

EMERSE Community Meeting

em^{ER}rse

2024-May-15

<https://project-emerse.org>

Plan for today

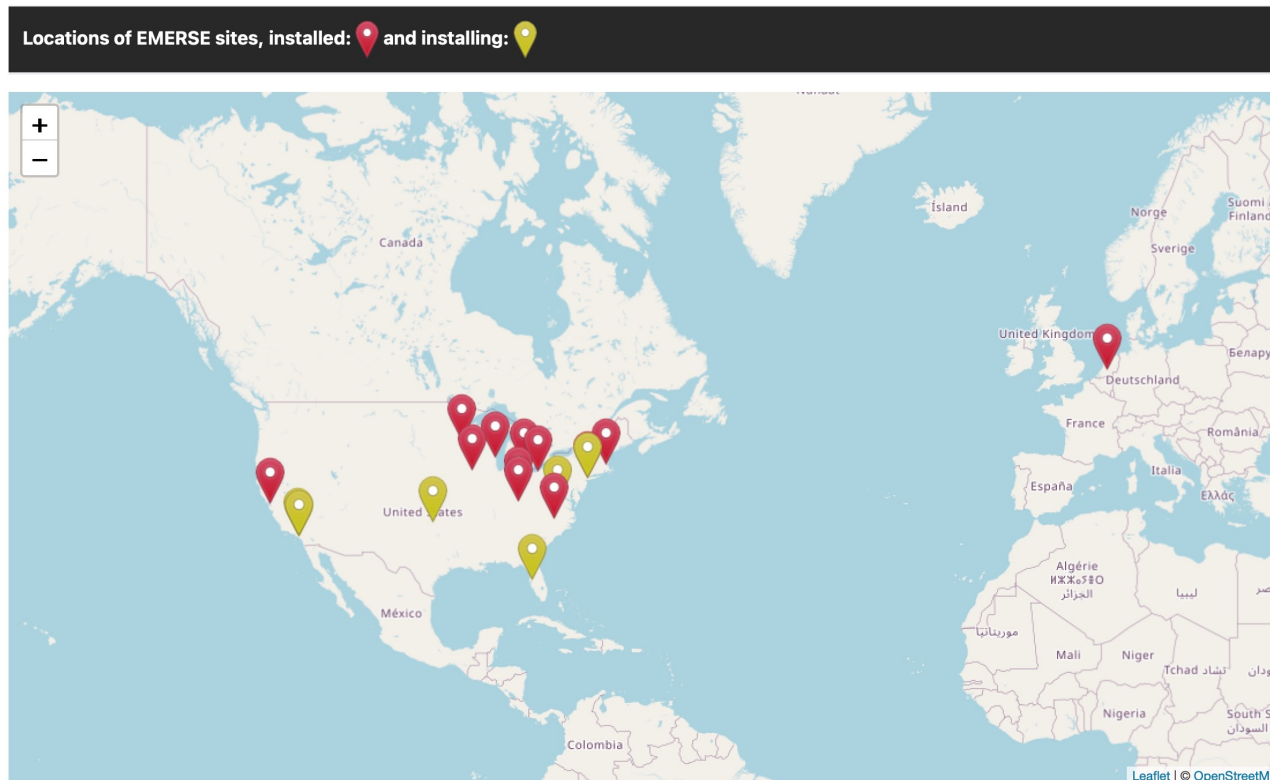
- Welcome and Housekeeping
- Announcements and Updates
- Guest speaker: Jackie Gravitt, “Implementing EMERSE at UVA: Lessons Learned”
- Discussion: Considerations for implementing and supporting EMERSE
- Open Forum
- Adjourn

Housekeeping

- *Zoom Meeting*
 - More opportunities for interaction
- Please stay muted, unless you would like to ask a question or make a comment
- Feel free to use the chat function to type questions or provide comments
- We will answer questions throughout
- We will record this meeting; it will be available where prior recordings are located:
 - <https://project-emerse.org/presentations.html>
 - <https://project-emerse.org/community.html>

