Using EMERSE to interrogate free text EHR data for COVID-19 and beyond



David Hanauer, MD, MS University of Michigan

hanauer@umich.edu

Twitter: @DavidHanauer2

Harvard Virtual Symposium June 11, 2020

If you're thinking of taking notes



these slides can be found at:

link will be on most slides



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



Designed to be self-serve



What is EMERSE?

- A research tool, available at no cost
 - Apache 2.0 licensing
- Picking the right tool for your research is important
 - multiple tools are often needed



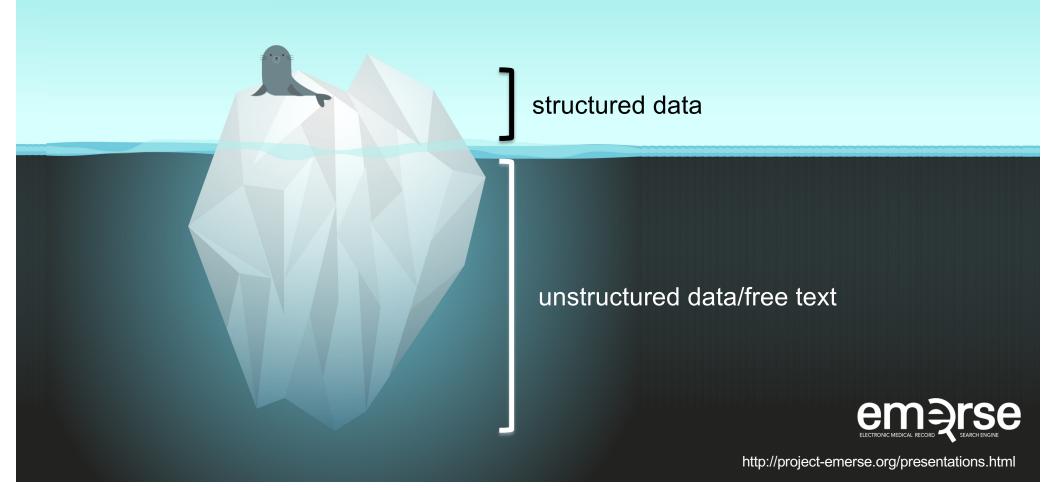


Unstructured vs Structured Data

EMERSE is for this	i2b2/4CE have foc	used on this
Unstructured Data (free text)	Structured Data	
Mrs. Jones is a 56 year old female with a h/o HTN, T2DM, and, and high cholesterol who comes to the clinic today with a 3 day h/o coughing, fever, and severe headache.	WBC: Total cholesterol: Weight: AST: ALT:	5.6 182 67.4 30 52



80% of EHR data are in unstructured free text



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)

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

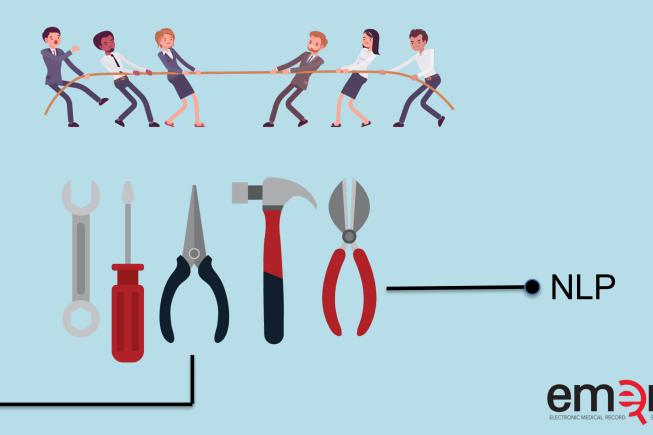


NLP: artificial intelligence



EMERSE vs NLP

EMERSE



With 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 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



Find cohorts

...or Bainbridge-Ropers Syndrome

ICD-10-CM Code **Q87.0**

Congenital malformation syndromes predominantly affecting facial appearance

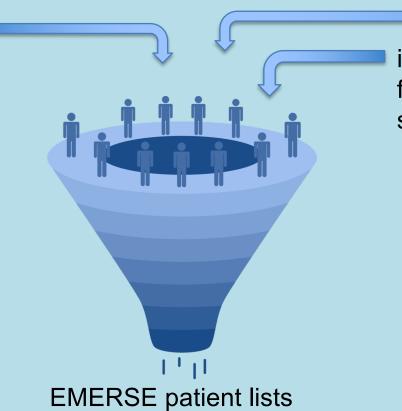
BILLABLE POA Exempt ICD-10 from 2011 - 2016

