

An introduction to the EMERSE text searching and natural language processing tool for electronic health record notes

Presentation for Columbia University Herbert Irving Comprehensive Cancer Center

October 29, 2025



Twitter/X: @projectEMERSE

Web: project-emerse.org

Email: hanauer@umich.edu
emerse-team@umich.edu

David Hanauer, MD, MS

Dept of Learning Health Sciences
University of Michigan

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

If you want to take notes or visit links



these slides can be found at:

this link will be
on most slides



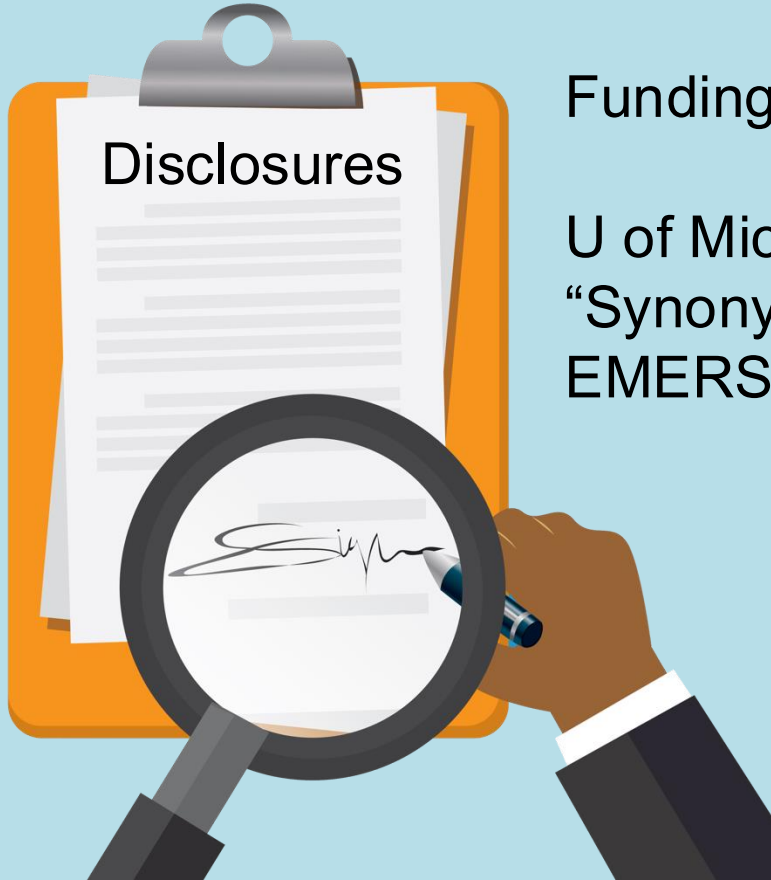
em**rse**
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

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

Disclosures

Funding: NIH (NCI-ITCR, NCATS-CTSA)

U of Michigan Royalties/Licensing:
“Synonyms” dataset—optional “plugin” for
EMERSE.



A hand holding a megaphone on the left side of the image. From the megaphone, several white and orange lines radiate outwards towards the top right, creating a sense of motion and emphasis. The background is split into a red area on the left and a black area on the right.

I am here to tell you
about EMERSE

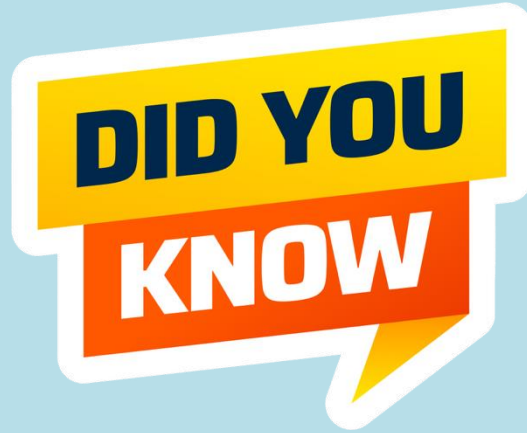
em_{er}se
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

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

2021 study out of UC Irvine: *Design, Implementation, and Usability of the Electronic Medical Record Search Engine (EMERSE) Tool*

<https://escholarship.org/uc/item/44p23878>

“Users unanimously responded that they would recommend the system to others, and...for a tool they found so useful, they believed that **far too few people both within and outside of their network knew about the tool’s existence.**”



EMERSE is available at HICCC

Unstructured vs Structured Data

EMERSE is for this...	...not this
<i>Unstructured Data (free-text)</i>	<i>Structured Data</i>
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

80% of EHR data are in unstructured free-text



structured data



unstructured data/free text

Most medical centers lack tools for free-text



And the ones that do exist aren't great



2024 study on EMR usability:

“The same three items (integration into workflow, **finding information**, and usability of alerts) received the highest number of 'poor' ratings among hospital and practice physicians.”