Community

- Updated page for EMERSE Community
 - <https://project-emerse.org/community.html>



Acknowledgements

- EMERSE has been supported by:
 - NCI ITCR program
 - NCATS via MICHR CTSA
 - Michigan Medicine
 - Department of Learning Health Sciences
 - Office of Research
 - Health Information Technology & Services

Research Collaborative

- Last meeting was April 29, 2024; next meeting TBD
- Initial research focus: pronoun usage in the EHR over time
 - Contributions from U Michigan, UCSF, U Virginia, U Cincinnati, U Iowa

<i>Terms:</i>	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
preferred pronouns	0	0	0	2	198	389	581	992	1,438	2,017	2,204	2,046
pronouns are	0	0	2	6	218	435	616	918	1,184	1,361	1,523	1,798
she her hers	0	0	0	19	23	22	95	532	1,390	3,687	3,837	5,429
she her they	0	0	2	2	0	2	2	5	13	28	38	68
they them their	0	0	2	4	5	21	36	63	125	265	223	202

- If interested in participating, please contact EMERSE-team@umich.edu

Publications

563

Papers

92

Abstracts

18

Other

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

Publications (new since last meeting)

- DeCamillo D, Herrel LA, Haymart B, Latfolla A, Barnes GD. Unexplained hematuria in direct oral anticoagulant use: a single-center retrospective case series. *Res Pract Thromb Haemost.* 2024;8:e102404 www.rpthjournal.org. PMID: Pending.
- Kang Y, Chao S, Battles A, Firn J. Clinical Ethics Consultation for Patients Impacted by Incarceration: A Single Center Retrospective Review. *J Correct Health Care.* 2024 Apr 9. doi: 10.1089/jchc.23.10.0083. Epub ahead of print. PMID: 38597931
- Drake VE, Smith C, Watkins MO, Rudy SF, Joseph AW, Stucken CL, Brenner MJ, Kim JC, Moyer JS. Outcomes of Autologous Versus Irradiated Homologous Costal Cartilage Graft in Rhinoplasty. *Facial Plast Surg Aesthet Med.* 2024 Mar 19. doi: 10.1089/fpsam.2023.0334. Epub ahead of print. PMID: 38502836
- Pradhan S, Strohacker C, Schachtner S, Palm K, Trauth A, Gao Z, Marcuccio E. Management of Hematochezia in Infants with Congenital Heart Disease Admitted to the Acute Care Cardiology Unit: A Multicenter Retrospective Pilot Study. *J Pediatr.* 2024 Feb 27;269:113992. doi: 10.1016/j.jpeds.2024.113992. Epub ahead of print. PMID: 38417782
- Lopez-Medina AI, Campos-Staffico AM, A Chahal CA, Volkens I, Jacoby JP, Berenfeld O, Luzum JA. Genetic risk factors for drug-induced long QT syndrome: findings from a large real-world case-control study. *Pharmacogenomics.* 2024 Feb;25(3):117-131. doi: 10.2217/pgs-2023-0229. Epub 2024 Mar 20. PMCID: PMC10964839. PMID: 38506312
- Harake ES, Linzey JR, Jiang C, Joshi RS, Zaki MM, Jones JC, Khalsa SSS, Lee JH, Wilseck Z, Joseph JR, Hollon TC, Park P. Development and validation of an artificial intelligence model to accurately predict spinopelvic parameters. *J Neurosurg Spine.* 2024 Mar 29:1-9. doi: 10.3171/2024.1.SPINE231252. Epub ahead of print. PMID: 38552236
- Haring CT, Heft Neal ME, Jaffe CA, Shuman AG, Rosko AJ, Spector ME. Association of preoperative thyroid hormone replacement with perioperative complications after major abdominal surgery. *Am J Surg.* 2024 Jan 23:S0002-9610(24)00020-5. doi: 10.1016/j.amjsurg.2024.01.018. Epub ahead of print. PMID: 38311517
- Holste KG, Makar KG, Ulma RM, et al. Advantages of an Image-Guided Subcranial Technique for Le Fort III Osteotomies for Midface Advancement: A Case Series. *FACE.* 2024;0(0). doi:10.1177/27325016241227446. PMID: Pending
- Ahmad RA, Orelaru F, Arora A, Ling C, Kim KM, Fukuhara S, Patel H, Deeb GM, Yang B. Acute type A intramural hematoma: The less-deadly acute aortic syndrome? *J Thorac Cardiovasc Surg.* 2024 Jan 25:S0022-5223(24)00090-4. doi: 10.1016/j.jtcvs.2024.01.032. Epub ahead of print. PMID: 38280668
- Roseland ME, Ma T, Shampain KL, Stein EB, Wasnik AP, Curci NE, Sciallis AP, Uppal S, Johnson TD, Maturen KE. Neoadjuvant chemotherapy for high-grade serous ovarian cancer: radiologic-pathologic correlation of response assessment and predictors of progression. *Abdom Radiol (NY).* 2024 Mar 13. doi: 10.1007/s00261-024-04215-w. Epub ahead of print. PMID: 38478037

