EMERSE: A practical tool for unlocking the free text in clinical notes for research and more

emerse

ELECTRONIC MEDICAL RECORD SEARCH ENGINE

David Hanauer, MD, MS

University of Michigan

hanauer@umich.edu

Twitter: @DavidHanauer2

CTSI-CN Informatics Seminar
August 4, 2020

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", for which David Hanauer is entitled to a portion of the license fees

About me

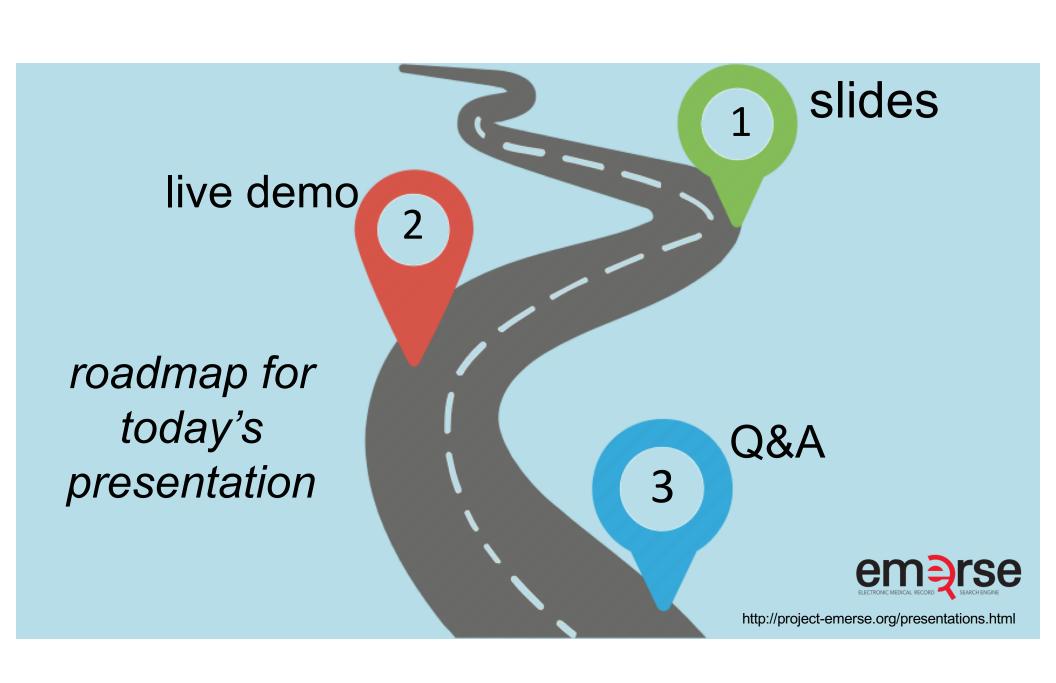My work focuses on clinical informatics

I have tweeted only one time

I do my own research and I support the research through our CTSA

I once published a paper that earned me a 2014 "Ig Nobel" award

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0070585

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

slides
1

live demo
2

roadmap for today's presentation

Q&A
3

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
- Explain how access is obtained
- Show a live demo of EMERSE

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

What is EMERSE?

A research tool, available at no cost

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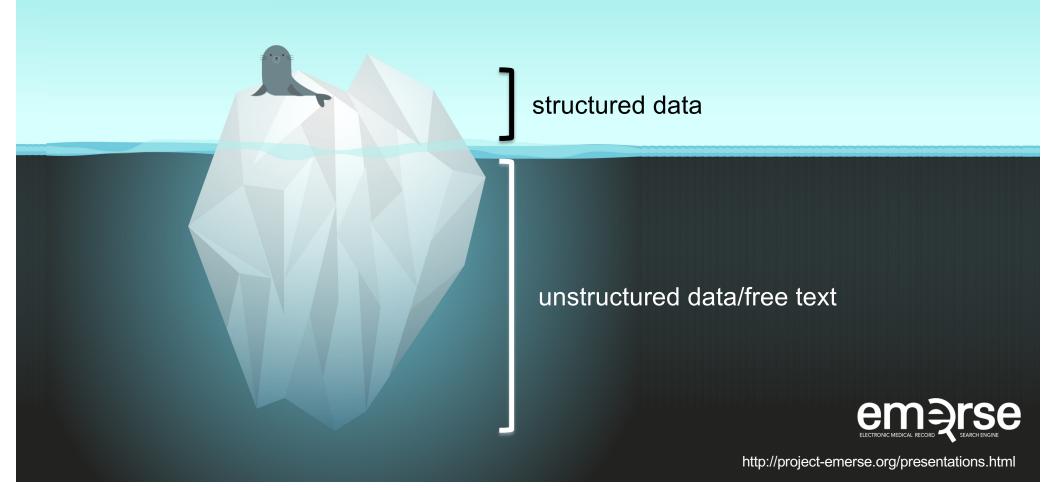- – picking the right tool for your research is important; multiple tools are often needed