<https://www.nature.com/articles/s41746-025-01657-4>

Do we even need search anymore?



Yes!

Perspective

Search still matters: information retrieval in the era of generative AI

William Hersh , MD*

<https://pmc.ncbi.nlm.nih.gov/articles/PMC11339511/pdf/ocae014.pdf>
2024

The EMERSE solution



- A system “for the people”
- Users search the notes on their own
 - No need to wait in a queue for an analyst or a data scientist
- Easy-to-use for non-technical users
- Unlike with some EHRs, EMERSE can search across all notes and all patients at once
- Continuous refinements for almost 20 years



The EMERSE solution

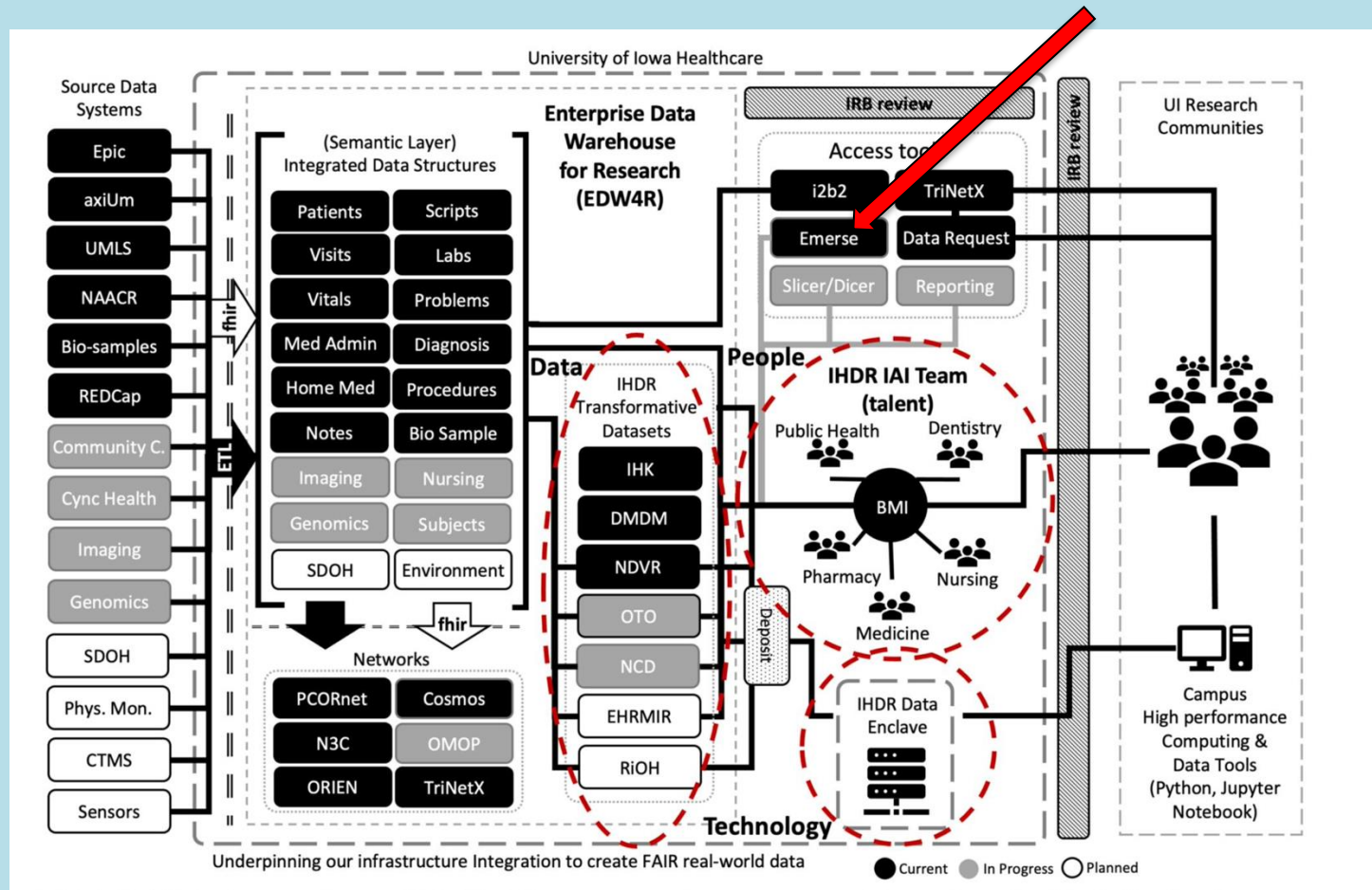


- A system “for the admins”
- Enterprise grade, easy to support
- Configurable with granular roles/privileges to control access (e.g, all patients vs. some patients)
- Data are kept secure within a centralized, audited system *at your own site*
 - No need to download/store the data elsewhere

EMERSE is a useful tool

- It will be one of many you need to build/support your project(s)





We think EMERSE is really good

- We're proud of what we created
- We think it's better than other software tools that cost \$\$\$

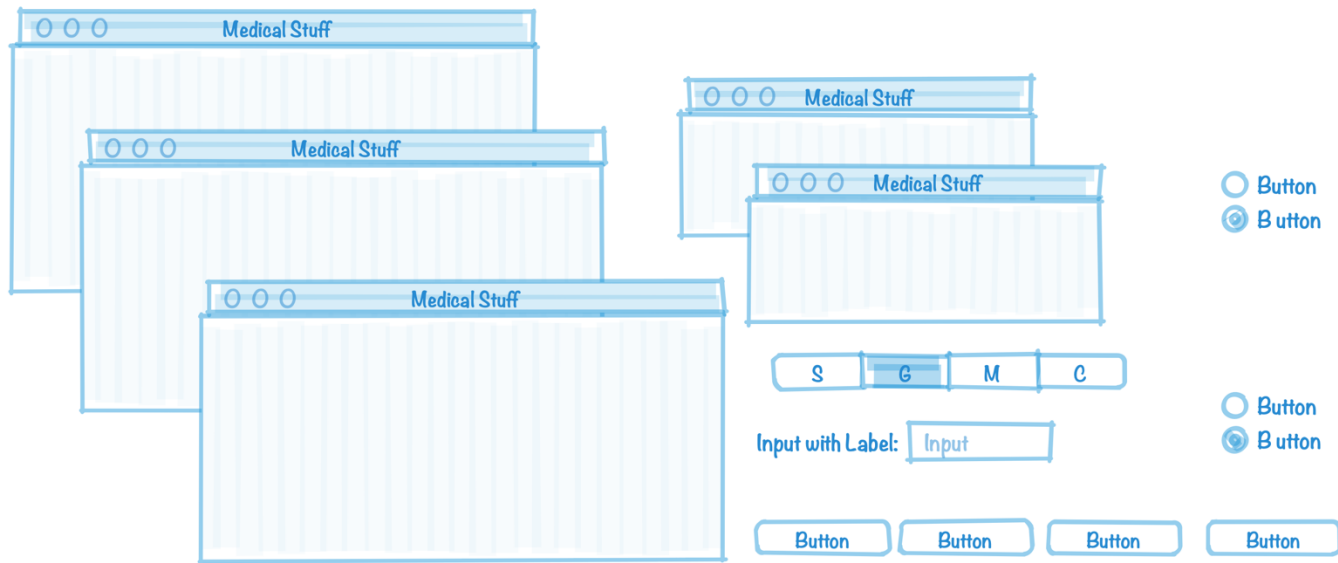


Why do we think it's better?

We've been laser-focused on it

EHRs have thousands of features

ACME EHR



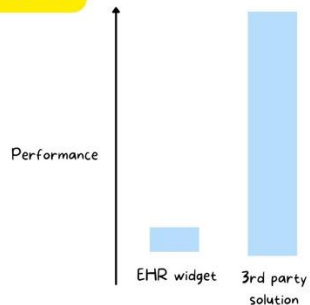
The EMERSE team has invested \$millions and 20 years on just this one feature (search) because it's so important

Why do we think it's better?

It has to be better

It's not a health system's fault for hoping an EHR can provide everything.

3rd party solutions must be 10x better on a differentiator that truly matters.



EMERSE is a “best of breed” product and must stay ahead of integrated EHR tools to remain relevant

Why do we think it's better?

We've studied it

- “new users of the EMERSE system are able to complete basic but critical workflow tasks in the system with a high rate of success...are **highly satisfied with the interface**, and have highly positive perceptions of its expected utility in their work.”
- “survey results indicate **very high ratings of usability and satisfaction** with the EMERSE system by new users after only one session of use.”

Why do we think it's better?

Our users tell us

Thank you for...this important tool which is **proving extremely valuable** in enhancing patient safety and quality of care delivered at Michigan Medicine.

..continued gratitude for the availability of this powerful research tool. As always, **thank you for this innovation!**

EMERSE is **working out great for casefinding**...have found quite a bit of cases that we would have missed otherwise. **Thank you and everyone on your team!**

To say **"it is the most useful tool I use in my job"** would be an understatement, so **thank you** for creating and sharing it with others!!

EMERSE is **an absolute gem** @umichmedicine. The functionality is very friendly and it saved hours of time during the data collection process. **Thank you to the team** that created this powerful research tool!



Why do we think it's better?

Researchers mention it in their publications

- “Reviewers used the EMERSE search tool to **ensure thorough review** of the available documentation...” [PMID 36119396]
- “the **tool avoids the pitfalls of diagnostic inaccuracy** seen with tools querying on ICD and billing codes...” [PMID 36114099]
- “[EMERSE] provides software features to **comprehensively scan all clinical documents**...for keywords and phrases to ensure that even rarely mentioned events are detected.” [PMID 36550198]
- “...information [was] captured via EMERSE...in order to **obtain the most accurate and complete information** per patient.” [PMID 36752520]

Why do we think it's better?

Top-tier medical centers use it

Site	Site
U of Michigan	Case Western Reserve U
Harvard U – Dana Farber Cancer Center	Utrecht University, Netherlands
Columbia U Cancer Center	U of Iowa
U of North Carolina – Chapel Hill	Weill Cornell Medical Center
U of California – San Francisco	U of Virginia
U of Kentucky	Moffitt Cancer Center, Tampa, FL*
U of Cincinnati	Children's Health Orange County, Calif.*

*Currently installing

What can EMERSE do?

Lots of things!



EMERSE for Cardiac Surgery Research

David Hanauer

Linda Farhat, Clinical Research Coordinator
Department of Cardiac Surgery, Michigan Medicine

01:28

Cardiac Surgery Research

"It allows us to basically target exactly what we need to know."

--Linda Farhat, Clinical Research Coordinator



EMERSE for Quality Analytics

David Hanauer

Andrew Heiler, Clinical Quality Coordinator

01:18

Quality Analytics

"I can't imagine having to do what I do without having EMERSE."

--Andrew Heiler, Quality Analytics Coordinator



EMERSE for Infection Control

David Hanauer

Jayna Berger Heiler, Infection Prevention Coach

01:33

Infection Prevention

"The value of EMERSE is immeasurable."

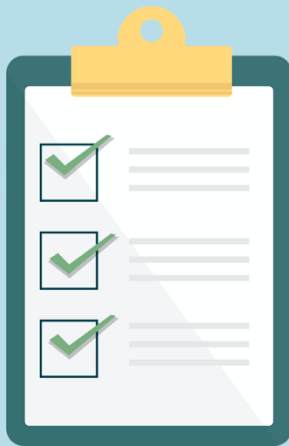
--Jayna Berger Heiler, Infection Prevention Coach

Watch our videos: https://project-emerse.org/use_cases.html

Find cohorts

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

- diseases
- drugs
- symptoms
- anything*



*if it is mentioned

Find cohorts

It's perfect for finding rare things...



...like rare cancers such as
cutaneous leiomyosarcoma

See this talk for more details:

<https://vimeo.com/677482835>

“Using EMERSE to Improve Research
Involving Rare Cancers”



Highlight documents for chart review

Thoracocentesis confirmed the recurrence of mantle cell lymphoma. Disease restaging work-up revealed multicompartiment 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.