Publications (“The largest”)

- “Our report describes the successful technical advantages of using Stealth navigation for subcranial Le Fort III osteotomies and presents, to our knowledge, **the largest case series of patients with long-term follow-up.**” [doi:10.1177/27325016241227446]
- “...this is **the largest study that investigated the association of candidate genetic variants with the risk of diLQTS.**” [PMID 38506312]

Upcoming software release

- Version 7.x
- Release date: unknown, but we're getting close!
- Lots of complex changes under the hood
 - So many changes, it will be challenging to cover them all
 - Many are discussed in our GitHub issues (<https://github.com/project-emerse/emerse/issues>)
 - More powerful
 - Many UI improvements
 - Undergoing lots of testing, bug fixes
- Will have basic NLP capabilities baked in
 - Negation, named entity recognition, mapping to UMLS concepts
- UMLS license will be required (one per site; free)
- Will require upgrading, re-indexing, re-uploading synonyms

Upcoming software release

Temporary Filters Saved Filters

Temporary

- Name/Description
- Manage**
- Share
- Save
- Clear
- Filter Details

Patient Demographics

- Birth Date
- Ethnicity
- Race
- Sex
- Deceased

General

- Source
- Report Date
- Document Type
- Age in days, on document date
- Age in years, on document date**
- Encounter ID

Source: Main EHR

Match documents where Age in years, on document date represents how old the patient was at the time the document was written, and is in the following range:

From
_____ years _____ months

To
_____ years _____ months

No filter applied

CLEAR

Upcoming software release

The screenshot displays the EMERSE software interface. At the top left is the EMERSE logo. Below it, the 'Patients' section shows 'All Local Patients (10,000)'. The 'Filters' section has a text input 'Age in days, on document date'. The 'Terms' section contains a row of colored buttons: 'swimming' (pink), 'softball' (red), 'baseball' (red), 'nausea' (green), 'headache' (green), 'stomach pain' (cyan), 'abdominal pain' (cyan), 'chest pain' (yellow), 'pain in chest' (yellow), 'pain thoracic' (yellow), 'pains, chest' (yellow), and 'thorax pain' (yellow). Below the terms are two buttons: 'HIGHLIGHT DOCUMENTS' and 'FIND PATIENTS'. A navigation bar includes 'Temporary Terms', 'Saved Terms', 'Advanced Terms', and 'Synonym Preferences'. The main area is split into a left sidebar and a right pane. The sidebar lists actions: 'Name/Description', 'Edit', 'Share', 'Save', 'Clear/Delete', 'Export', 'Bulk Upload', and 'Query Detail' (highlighted). The right pane shows a 'Term Text Only' toggle (checked) and a visual representation of the search query: 'swimming' followed by 'AND (softball OR baseball)', 'AND (nausea OR headache)', 'AND (stomach pain OR abdominal pain)', and 'AND (chest pain OR pain in chest OR pain thoracic OR pains, chest OR thorax pain)'.