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 Total cholesterol: 182 Weight: 67.4 AST: 30 ALT: 52

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

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

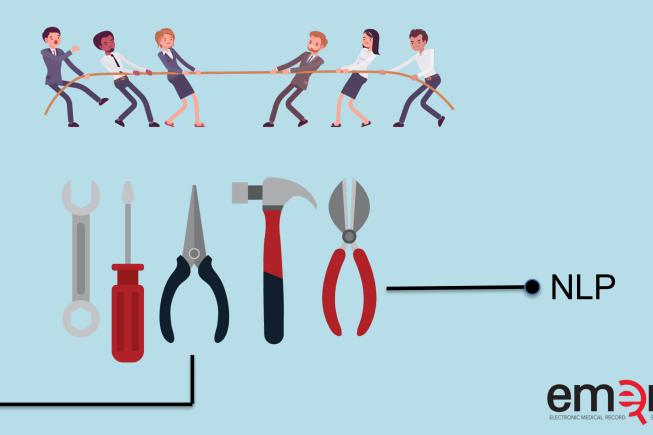# EMERSE vs NLP

EMERSE: *augmented* intelligence

NLP: *artificial* intelligence

EMERSE vs NLP

NLP

EMERSE

Find cohorts

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

– diseases

– drugs

– symptoms

– anything!

Find cohorts

It's perfect for finding rare things…

…like rare cancers

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
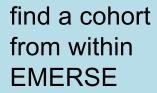
ICD-10-CM Code Q87.0

Congenital malformation syndromes predominantly affecting facial appearance

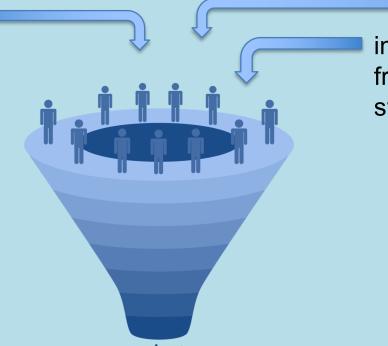BILLABLE | POA Exempt | ICD-10 from 2011 - 2016

…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
(up to 100,000 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.

em𝐞rse
ELECTRONIC MEDICAL RECORD   SEARCH ENGINE

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

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

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

COVID-19 terms

coronavirus disease 2019
coronavirus disease 19
coronavirus disease
COVID-19 disease
COVID19 disease
COVID disease
nCoV-2019
nCoV-19
nCoV19
19-nCoV
COVID-19 novel
COVID19 novel
COVID novel
coronavirus COVID-19
coronavirus COVID-2019
coronavirus COVID19
coronavirus COVID2019
coronavirus COVID
COVID-19 coronavirus
COVID19 coronavirus
COVID-2019 coronavirus
COVID2019 coronavirus
COVID coronavirus
COVID-19 virus
COVID19 virus
COVID-2019 virus
COVID2019 virus
COVID virus
SARSCoV-2
SARSCoV2
SARSCoV

novel coronavirus (COVID-19)
novel coronavirus (COVID19)
novel coronavirus (COVID)
novel coronavirus 2019
novel coronavirus
novel 2019 coronavirus
novel 2019 CoV
new coronavirus
new 2019 coronavirus
new 2019 CoV
2019 novel coronavirus
2019 novel CoV
2019-nCoV acute respiratory disease
19 novel coronavirus acute respiratory
19 novel coronavirus
19 novel CoV
19 coronavirus acute respiratory disease
19-nCoV acute respiratory disease
nCoV acute respiratory disease
SARS-CoV-2
SARS-CoV2
SARS-CoV
SARS-CoV-19
SARS-CoV-2019
2019 coronavirus-2
2019 coronavirus
19 coronavirus-2
19 coronavirus
SARS-coronavirus 2
SARS-coronavirus
Corona Virus Disease 19

Corona Virus Disease 2019
COVID-19 infections
COVID19 infections
COVID-19 infection
COVID19 infection
COVID-19
COVID19
CV19
CV-19
COVID-2019 infections
COVID2019 infections
COVID-2019 infection
COVID2019 infection
COVID-2019
CV2019
CV-2019
2019-nCoV
Wuhan coronavirus
Wuhan virus
coronavirus
corona-virus
Chinese virus
Chinese coronavirus
COVID2019
corona 2019
novel corona
U07.1
CORVID-19 [spelling alternative]
CORVID19 [spelling alternative]
CORVID [spelling alternative]
CORVIN-19 [spelling alternative]

CORVIN19 [spelling alternative]
CORVIN [spelling alternative]
COVIN-19 [spelling alternative]
COVIN19 [spelling alternative]
COVIN [spelling alternative]
COAVID-19 [spelling alternative]
COAVID19 [spelling alternative]
COAVID [spelling alternative]
COVD-19 [spelling alternative]
COVD19 [spelling alternative]
COVD [spelling alternative]
CAVIN-19 [spelling alternative]
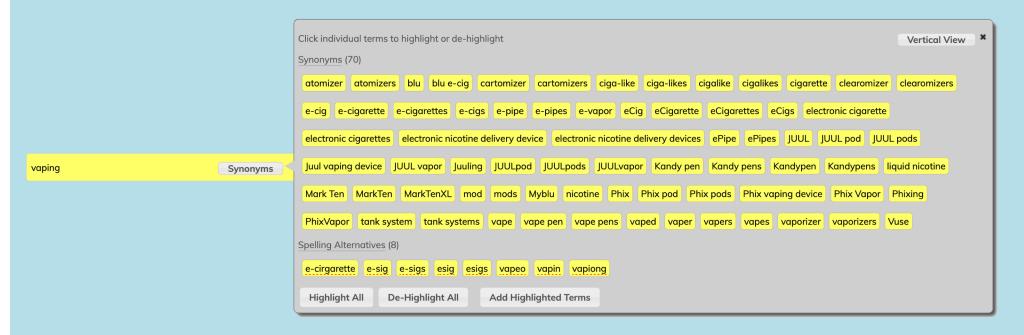CAVIN19 [spelling alternative]
CAVIN [spelling alternative]
novel coronvirus [spelling alternative]
CODIV-19 [spelling alternative]
CODIV19 [spelling alternative]
CODIV [spelling alternative]
1COVID19 [spelling alternative]

More at:
http://project-emerse.org/synonyms_covid19.html

emerse
ELECTRONIC MEDICAL RECORD   SEARCH ENGINE

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

EMERSE query expansion

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

EMERSE query expansion

- Can use "EMERSE Synonyms" or upload your own dataset (e.g., ICD-10, human phenotype ontology, MeSH terms)

- Simple file format
  - http://project-emerse.org/documentation/administrator_guide.html#truepreparing-a-synonyms-file

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

EMERSE Synonyms

- Manually curated over past 12+ years
- Terms/phrases not found in standard vocabularies, taxonomies, ontologies (but found in clinical notes)
- Academic and non-academic licensing available*
- > 1.1 million unique rows

  *David Hanauer is entitled to a portion of the license fees

- Details are at:

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

- To explore a subset, visit:

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

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

EMERSE Synonyms

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

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

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

What's in EMERSE?

Free text notes

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

@ Michigan: 2.5+ million patients
                170+ million notes
                1995-present

What's not in EMERSE?

Structured data
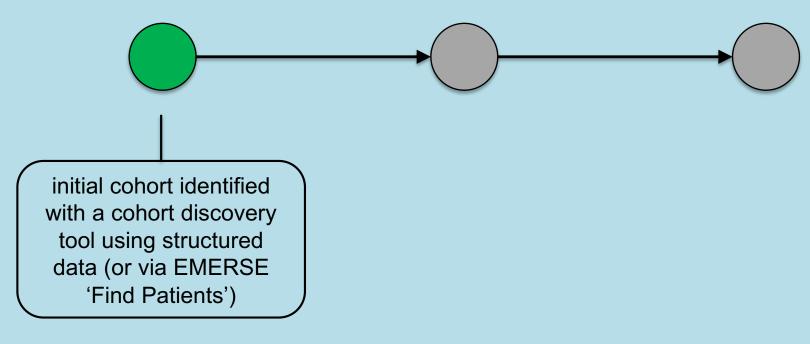
- discrete lab values

- vital signs

- billing codes

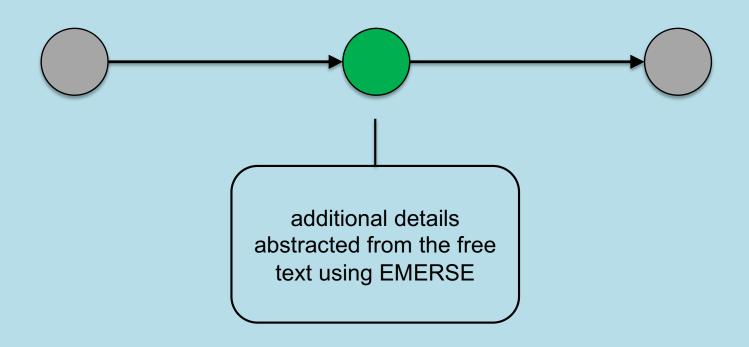Other tools exist for
structured data (DataDirect, i2b2, etc.)

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

additional details abstracted from the free text using EMERSE

Typical workflow

Data recorded in an electronic data capture system (e.g., REDCap)
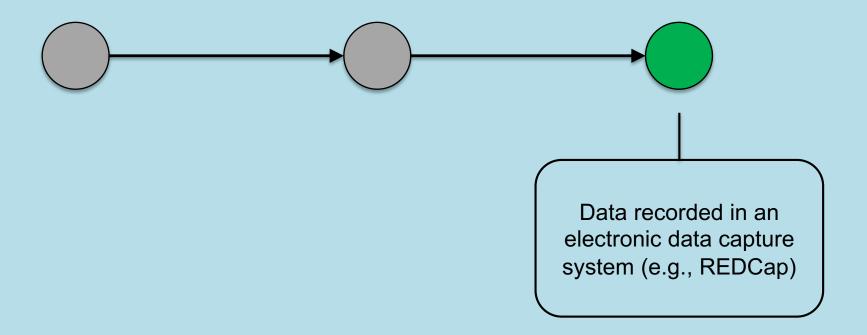
Publications using EMERSE

354 papers

54 abstracts
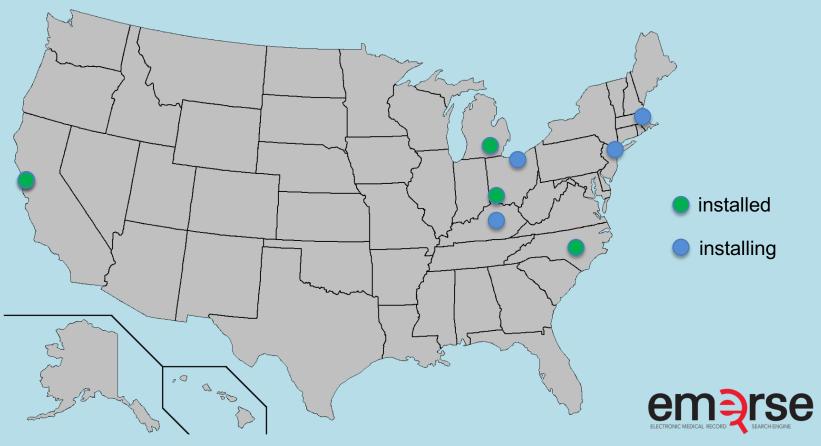
Full list at:
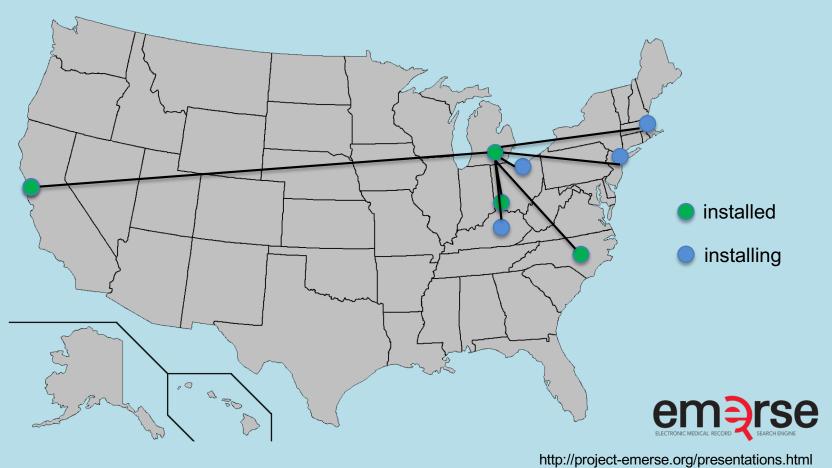
http://project-emerse.org/publications.html

Where is EMERSE?

installed

installing

The future…

● installed
● installing

Who can obtain access?
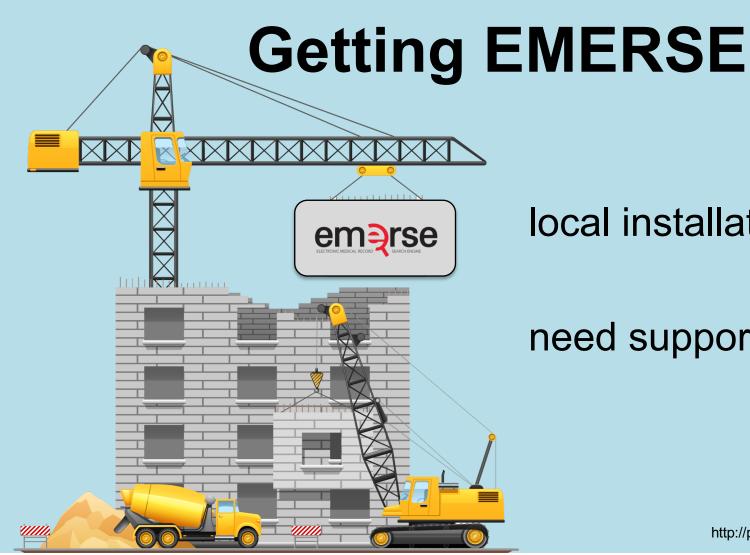
@ Michigan

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

emerse
ELECTRONIC MEDICAL RECORD   SEARCH ENGINE

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

Getting EMERSE

local installation required

need support of IT team

Under the Hood

 EMERSE application
(Java, Javascript)

 Apache Solr

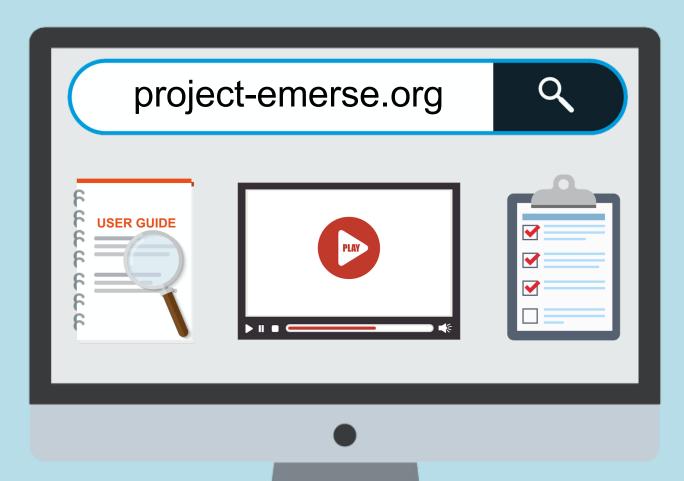 Apache ActiveMQ

 Oracle

Test driving EMERSE

online demo

videos on website

virtual machine
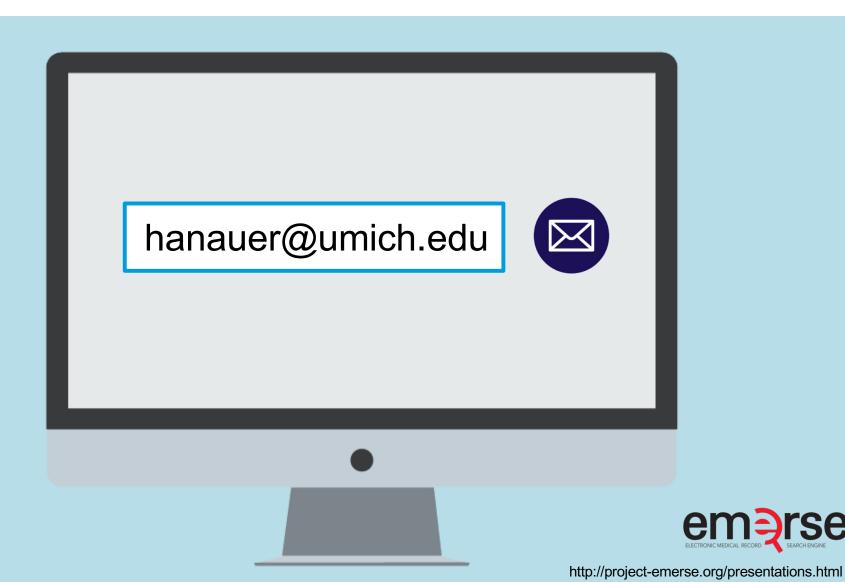
　- could be used for

individual projects

project-emerse.org

USER GUIDE

PLAY

emerse
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

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

hanauer@umich.edu