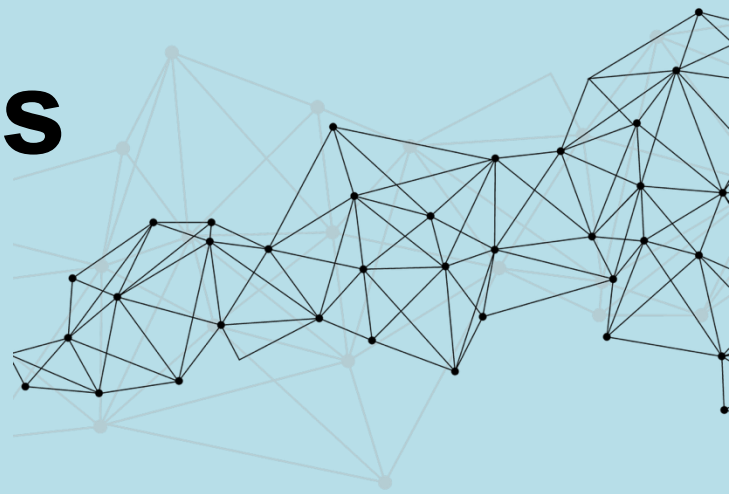


EMERSE is $\equiv$ *fast*

Query to identify all patients with the following	Reporting DB 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

Synonyms



- Used for query expansions
- User-controlled
- Multiple datasets can be included
- EMERSE Synonyms
 - acronyms, abbreviations, professional/consumer terms, misspellings, trade/generic drug names, species, chemo regimens, phrase variations, malapropisms, idioms, neologisms, organizations, companies, & more
 - 2.6 million unique entries

Synonyms for "smoker"



Click individual terms to highlight or de-highlight.

Synonyms (126)

[SORT A TO Z](#)

[HIGHLIGHT ALL](#)

[HIGHLIGHT NONE](#)

- abuse nicotine
- abuses nicotine
- addicted to nicotine
- addicted to tobacco
- addiction to nicotine
- addiction to tobacco
- ash tray
- ash trays
- ashtray
- ashtrays
- chain smoker
- chain smokers
- chain smoking
- chainsmoker
- chainsmokers
- chainsmoking
- cigar
- cigar day
- cigar per day
- cigarette day
- cigarette pack day
- cigarette packs each day
- cigarette packs every day
- cigarette packs per day
- cigarette per day
- cigarette smoke
- cigarette smoker
- cigarette smokers
- cigarette smoking
- cigarettes
- cigarettes day
- cigarettes per day
- cigars
- cigars days
- cigars per day
- continue to smoke
- continued to smoke
- continues to smoke
- continuing to smoke
- current every day smoker
- current everyday smoker
- current smoker
- currently smokes
- dependence on cigarettes
- dependence on nicotine
- dependence on tobacco
- dependent on cigarettes
- dependent on nicotine
- dependent on tobacco
- nicotine
- nicotine abuse
- nicotine abuser
- nicotine abusers
- nicotine addiction
- nicotine addictions
- nicotine dependence
- nicotine dependent
- pack a day
- pack day
- pack each day
- pack history
- pack per day
- pack year
- pack year smoker
- pack years
- pack yr
- pack yrs
- packs
- packs a day
- packs day
- packs each day
- packs of cigarette day
- packs of cigarette each day
- packs of cigarette per day
- packs of cigarettes day
- packs of cigarettes each day
- packs of cigarettes every day
- packs of cigarettes per day
- packs per day
- packs per year
- packyear
- packyears
- packyr
- packyrs
- pipe
- pipes
- pk year
- pk years
- pk yr
- pk yrs

Synonyms for "social determinant of health"



Click individual terms to highlight or de-highlight.

Synonyms (364)

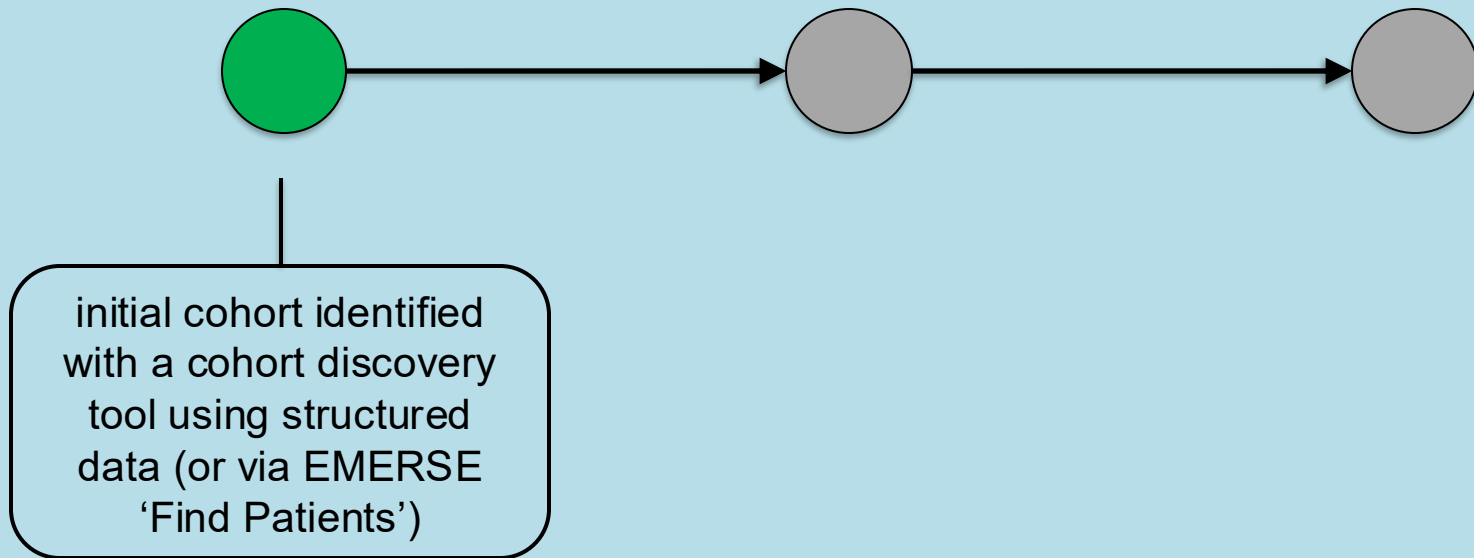
[SORT A TO Z](#)

[HIGHLIGHT ALL](#)

[HIGHLIGHT NONE](#)

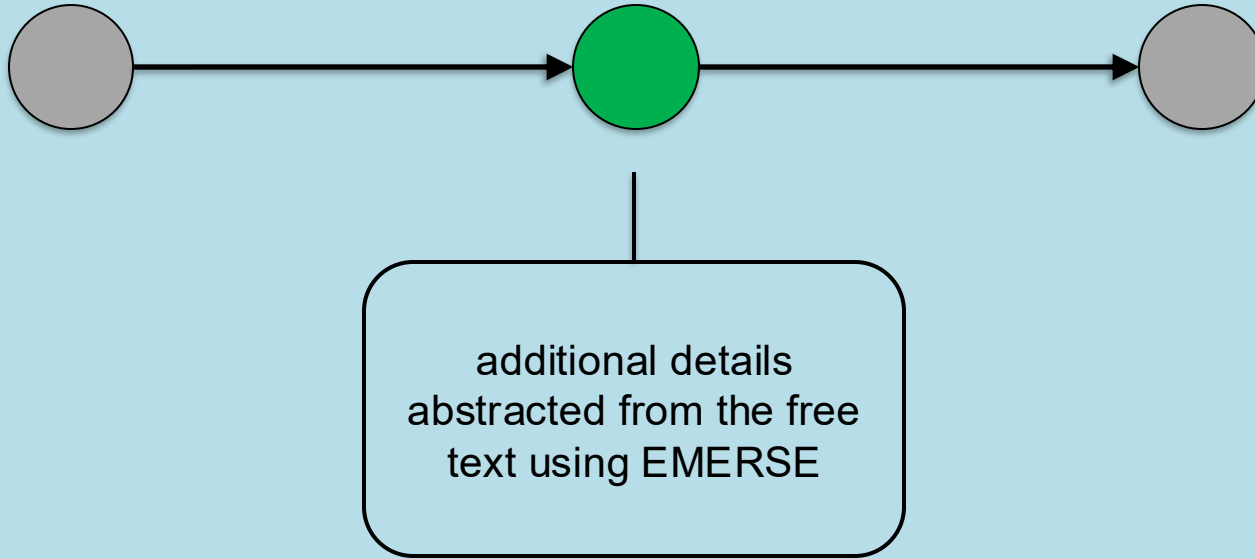
abused academic problem adoptive parent afford affordable medication affordable prescription
alcoholism amphetamine abuse antisocial appears depressed appears sad appears to be depressed
appears to be sad assisted living facilit assisted living facility assistive device bankruptcy banned
bath bench bath not on first floor bed and bath are not on the first floor boarding house burglary
care giver support cared for by a friend cared for by a relative charity care cheaper childhood abuse
chipped paint cigarettes cluttered co pay cocaine abuse common law partnership community support
computer illiterate condominium copay correctional facility court date court mandated court ordered
crime victim death of family member death of friend death of partner dementia detox
difficult working conditions difficulty maintaining employment disability income disabled partner
disabled partners disabled relative disorderly divorced drug seeking drunk dysthymia dysthymic
economic problems education level education problem education problems elderly parents
employment difficulties employment income insecurity employment insecurity estranged evict evicted
experiencing discrimination exposed to violence extended care facility extended tub bench feeling down
feels down felony filthy living finances financial concerns financial constraints financial difficulties
financial issues financial strain financial stress financial stressors financially fixed income flooded
food bank food insecure food insecurity food pantries food stamps foreclosed forgery forging
foster care foster family fruit and vegetable intake gender identity general financial insecurity
grab bars present group home gunpoint hand held shower health insurance healthcare problem

Typical workflow

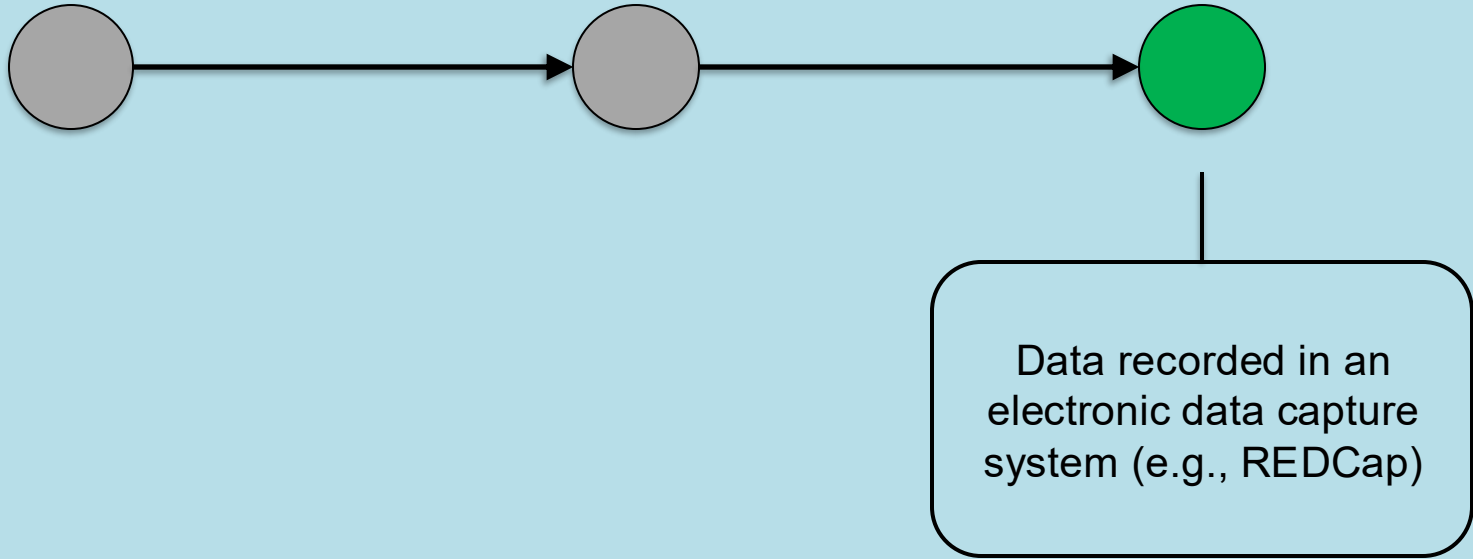


Cohort discovery tools:
i2b2/ENACT, Leaf, etc.

Typical workflow



Typical workflow



Publications using EMERSE

701

papers and abstracts



Full list at:

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

Recently released...

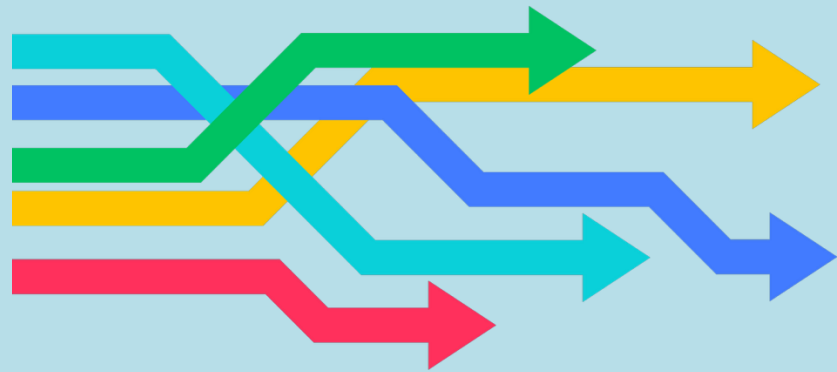


Incorporation of NLP features

- negation
- uncertainty
- subject (patient vs other)
- history of
- named entity recognition/mapping to ontologies

If you don't like our built-in NLP, use your own **em****rse**
ELECTRONIC MEDICAL RECORD SEARCH ENGINE

Future Work

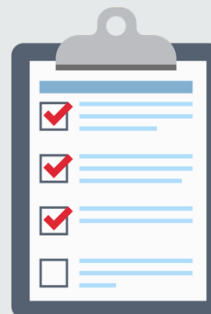
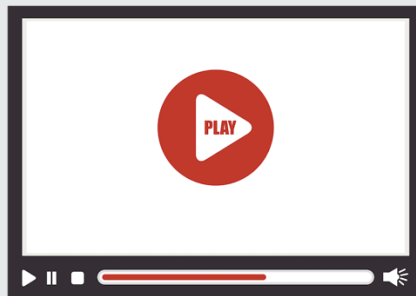


Data extraction from templated notes

Integration with ChatGPT or similar tools

& more...

project-emerse.org



Interested in EMERSE?



Contact us to schedule a time with your team for:

- Discussions about usage strategies
- Training
- Live demonstrations (abstractors, IT teams, etc)

Community Meeting



NEXT WEEK: Wed, November 5, 2024, 1-2 PM ET

Open to everyone, registration link at

<http://project-emerse.org/>



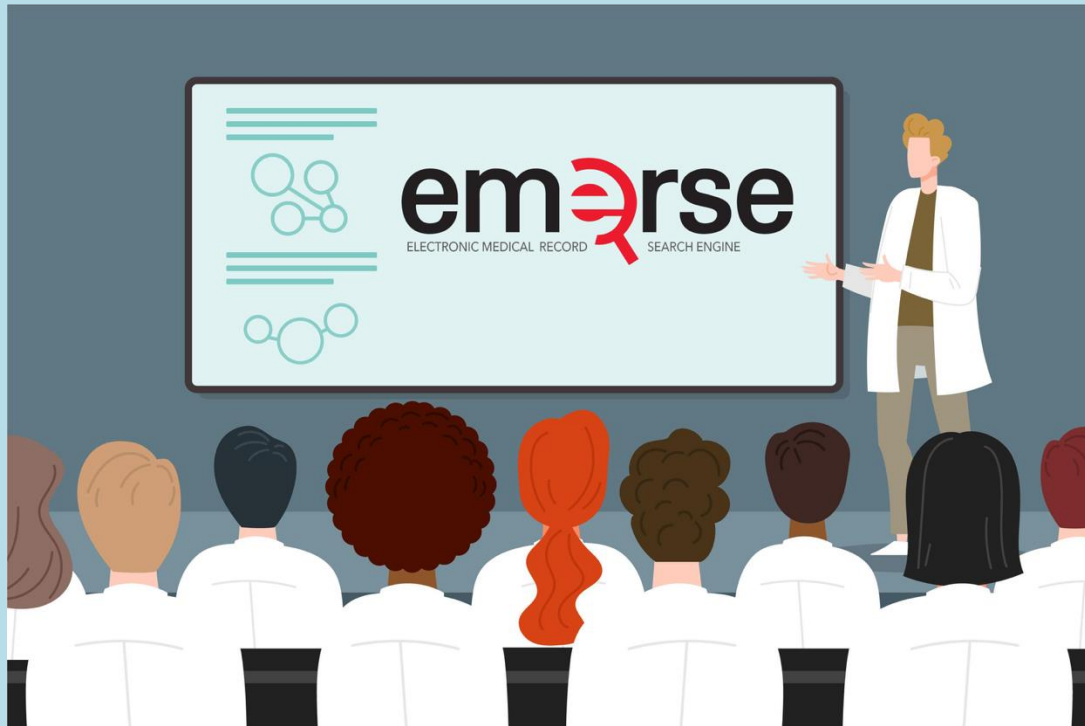
<http://project-emerse.org/presentations.html>

EMERSE-team@umich.edu



Lisa Ferguson
David Hanauer
Kellen McClain
Guan Wang

Live Demo



- No real names
- No PHI
- Publicly available
- Abstracts & case reports