EMERSE

Patients All Local Patients (10,000)

Filters Age in days, on document date

Terms swimming softball baseball nausea headache stomach pain abdominal pain chest pain pain in chest pain thoracic pains, chest thorax pain

Results HIGHLIGHT DOCUMENTS FIND PATIENTS

Temporary Terms Saved Terms Advanced Terms Synonym Preferences

Temporary Terms

- Name/Description
- Edit
- Share
- Save
- Clear/Delete
- Export
- Bulk Upload
- Query Detail

Term Text Only

swimming

AND

(softball OR baseball)

AND

(nausea OR headache)

AND

(stomach pain OR abdominal pain)

AND

(chest pain OR pain in chest OR pain thoracic OR pains, chest OR thorax pain)

Upcoming software release



Patients Demo List (16)

Filters Birth Date Source Department in Main EHR Imaging Modality in Radiology

Terms

Results HIGHLIGHT DOCUMENTS FIND PATIENTS

Temporary Filters Saved Filters

Temporary

Name/Description

Manage

Share

Save

Clear

Filter Details

Certain filters may be added by the system in order to maintain the correctness of filtering.

Patient Filter(s):
AND Birth Date **BETWEEN** [01/09/2012 TO 01/30/2024]

Document Filter(s):

Source **INCLUDES** (Pathology, Radiology)

AND

(**NOT** Source: Main EHR **OR** Department **INCLUDES** (Internal Medicine))

AND

(**NOT** Source: Radiology **OR** Imaging Modality **INCLUDES** (CT, MRI))



Guest speaker

“Implementing EMERSE at UVA: Lessons Learned”



Jackie Gravitt

Health Data Engineer

Analytics and Reporting Team (ART) Data Operations

University of Virginia (UVA) Health

Implementing EMERSE at the University of Virginia

Lessons Learned

About Me

Jackie Gravitt

- Engineering Staff Specialist, Health Data
- Started with UVA Health in May 2023
- UVA Alumna – Systems Engineering and Psychology



- Longtime Michigan fan



UVA Health

About Us

- Academic Medical Center in Charlottesville, Virginia
- Also encompasses 3 community hospitals
 - 1000+ Beds Across 4 Hospitals
- Level 1 Trauma Center, Transplant Center, Comprehensive Cancer Center
- #1 Children's Hospital in Virginia



EMERSE Team at UVA

Team	Leaders	Work Team
Analytics & Performance Evaluation	Christine Kelly-Fisher	
Data Operations	Kapil Parkhi	Jackie Gravitt
RCTA – Research & Clinical Trial Analytics	Deb Green	Dave Guan, Fauzia Khan, Misha Zimmel, Rob Pates
HIT – Health IT	Ron Clarke	Shayla Gustafson

UVA Strategy

■ Background

- 2020 POC – desktop machine - completed
 - Prototype with data science institute partners Don Brown and Saurav Sengupta
 - Completed install on Windows machine with EMERSE dummy data and then with an export of UVA data

■ Project Plan

- Install on desktop with dummy data
- Install on test server with dummy data
- Populate test server with limited set of real data
 - ED Notes – Notes, Provider Notes, Triage Notes for last x many days
 - Weekly updates
- Install on production with all notes
 - Weekly updates
- Backfill on production with historical notes

EMERSE Components

University of Michigan





- Database
 - Oracle (prod)
 - Maria DB (test)
- Server
 - Linux
- Indexing Code
 - Java application that pushes to Solr via SolrJ API

University of Virginia

- Database
 - SQL Server
- Server
 - Windows
- Indexing Code
 - Python script using pysolr run via JAMS job

Database - SQL

- Database install – straight forward, Michigan provides scripts
 - Need to download SQL drivers

