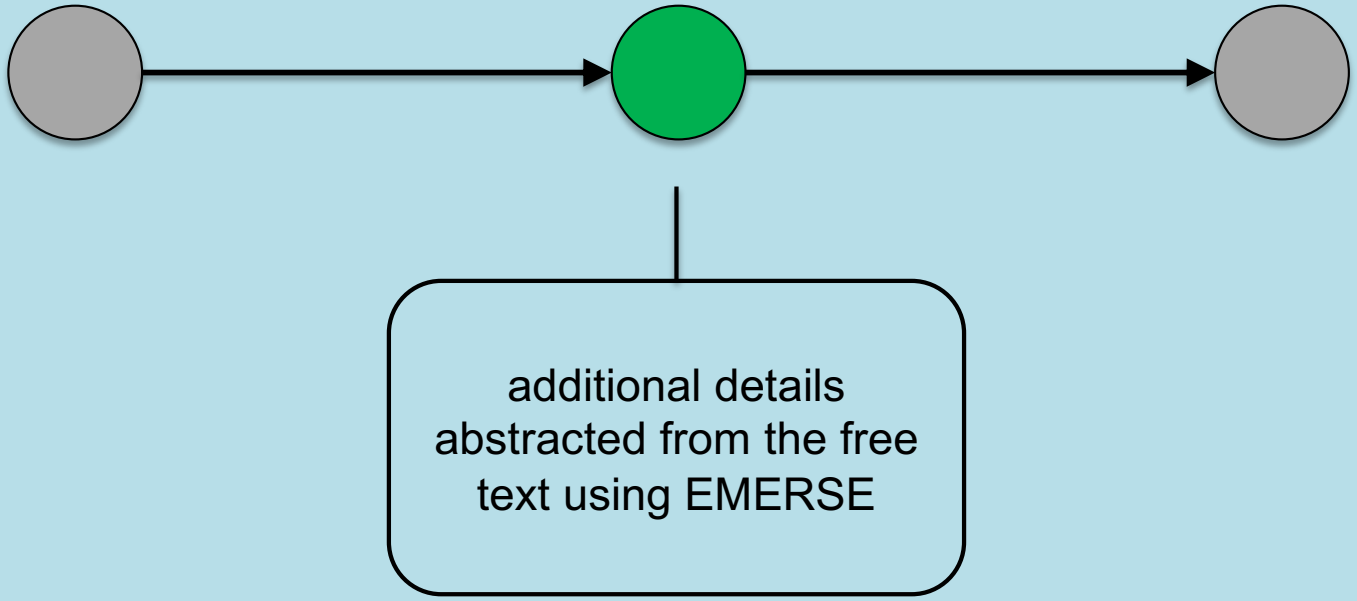
initial cohort identified with a cohort discovery tool using structured data (or via EMERSE 'Find Patients')

Cohort discovery tools:
DataDirect @ Michigan
i2b2, LEAF, others elsewhere

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

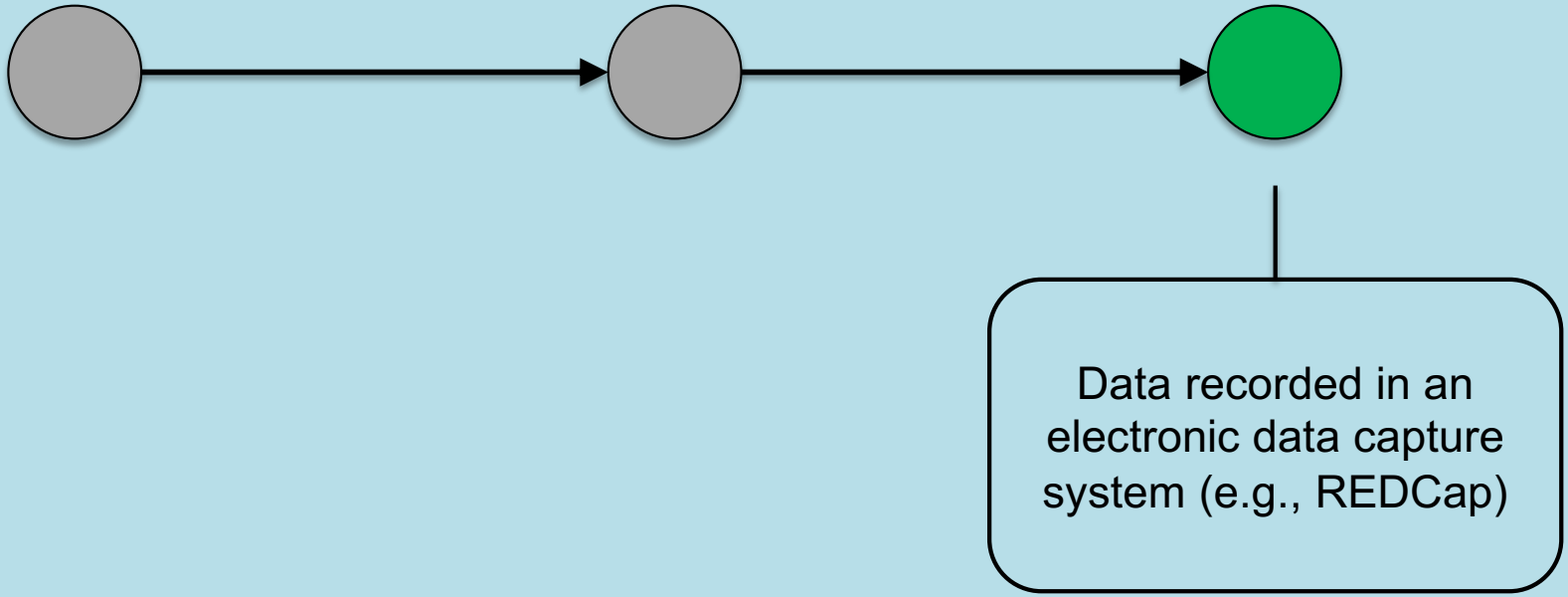
# Typical workflow




emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Typical workflow




emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Publications using EMERSE

**446** papers

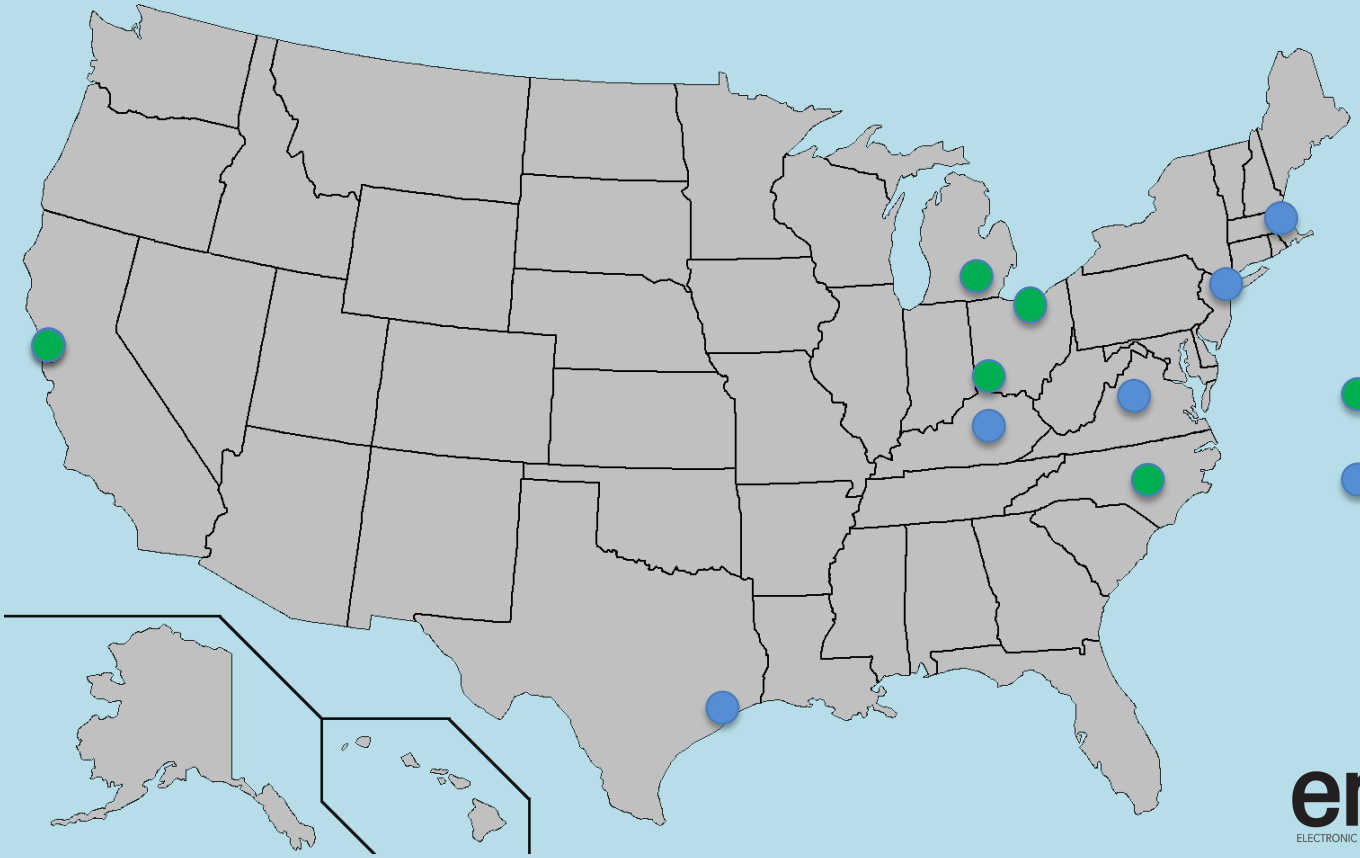
**81** abstracts

Full list at:

http://project-emerse.org/publications.html

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

# Where is EMERSE?



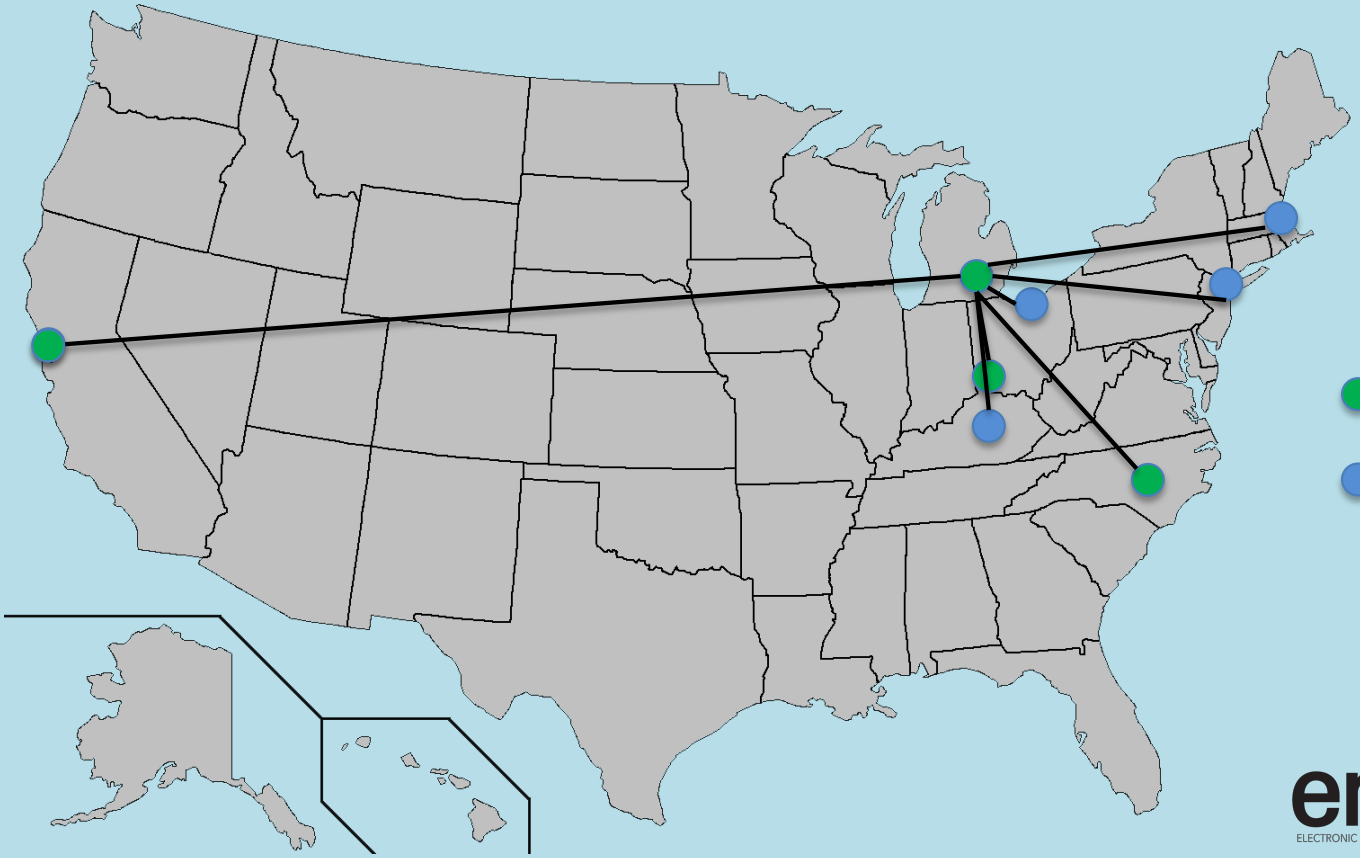
installed

installing

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# The future…




emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Who can obtain access?

**@ Michigan**

Anyone with EHR access, BUT:

- you need a valid reason such as research, operations, quality
- “curiosity” is not a valid reason*
- all usage is tracked

**@ elsewhere**

Will depend on each site’s local rules

*but we do have a demo site for curious users

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Getting EMERSE

local installation required

need support of IT team

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Under the Hood

- EMERSE application (Java, Javascript)
- Apache Solr
- Database

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

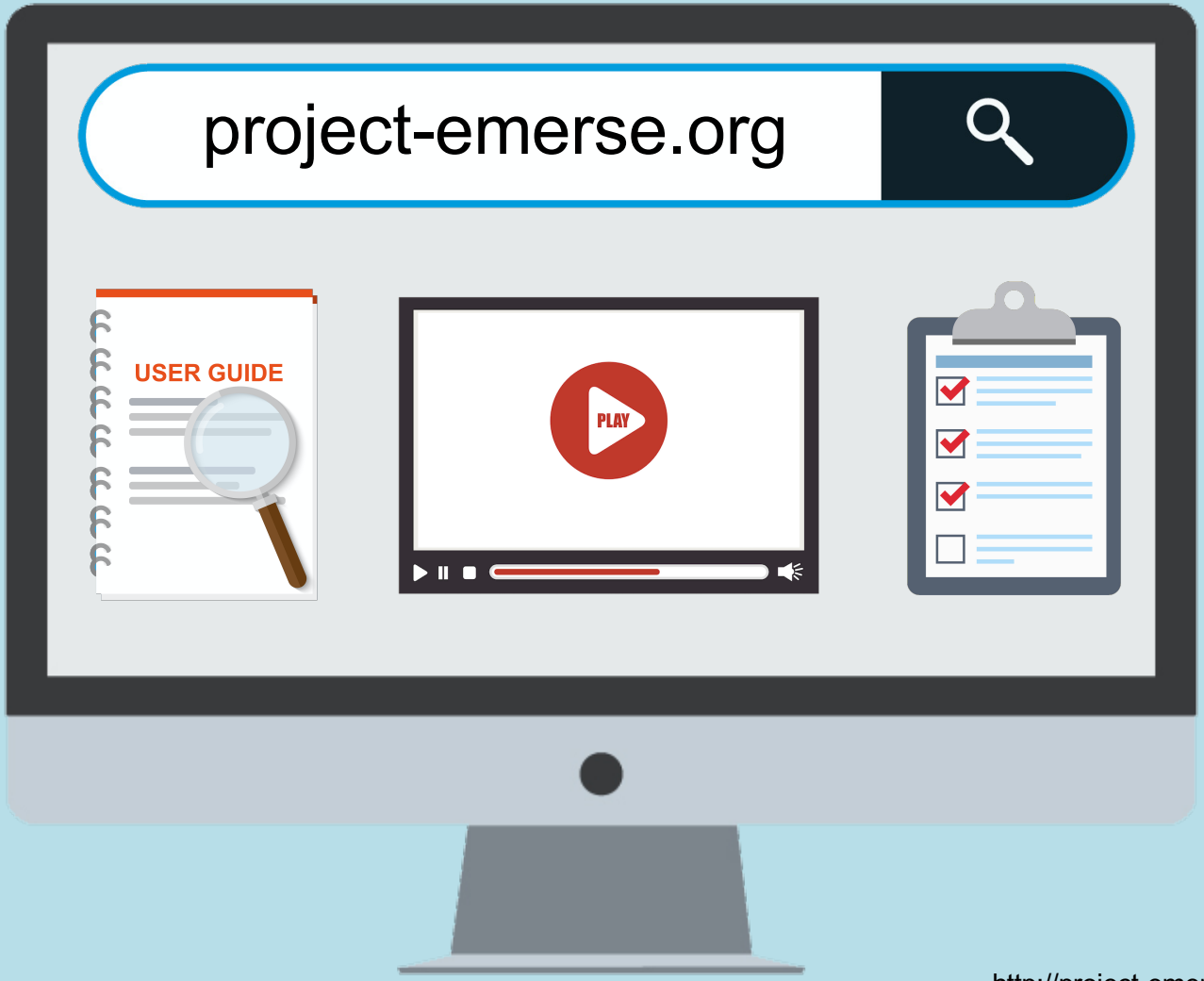
# Test driving EMERSE

online demo (contact us)

videos on website

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html




emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

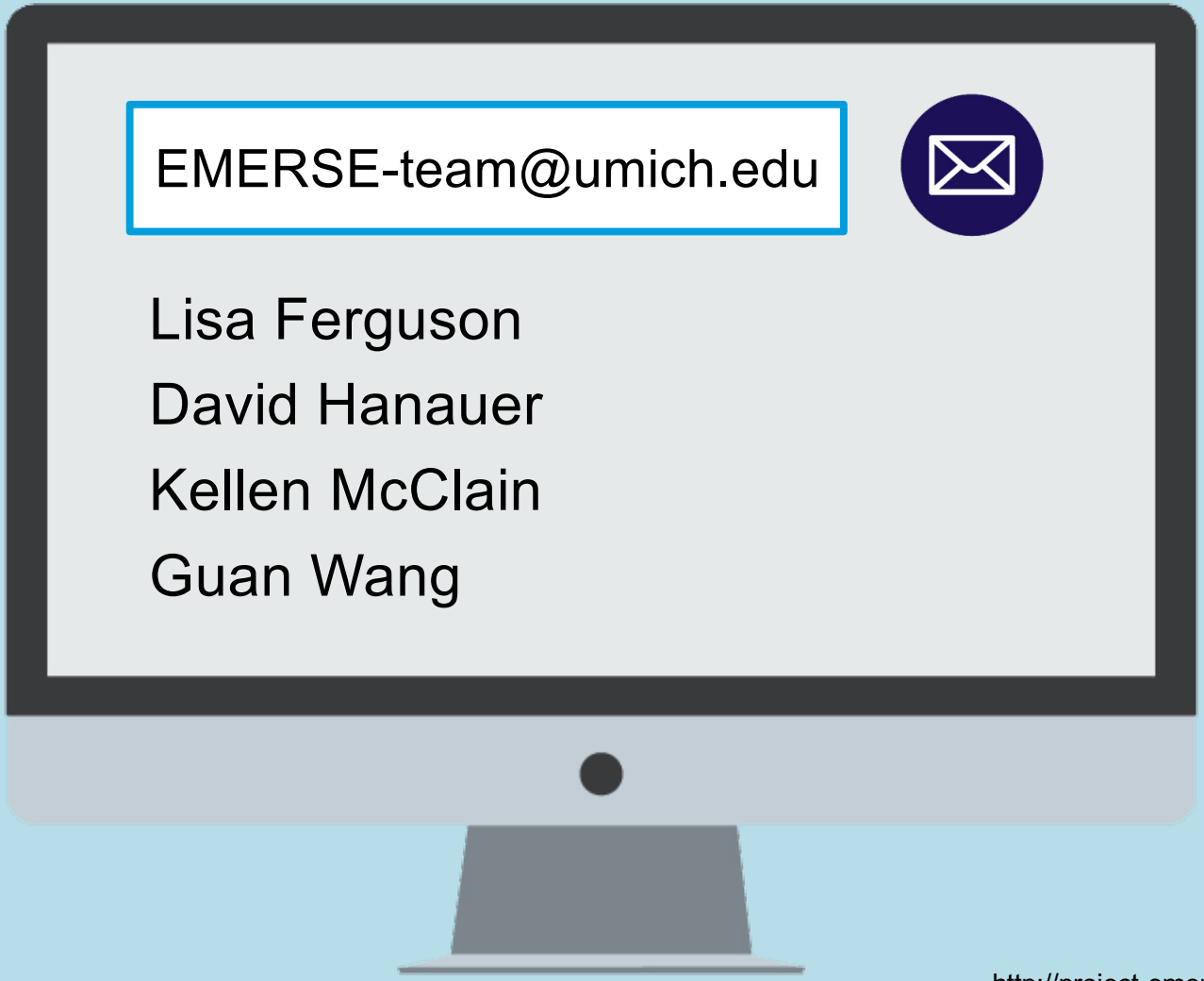
http://project-emerse.org/presentations.html

Twitter: @projectEMERSE

publications
software releases
announcements
webinars

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html




emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Acknowledgements

U of Michigan Rogel Cancer Center

U of Michigan Institute for Clinical and Health Research (MICHR)

U of Michigan Health Information and Technology Services (HITS)

NCI Information Technology for Cancer Research (ITCR)

U of Michigan Department of Learning Health Sciences

U of Michigan Office of Research

Our Partner Institutions

Our Users

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

*roadmap for today's presentation*

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html

# Live demonstration

- 10,000 simulated patients
- 500,000 simulated documents
- Look and options will differ depending on version and local setup (some components may not be available to all users/sites)

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

http://project-emerse.org/presentations.html