...or hundreds of other disorders with no specific code



Working with cohorts

find a cohort from within EMERSE



(up to 100,000 per list)

import list from another system (such as **i2b2**)

emerse search engine

upload a list

Highlight documents for chart review

A chest x-ray was obtained and revealed the presence of patchy consolidations in bilateral lower lobes. CT chest/abdomen/pelvis (Fig. 1) was notable for diffuse ground-glass opacities in the periphery of bilateral lungs as well as splenomegaly with a splenic hypodensity measuring 8.6 × 0.7 cm and surrounding peri-splenic edema, consistent with a splenic infarct. Hydroxychloroquine and a one-time dose of Tocilizumab were administered along with supportive measures. A continuous heparin infusion was also initiated and patient was admitted to the general medical ward for further workup.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7256550/





Chart review



Recommended on the 4CE Slack channel for validating the structured COVID-19 data.

Tools like EMERSE could help.





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	



Additional COVID studies using EMERSE



"Ocular Manifestations in COVID-19 Patients"





Additional COVID studies using EMERSE



"Abdominal CT in COVID-19 Patients: Incidence, Indications, and Findings"





COVID-19 terms

coronavirus disease 2019
coronavirus disease 19
coronavirus disease 19
coronavirus disease
COVID-19 disease
COVID19 disease
COVID disease
nCoV-2019
nCoV-19
nCoV-19
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
COVID-19 virus
COVID-2019 virus
COVID-2019 virus
COVID2019 virus
COVID2019 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

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 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

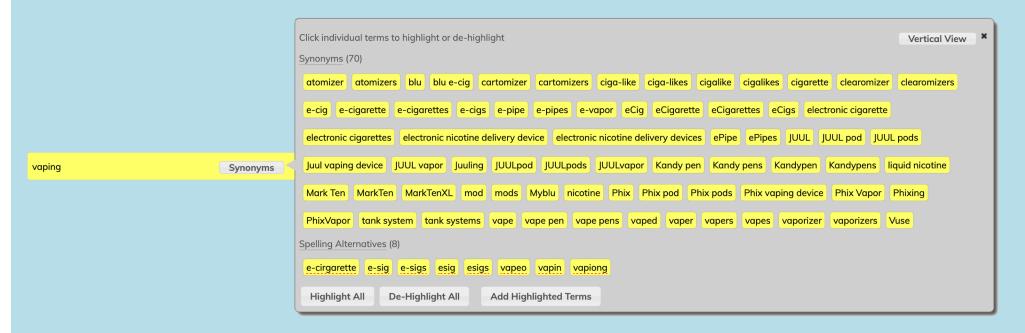
U07.1

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 query expansion





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 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/

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

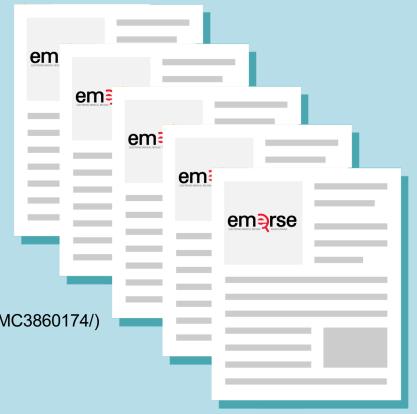
Category	Example
True synonyms	frigid cold
Phrasing variations	lido w/ epi epi-lidocaine
Malapropisms/ usage errors	beast mass prostrate cancer
Root word variations	adenotonsillectomy tonsilloadenoidectomy
Hyphenations	pseudosarcomatous pseudo-sarcomatous
Idioms	under the knife under the weather
Neologisms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

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/)