 6.6.0.mariadb.sql	1.97 MB	Feb 12
 6.6.0.mssql.sql	1.98 MB	Feb 12
 6.6.0.oracle.sql	5.84 MB	Feb 12
 6.6.0.postgresql.sql	2.05 MB	Feb 12

- Patients – weekly load
 - Loaded from Clarity via SSIS Package/SQL stored procedure
- Research Studies and Members – daily load
 - Python script pulls information from API and lands in staging table
 - SQL stored procedure moves from staging into EMERSE research tables

Server/Desktop - Windows

- Set system variables
 - JAVA_HOME, JRE_HOME, CATALINA_HOME, SOLR_HOME
- Tomcat
 - Set to run on port 8090
 - Change default credentials
 - Change allowed IPs
 - Configure Tomcat to run as a Windows service and to logon using local system account
 - Change initial and memory pools for Java From 128MB and 256MB to 1536 and 2048 MB, respectively
 - Change default settings for non-existent resources to not show report or server information
 - Configure Tomcat to enable HTTP Script Transport Security (HSTS)
 - Include SQL drivers in lib folder
 - NOTE: After SSL is configured, you will have to access the host via port 443

Server/Desktop – Windows (cont)

- Solr
 - Configure to run as a Windows service
 - Enable basic security, create user accounts, and enable roles
 - Configure audit logging
- Set Up SSL
 - Generate keystore file and certificates
 - Configure Solr to use it once set up

Indexing – Python

- SQL query loads data from Clarity
 - Goes against week old copy of Clarity

```
data_source = set_data_source('xxx')
print("Wait, loading SQL.....")
#sql_text = "select top 10 * FROM CLARITY.dbo.HNO_NOTE_TEXT"
sql_text = f'''
WITH Subquery as
(SELECT
    inf.NOTE_ID
    ,COALESCE(inf.UPDATE_DATE, '1/1/1900') AS LAST_UPDATED
    ,COALESCE(iptyp.name, 'Unknown') AS Note_Type_Desc
```

```
-----
where inf.DELETED_CAT_C is NULL
    and inf.PAT_ENC_CSN_ID is not NULL
        AND inf.DATE_OF_SERVIC_DTTM > DATEADD(DAY, -30, GETDATE())
        AND IP_NOTE_TYPE_C IN ('6', '8', '19')
    )
```

```
notes_df = pd.read_sql_query(sql_text, data_source)
```


Indexing – Python (cont)

- Pushes to Solr using pysolr in chunks of 1000

```
import pysolr
import json

solr = pysolr.Solr('https://xxx:8983/solr/documents/', always_commit=True)
display(solr)

display("Total number of documents to be added:")
num_of_rec=len(notes_df.index)
display(num_of_rec)
i=0

while (len(notes_df)>0):
    display(notes_df.head(1))
    notes_df_1000, notes_df = (notes_df[:1000], notes_df[1000:]) if len(notes_df) > 1000 else (notes_df, pd.DataFrame())
    notes_js1=notes_df_1000.to_json(orient='records',date_format='iso') #js1 is a json style string
    notes_js2=json.loads(notes_js1) #js2 is a list
    if len(notes_df)>0:
        print("adding", (1000*i+1), "to", (1000*(i+1)), "please wait ...")
    else:
        print("adding", (1000*i+1), "to", num_of_rec, "please wait ...")
    result = solr.add(notes_js2)
    display(result)
    i=i+1
```

Indexing – Python (cont)

- Code worked!
 - ...but load times were growing with each run – time to load 100k rows:
 - 10 minutes → 17 minutes → 25 minutes
 - Sufficient for weekly loads, but we wanted to load years worth of older data
- Asked ChatGPT for some help
 - Rewrote with threads and queues
 - Was able to load 7 million documents in the time it had taken to do 744k the with the old script

