

Hem/Onc Clinical Trials Course

David Hanauer

hanauer@umich.edu

July 21, 2020

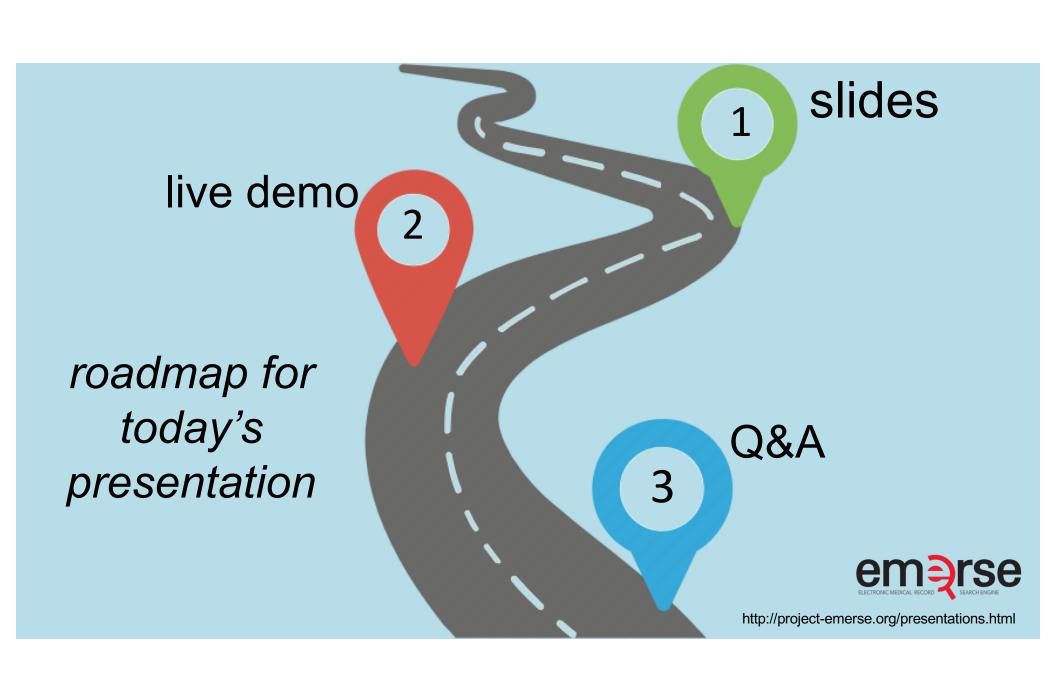
If you're thinking of taking notes



these slides can be found at:

link will be on most slides







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)







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



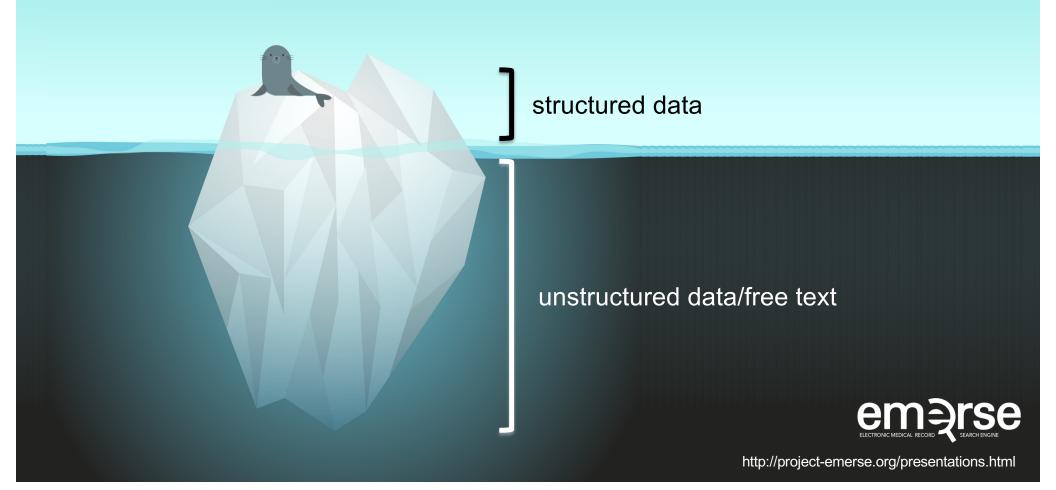


Unstructured vs Structured Data

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: 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

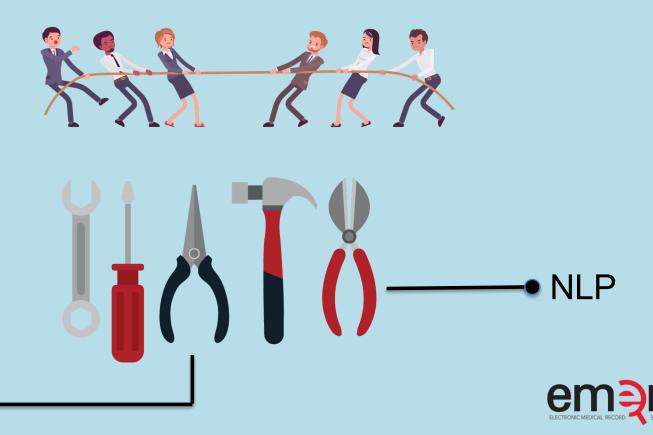


NLP: artificial intelligence



EMERSE vs NLP

EMERSE



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

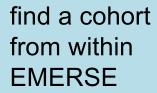
▶ ICD-10-CM Diagnosis Code C49.6 | [convert to ICD-9-CM]

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



Working with cohorts





EMERSE patient lists

(up to 100,000 per list)

import list from another system

upload a list of patients





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



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)





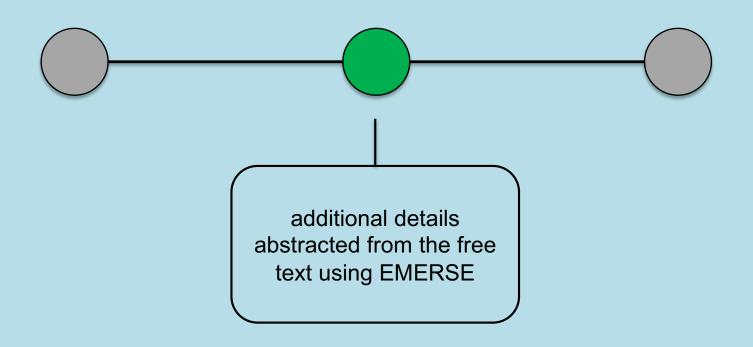
Typical workflow

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

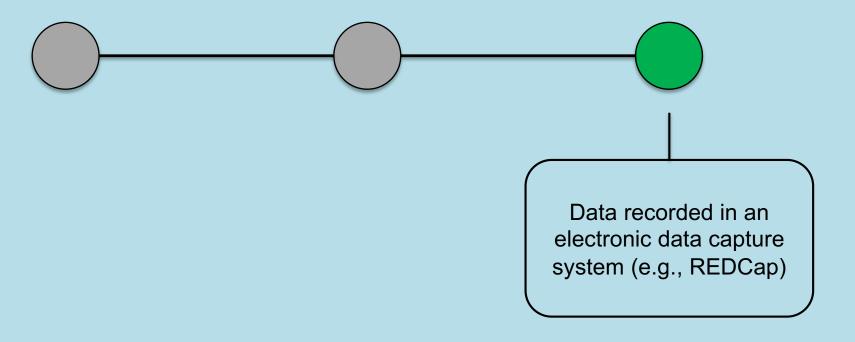


Typical workflow





Typical workflow





Publications using EMERSE

346 papers

52 abstracts

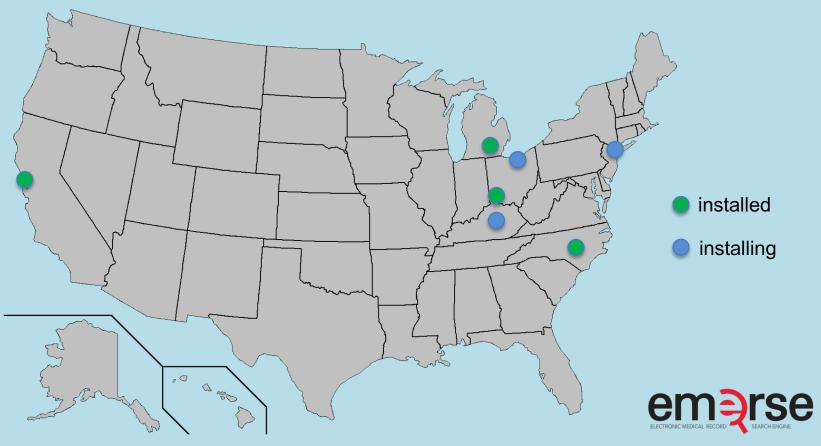


Full list at:

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



Where is EMERSE?



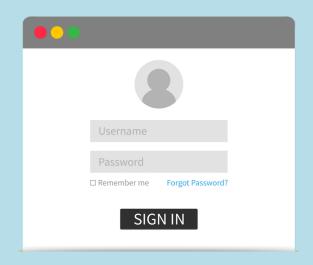
The future...



Getting access to EMERSE

@ Michigan

- Research: Data Office (need IRB approval)
- Non-research: Through Michart (contact David Hanauer for details)





Who can obtain access?

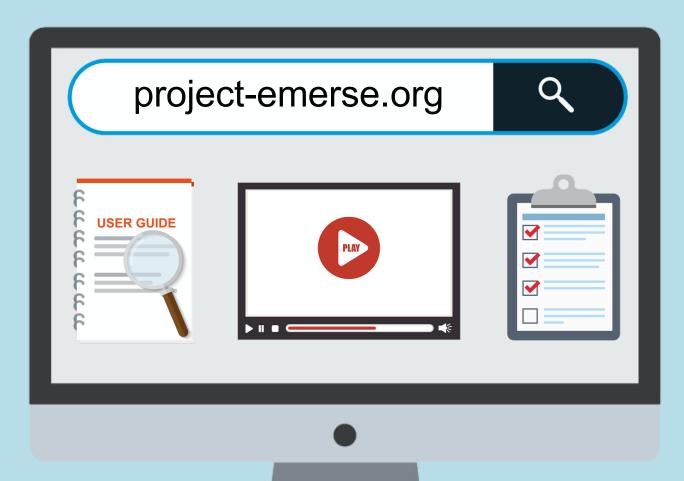
@ Michigan

Anyone with EHR access, BUT:

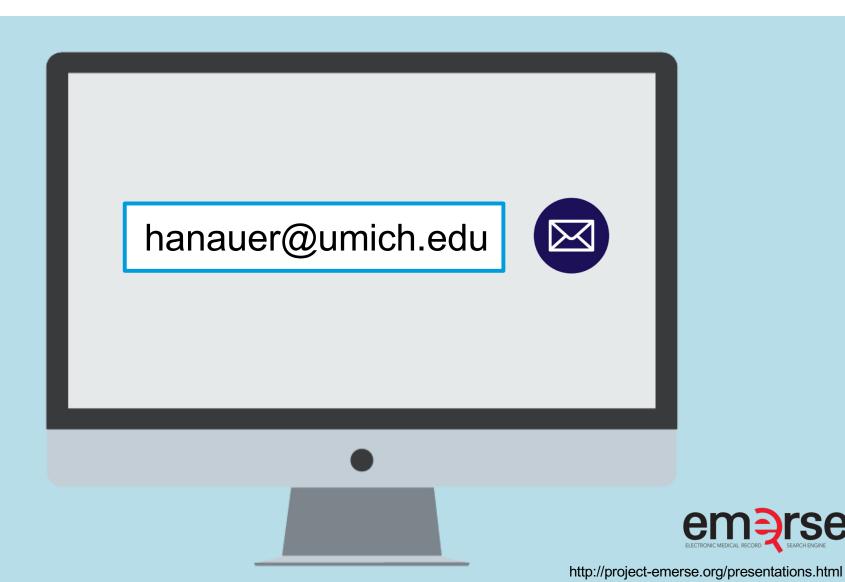
- you need a valid reason such as research, operations, quality
- "curiosity" is not a valid reason
- all usage is tracked











Acknowledgements





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



U of Michigan Health Information and Technology Services (HITS)



NCI Information Technology for Cancer Research (ITCR)