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

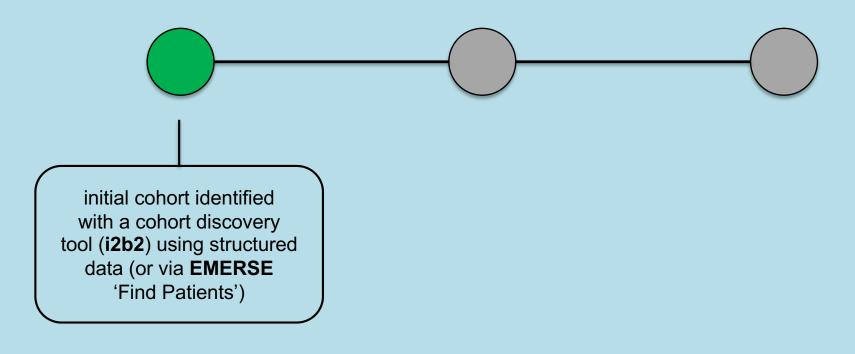
- discrete lab values
- vital signs
- billing codes

Other tools such as **i2b2** exist for structured data



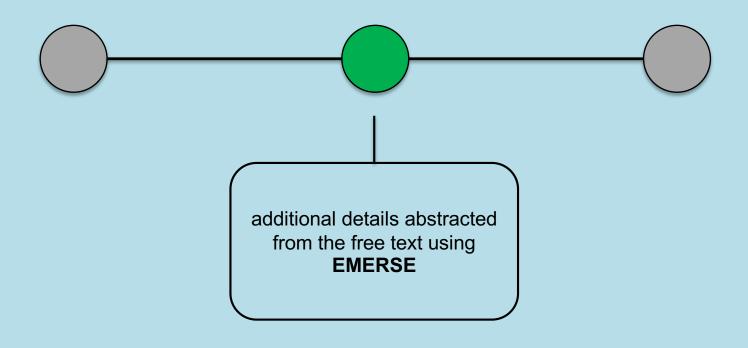


Typical workflow



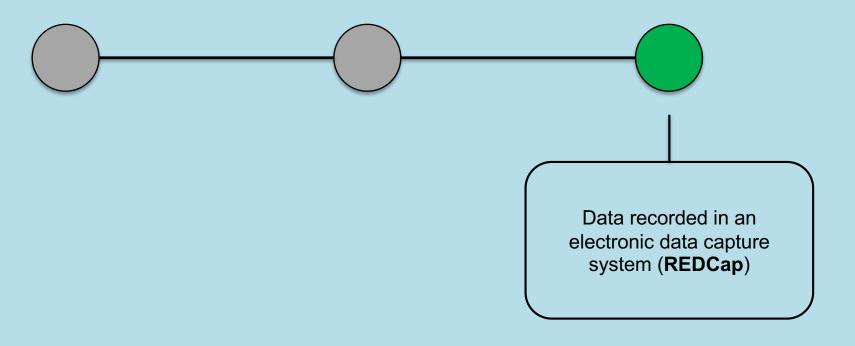


Typical workflow





Typical workflow





EMERSE and i2b2 could go together like...



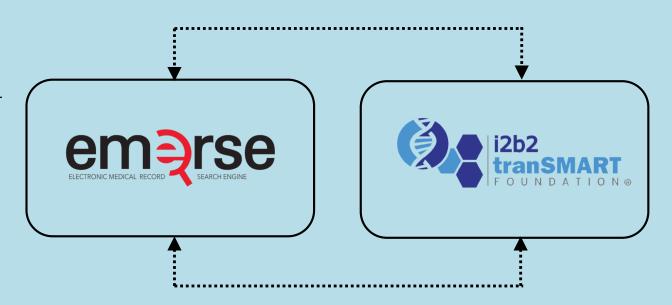




EMERSE and i2b2 could go together like...

i2b2 queries EMERSE/Solr API for list of patients as additional criteria e.g.: structured data & note contains phrase: _____

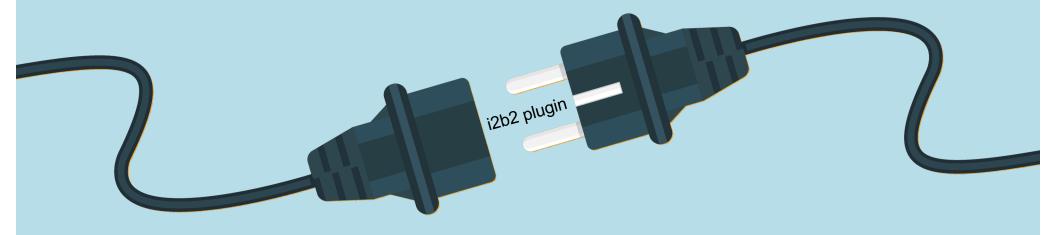
i2b2 sends patient cohort to EMERSE (this has been done before)



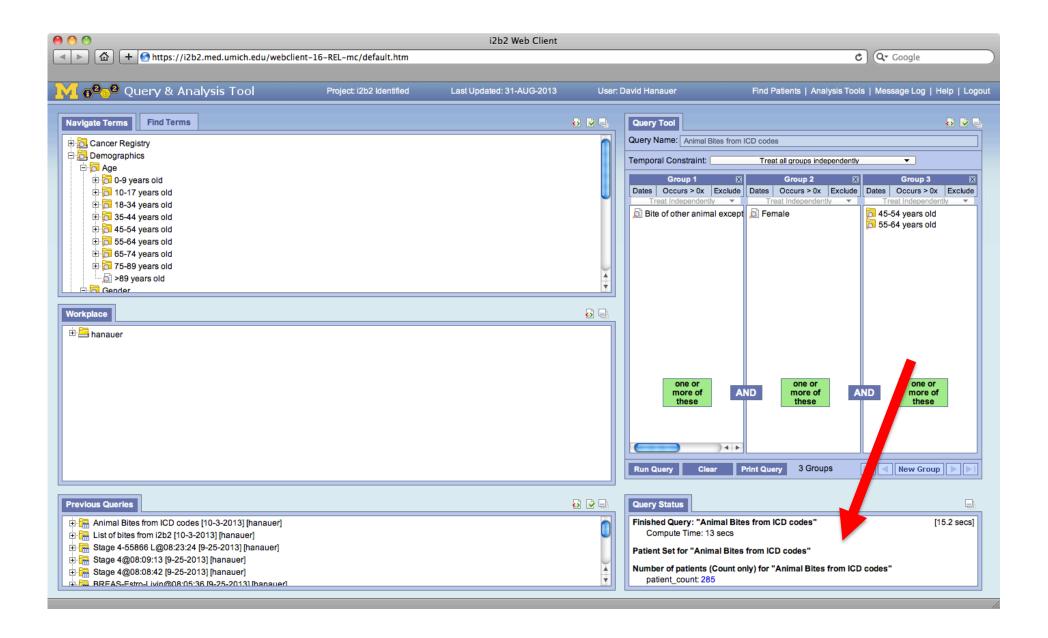
EMERSE queries i2b2 to narrow cohort using structured data

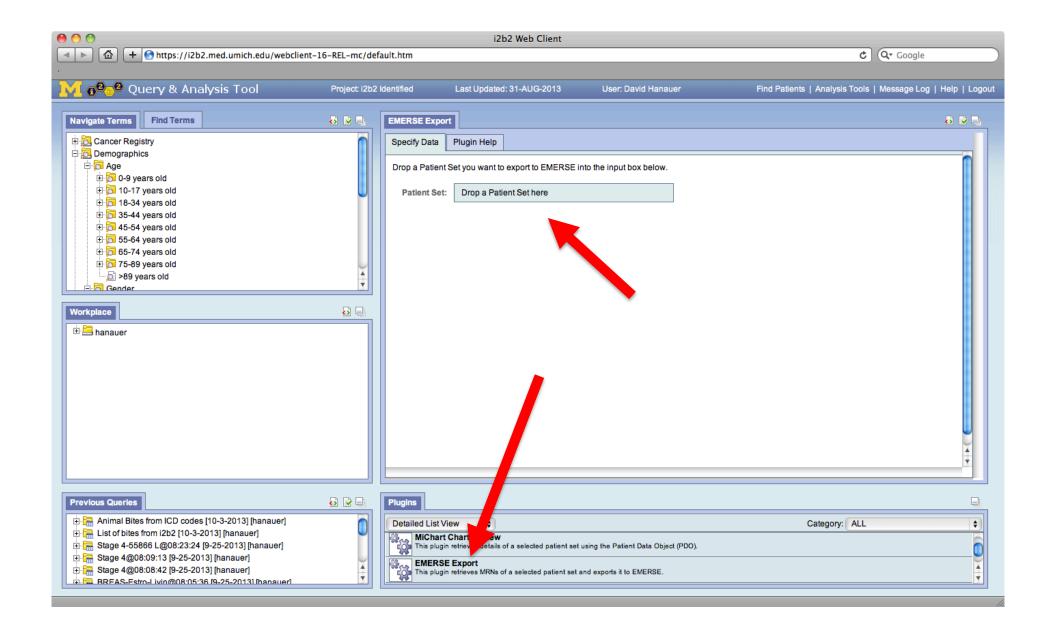


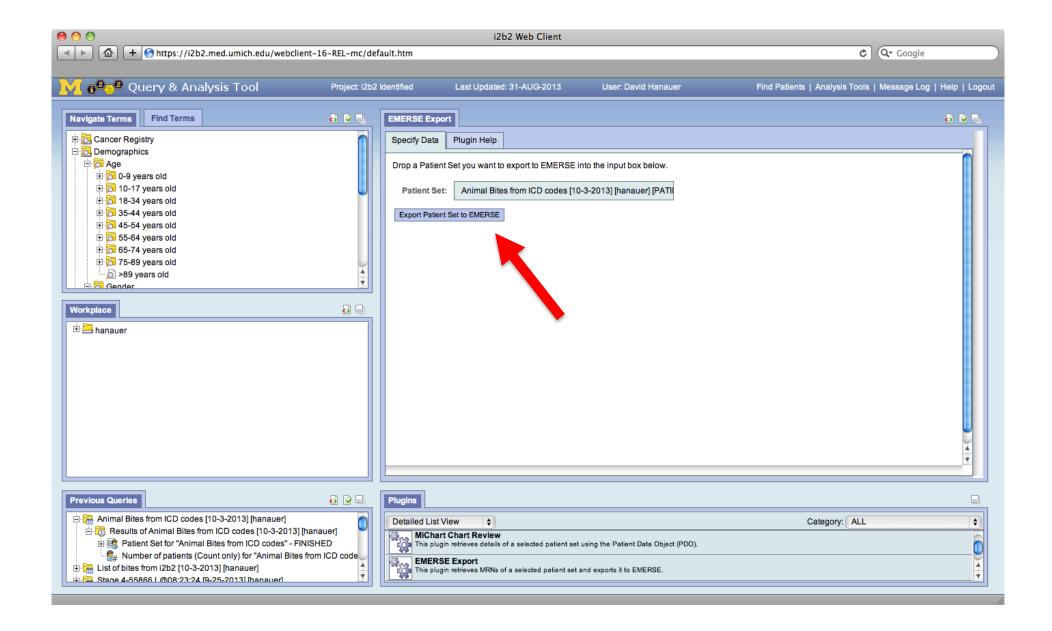
EMERSE and i2b2 have been connected in the past

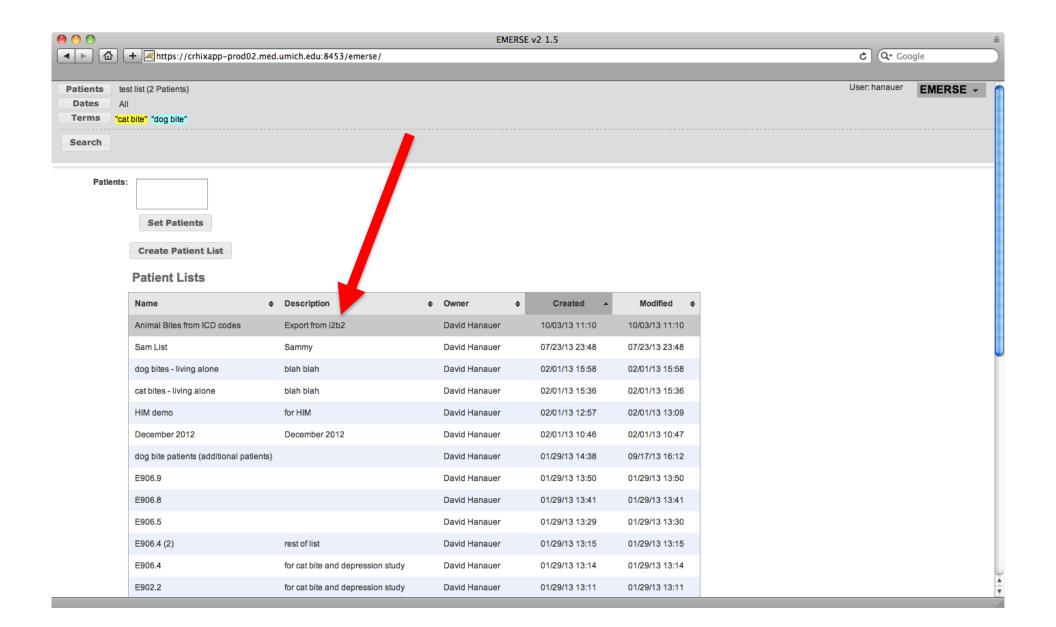


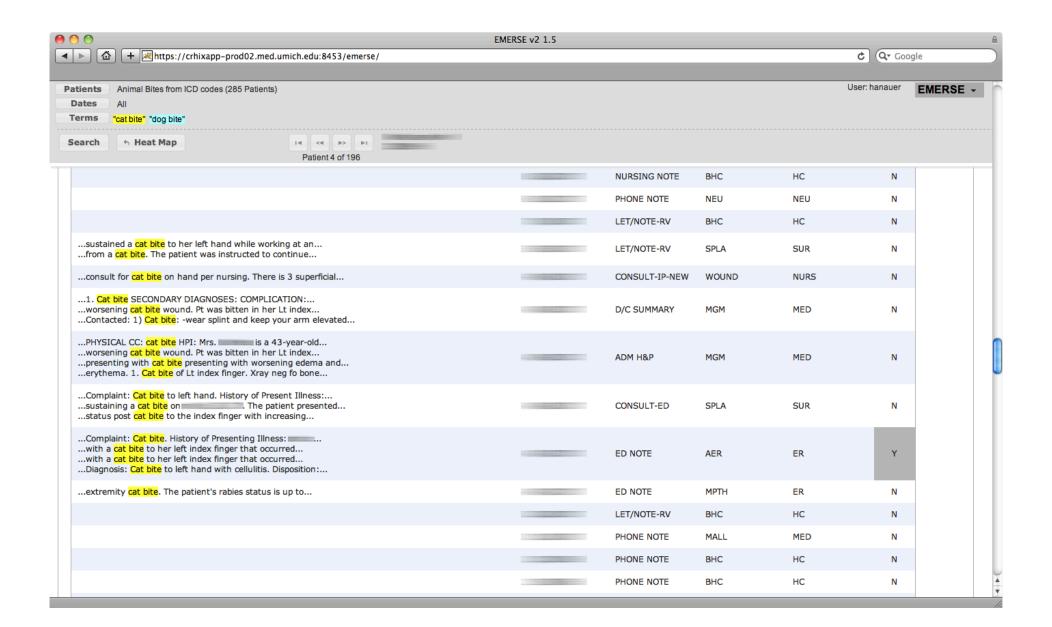












Publications using EMERSE

354 papers54 abstracts

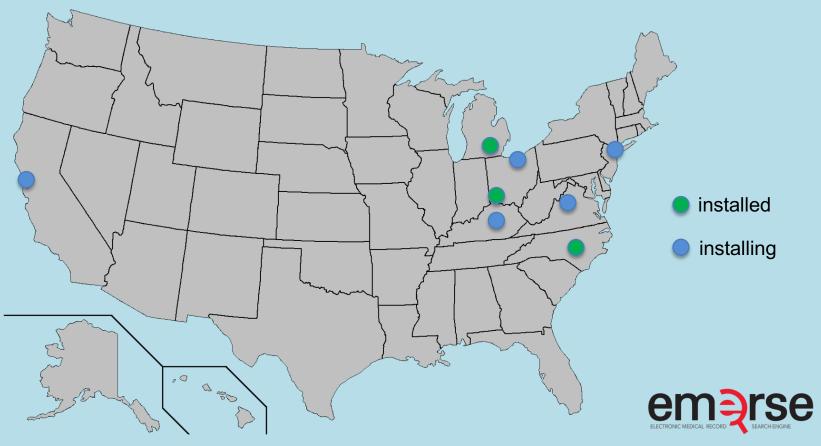


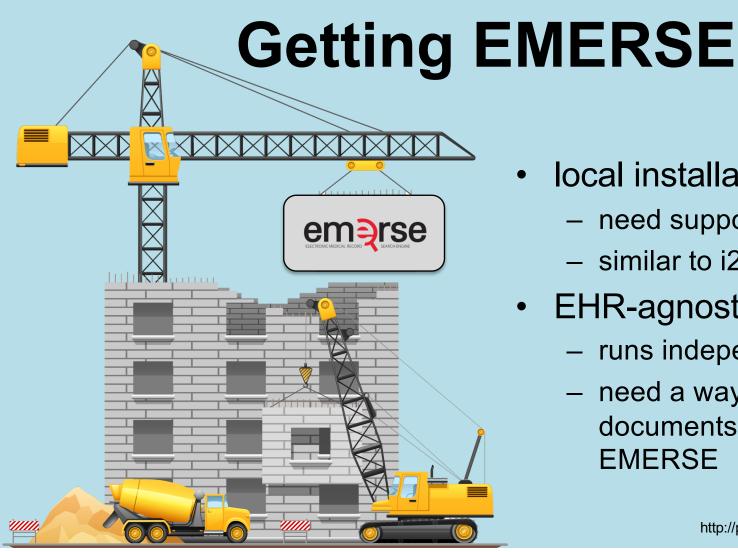
Full list can be found at:

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



Where is EMERSE?





- local installation required
 - need support of IT team
 - similar to i2b2
- **EHR-agnostic**
 - runs independent of EHR
 - need a way to extract documents to send to **EMERSE**

Test driving EMERSE

em₃rse

online demo videos on website virtual machine

> could be used for individual projects





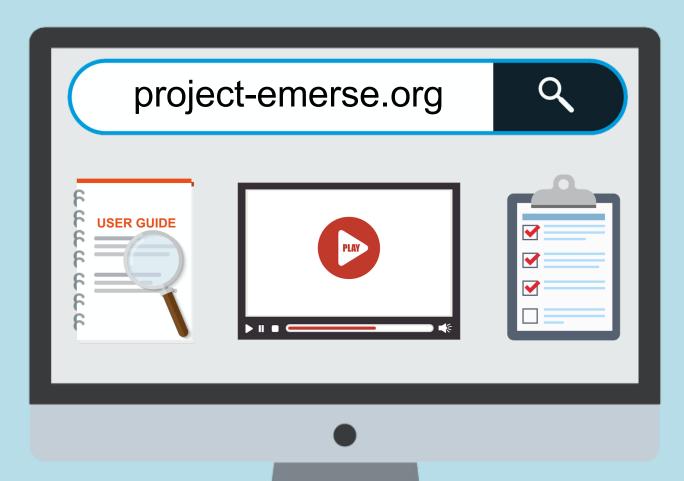
EMERSE "101"



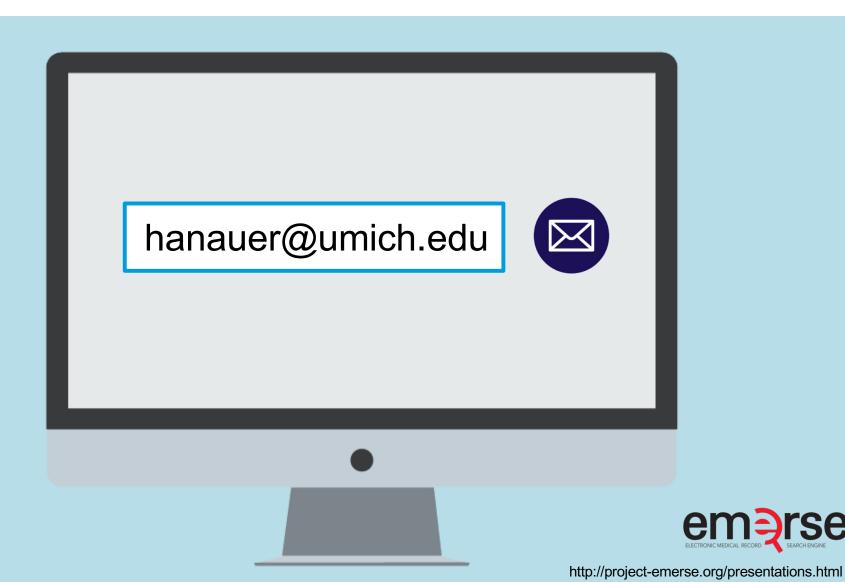
- live training
- open to anyone
- June 25, 2020, 10 AM EST
- register here:

https://bit.ly/2XPjQ6c









Acknowledgements



U of Michigan Rogel Cancer Center [P30CA046592]



Michigan Institute for Clinical and Health Research (MICHR) [UL1TR000433]



U of Michigan Health Information and Technology Services (HITS)



NCI Information Technology for Cancer Research (ITCR) [U24CA204863]