Indexing – Python (cont)

```
# Worker function to process chunks of data
def process_chunk(q):
    while True:
        chunk = q.get()
        print("Length of the queue:", q.qsize())
        logger.info(f'''Length of the queue: {q.qsize()}''')
        if chunk is None:
            break
        push_data_to_solr(chunk, solr_url, username, password)
        q.task_done()
```

```
while len(notes_df) > 0:
    chunk = notes_df.head(chunk_size)
    documents = chunk.to_dict(orient='records')
    logger.info(f'''Processing {len(chunk)} records, {len(notes_df)} remaining''')
    queue.put(documents)
    print("Length of the queue:", queue.qsize())
    notes_df = notes_df.iloc[chunk_size:]
```

```
# Push data to Solr using threading
chunk_size = 10000
logger.info(f'''Total Notes to process: {len(notes_df)} ''')
queue = Queue()
threads = []
for _ in range(4): # Adjust the number of threads as needed
    t = threading.Thread(target=process_chunk, args=(queue,))
    t.start()
    logger.info(f'''Name of the thread: {threading.current_thread()}''')
    logger.info(f'''Active count of thread: {threading.active_count()}''')
    threads.append(t)
print("Name of thread:", threading.current_thread())
print("Active count of thread:", threading.active_count())
```

```
# Wait for all threads to finish
logger.info(f'''Before join''')
queue.join()
logger.info(f'''After join''')

# Stop workers
for _ in range(4):
    queue.put(None)
for t in threads:
    logger.info(f'''Before join2''')
    t.join()
    logger.info(f'''After join2''')
```

Beyond the Basics

- Access
 - Access will be granted to those with IRB approval for a particular study
 - Using REDCap to request/approve
- Auditing Views
 - Created views to see who is accessing, attesting, searching

USER_ID	NAME	START_TIME	END_TIME	Session_Hours	ACTIV...	session_attestation_id	AttestationType	AttestationCode	descr...	SEARCH_TIME	search
rl	Sha...	2024-04-16 11...	2024-04-16 ...	0	1	2753	Common	ADMIN	Admi...	2024-04-16 11:15:38.673	{"useConjunction":true,"bundle":["diabetic"]}
n	Mik...	2024-04-16 11...	2024-04-16 ...	0	1	2752	Research Study	HSR	Can...	2024-04-16 11:15:15.433	{"useConjunction":true,"bundle":["lanreotide"]}
n	Mik...	2024-04-16 11...	2024-04-16 ...	0	1	2752	Research Study	HSR	Can...	2024-04-16 11:15:14.903	{"useConjunction":true,"bundle":["lanreotide"]}
d	Deb...	2024-04-12 11...	2024-04-12 ...	0	1	2706	Free Text	NULL	TFS...	2024-04-12 11:31:58.613	{"useConjunction":true,"advancedSearch":"\\induced psc
d	Deb...	2024-04-12 11...	2024-04-12 ...	0	1	2706	Free Text	NULL	TFS...	2024-04-12 11:31:58.427	{"useConjunction":true,"advancedSearch":"\\induced psc

	LOGIN_ID	USER_...	NAME	AttestationType	AttestationCode	DESCRIPTION	AttestationCount
1	2	j	Ja	Common	ADMIN	Adm	9
2	5	y	D	Free Text	NULL	DE	5
3	1	t	T	Research Study	HSR	EM	4
4	1	c	C	Free Text	NULL	Be	3
5	1	c	D	Free Text	NULL	TFS	2
6	1	c	D	Research Study	HSR	EM	1

Pilot Projects

- Pronouns in EHR with EMERSE partners
 - Christine and Deb searched according to study protocol
 - Manually updated aggregate data on study spreadsheet, validated results, submitted to the researcher
- Psoriasis study where needed information was in patient notes
 - Researcher submitted request to Data Request team; decided EMERSE might be a good fit
 - Researcher included diagnosis codes as a starting point, then required study team to manually review patient notes for specific phrases to meet inclusion
 - Deb and Harrison (on her team) did searches, exported patient lists, validated results, submitted to the researcher



Open Discussion: Implementation & Support

Wrap up

- THANK YOU
- Next Meeting: TBD (Sep/Oct 2024)
- Please complete meeting survey:

