# Advanced Development and Dissemination of EMERSE for Cancer Phenotyping from Medical Records

David Hanauer University of Michigan ITCR Annual Meeting May 23, 2018

#### EHRs ARE NOW PERVASIVE



Source: https://dashboard.healthit.gov/quickstats/pages/FIG-Hospital-EHR-Adoption.php

### 80% OF EHR DATA UNSTRUCTURED



UnstruturedStructured

http://www-07.ibm.com/hk/watson/health/

## STRUCTURED VS UNSTRUCTURED

#### **Structured Data**



Lab results



Prescribed medications

**Billing codes** 



\$

**Problem lists** 

#### **Unstructured Data**



Radiology images and reports



Pathology images and reports



**Clinical notes** 

## STRUCTURED VS UNSTRUCTURED

The unstructured, or free text, data are really important!

#### **Unstructured Data**



Radiology images and reports



Pathology images and reports



**Clinical notes** 

## FREE TEXT USES



P

#### eligibility determination



adverse event detection



data abstraction



## **NEED FOR FREE TEXT**



eligibility determination

Free text needed for resolving...

...**59%** of eligibility criteria for a chronic lymphocytic leukemia (CLL) clinical trial

...77% of eligibility criteria for a prostate cancer clinical trial

PMCID: PMC4333685

## DON'T IGNORE THE FREE TEXT



Accuracy of the medication list in the electronic health record—implications for care, research, and improvement

"There was variation by center, from 90% patients with complete agreement between the medication list and clinical narrative to 50% agreement."

"Our findings have important implications for deciding whether to solely rely on medication order data as a proxy for medication usage in research or clinical decision support."

https://academic.oup.com/jamia/advance-article/doi/10.1093/jamia/ocy027/4996457

### **RESEARCHERS NEED TOOLS**





## **RESEARCHERS NEED TOOLS**

#### **Structured Data**



Lab results



Prescribed medications



**Billing codes** 



**Problem lists** 







## **RESEARCHERS NEED TOOLS**









#### **Unstructured Data**



Radiology images and reports



Pathology images and reports



**Clinical notes** 

## **EMERSE**



#### <u>Electronic Me</u>dical <u>R</u>ecord <u>S</u>earch <u>E</u>ngine

- Developed at U of Michigan
- In use since 2005
- 198,000 user sessions
- 1,480 studies supported
- 164 (known) peer-reviewed publications

## **EMERSE – INFORMATION RETRIEVAL**

- Information Retrieval (IR) tool
  - Users familiar with Google, Bing, etc.
- Similar to, but has differences with, Natural Language Processing (NLP)
  - NLP can be powerful, but requires expertise and can be difficult to implement
     "Teaching a machine to read

"Teaching a machine to read a record is a lot harder than anyone thought"

• Flexible for most information needs

https://www.statnews.com/2017/09/05/watson-ibm-cancer/

Image: Bit	Patients All (2,351,932)			User: hanauer	EMERSE +
Terms * *pthelicid hemangioendathelisme*	Dates All Dates: 01/01/1900	) through 05/18/2018			
<ul> <li>Support the specific document summarizes are shown.</li> <li>Support the specif</li></ul>	Terms "Epithelioid hemangie	oendothelioma"			
All pottients matched the search arity         Dots, ranked document summories are shown.         To we patients to semporary Patient List and then run the search agent.         The patients to temporary Patient List and then run the search agent.         The patients to temporary Patient List         The trans         Charter to the patients to temporary Patient List and then run the search agent.         Charter to the patients to temporary Patient List and then run the search agent.         Charter to the patients to temporary Patient List and then run the search agent.         Charter to the patients to temporary Patient List and then run the search agent.         Charter to the patients to temporary Patient List and then run the search agent.         Charter to the patients to temporary Patient.         Charter to the patients to temporary Patients.         Charter to the patients to temporary Patients.         Charter to the patients to temporary Patients.         Charter to the patients the patients to the patients to the patients the patients to temporary Patients.         Charter to the patients the patients to the patient to temporary patients to the patients the					
<ul> <li>bit por univer decountent summaries or show.</li> <li>In traview these patients in more detail, move them to a Temporary Patient List and then run the search again.</li> <li>Invester search and the search again.</li> <li>Invester search adain.</li> <li>Invester search adain.<!--</td--><td></td><td></td><td></td><td></td><td></td></li></ul>					
116 patients matched the search criteria         100 top-onded document summaises are show.         To review the solution is more dockin, more than to a temporary Patient List and then run the search again.         Interview the solution is more dockin, more than to a temporary Patient List and then run the search again.         Interview the solution is more dockin, more than to a temporary Patient List and then run the search again.         Interview the solution is more docking the temporary Patient List.         Interview the solution is to temporary Patient List.         Revise Terms         Interview the solution is to temporary Patient List.         Interview the solution is to temporary Patient List.         Interview the solution is the solution is too patient is too temporary Patient List.         Interview the solution is too patient is too temporary Patient List.         Interview the solution is too patient is too temporary Patient List.         Interview the solution is too patient is too temporary Patient List.         Interview the solution is too patient is too					
100 top-orpled document summaries are shown.	116 patients mata	ched the search criteria		Summary	
To review these patients in more detail, move them to a Temporary Patient List     Move patients to Temporary Patient List   Revise Terms   Bott charts by Demographics _ Sort charts by Frequency   Cender   Conder   Female   Mole   And   Mole   And   Mole   And   Mole   (Difficient Terms) Terms Demographics _ Sort charts by Frequency Cender (Conder) (Cond	100 top-ranked document sur	nmaries are shown.			
Move patients to Temporary Patient List	To review these patients in mo	ore detail, move them to a Temporary Patient List	t and then run the search again.	suggestive of epithelioid hemangioendothelioma. See COMMENT includes epithelioid hemangioendothelioma, poorly differentiated	
Instrume   History of egitheliaid hemonogiendabeliam, About 30 History of egitheliaid hemonogiendabeliam, COMPAISON Conder Conder Fender Fender Mole 40 40 History Fighteliaid hemonogiendabeliam, CDI highlights History of egitheliaid hemonogiendabeliam, CDI highlights History egitheliaid hemonogiendabeliam, CDI highlights History of egitheliaid hemonogiendabeliam, CDI have History of egitheliaid hemonogiendabeliam, C	Move patients to Tempora	ry Patient List		considerations, <mark>epithelioid hemangioendothelioma</mark> is our primary	
Revise Terms      with harvon repitheliad hemongioendatheliam. This scon				History of <mark>epithelioid hemangioendothelioma</mark> . About 30	
Listophene Setupping Setup	Revise Terms			with known epithelioid hemangioendothelioma. This scan history of <mark>epithelioid hemangioendothelioma</mark> . Please note	
Sort chorts by Demographics _ Sort charts by Frequency      Gender      Female      Male      Unknown      Unknown      Unknown      Unknown      Consistent with epithelioid hemongioendotheliona. Immunohistochemical      Sert charts by Trequency      Consistent with epithelioid hemongioendotheliona. The itsets      diagnosis of epithelioid hemongioendotheliona. Charts by Frequency      Kennel				HISTORY Epitheliaid hamanajaandatheliama COMPARISON	
Sort charts by Demographics Sort charts by Frequency      Gender      Fencle      Fencle      Unknown      Unknown      Unknown      Conceptibility      Concepti				diagnosis of <mark>epithelioid hemangioendotheliona</mark> . There is	
GenderCOMPLAINT: Epitheliaid hemanajoendotheliama. The liver   angiosarcoma versus epitheliaid hemanajoendotheliama. The liver   angiosarcoma versus epitheliaid hemanajoendotheliama. The liver   angiosarcoma versus epitheliaid hemanajoendotheliama. CD3 heves   ande   Mole   Mole   Unknown   Unknown   Unknown		<ul> <li>Sort charts by Demographics Sort c</li> </ul>	harts by Frequency	diagnosis of <mark>epithelioid hemangioendothelioma</mark>	
Scritter     Femole     Mole	Gender			COMPLAINT: Epithelioid hemangioendothelioma of the liver	
Femaleconsistent with epithelioid hemangioendothelioma. The lotest   Maleman with epithelioid hemangioendothelioma cD31 highlights   Maleman with epithelioid hemangioendothelioma on thistory  man with epithelioid hemangioendothelioma cD31 highlights  man with epithelioid hemangioendothelioma on thistory  man with epithelioid hemangioendothelioma on thistory  man with epithelioid hemangioendothelioma class from  meetastatic epithelioid hemangioendothelioma. CD31 highlights  man with epithelioid hemangioendothelioma class from  meetastatic epithelioid hemangioendothelioma. CD31 highlights  man with epithelioid hemangioendothelioma. CD31 highlights  meetastatic epithelioid hemangioendothelioma. Lobs from  meetastatic epithelioid hemangioendothelioma. GROSS:  metastatic epithelioid hemangioendothelioma. GROSS:  metastatic epithelioid hemangioendothelioma. COMMENT  metastatic epithelioid hemangioendothelioma equ  metastatic epithelioid hemangioendothelioma  metastatic epithelioid heman	Gender			evaluated	
Fendelliver bibleioid hemangioendathelioma. CO31 highlights   Maleliver bibleioid hemangioendathelioma in the liver   Maleliver bibleioid hemangioendathelioma. and history   Interstatic epithelioid hemangioendathelioma. Labs from  edie bibleioid hemangioendathelioma. CTs have   Interstatic epithelioid hemangioendathelioma. CTs have				consistent with epithelioid hemangioendothelioma. The latest diagnosed epithelioid hemangioendothelioma of the liver	
Female /5     Male				liver biopsy) Epithelioid hemangioendothelioma. CD31 highlights	
Male        Male <td>Female</td> <td></td> <td>/5</td> <td>liver. Epithelioid hemangioendothelioma in the liver</td> <td></td>	Female		/5	liver. Epithelioid hemangioendothelioma in the liver	
Male Male Male Male Male Male Male Male				Since <mark>epithelioid hemangioendothelioma</mark> is indolent	
Male       40         Unknown      needae bopsy: Epithelioid hemangioendothelioma. Labs from				hepatic <mark>epithelioid hemangioendothelioma</mark> and history	
Male       40      results/plan for her epithelioid hemangioendothelioma. Labs from metastatic epithelioid hemangioendothelioma. CTs have         Unknown      piphelioid hemangioendothelioma. Immunohistochemical         Unknown      piphelioid hemangioendothelioma. GROSS:        piphelioid hemangioendothelioma. COMMENT:        piphelioid hemangioendothelioma with enlarged        piphelioid hemangioendothelioma seen in subsequent				needie biopsy: <mark>Epitnelioid nemangioendotnelioma</mark>	
Unknown       ): Epithelioid hemangioendothelioma. Immunohistochemical	Male	40		results/plan for her <mark>epithelioid hemangioendothelioma</mark> . Labs from	
Unknown					
Unknown Unknow				): <mark>Epithelioid hemangioendothelioma</mark> . Immunohistochemical ): <mark>Epithelioid hemangioendothelioma</mark> . Immunohistochemical	
Unknown Unknow				of case Enitheliaid homonologindetheliama (POSS)	
metastatic <mark>epithelioid hemangioendothelioma</mark> with enlarged Metastatic <mark>epithelioid hemangioendothelioma</mark> consistent with <mark>epithelioid hemangioendothelioma</mark> seen in subsequent	Unknown			): Epithelioid hemangioendothelioma. COMMENT:	
Metastatic <mark>epithelioid hemangioendothelioma</mark> consistent with <mark>epithelioid hemangioendothelioma</mark> seen in subsequent	GIRIOWI			metastatic epithelioid hemangioendothelioma with enlaraed	
consistent with <mark>epithelioid hemangioendothelioma</mark> seen in subsequent				Metastatic <mark>epithelioid hemangioendothelioma</mark>	
				consistent with <mark>epithelioid hemangioendothelioma</mark> seen in subsequent	t

Patients Den	mo List (new) (1	2 Patients)											User	emerse	EMERSE -
Dates All [	Dates: 02/15/20	08 through 09/28/2011													
Terms abo	domen therap	y lungs "cancer of the brain"	"cancer of brain"	"tumor of brain"	"pontine glioma"	glioma	craniopharyng	gioma "	meninges canc	er" "ma	ignancy of bra	n" "CNA ca"	"CNS malignancy	"tumor o	More
Overview															
Overview															
Sorted By: Insert	Order 🛟 A	scending 🛟					M	Numbers	Grayscale	Mosaic					
MRN	I	Patient Name		Carew	eb	Radio	ology		Pathology						
1000000	1049 E	Bloom, Harrison		7 of 2	20	6 0	of 30		15 of 20						
1000000	1047 F	Patel, Joshua		9 of 1	10	5 0	of 20		3 of 30						
1000000	036	SCOUTTEN, MARILYN		21 of	25	1 0	of 10		1 of 5						
1000000	1073 E	Errazuriz, Alberto		2 of	10	1 0	of 10		13 of 15						
1000000	1048 F	Fay, Pat		9 of <sup>-</sup>	10	1 0	of 10		2 of 10						
1000000	040 (	Chen, John		9 of <sup>-</sup>	10				4 of 20						
1000000	1035 I	LUCCHESSI, VINCENZO					of 15		2 of 10						
1000000	075 \$	Sarchand, Nandita		1 of	5	1	of 5		9 of 10						
1000000	1065 I	_adwig, Renske				2 0	of 10		5 of 5						
1000000	050	Jacobs, Kelly				2 0	of 10		5 of 5		-				
1000000	1072	Walsh, Alan				1 0	of 10								
1000000	1044	Taylor, Jonathon	_	3 of	5						-				
			$\mathbf{}$								1				
			(f	ake n	ames	5)									

Patients Demo List (ne	ew) (12 Patients)				User: emerse EMERSE -
Terms cd20 cd5	cd34 cd10 cd3 cd45 "marginal zone lymph	oma" "BCL 2" BEXXAR ritux	imab cytoxan "R CHOP" ar	nthracycline autologous stem cell transpla	nt" ["diffuse large cell lymphoma" ["salvage therapy More
Overview					
Overview			,	Numbers Gravscale Mosaic	
Sorted By: Insert Order	Ascending 🗘				
MRN	Patient Name	Careweb	Radiology	Pathology	
100000049	Bloom, Harrison	-		•	
100000047	Patel, Joshua		-	<b>.</b> .	
100000036	SCOUTTEN, MARILYN	•			
100000073	Errazuriz, Alberto			-	
100000048	Fay, Pat	12.2			
100000040	Chen, John				
100000035	LUCCHESSI, VINCENZO				

Patients Demo List (new) (12 Patients)	User: emerse EMERSE +
Dates All Dates: 02/15/2008 through 09/28/2011	
Terms cd20 cd5 cd34 cd10 cd3 cd45 "marginal zone lymphoma" "BCL 2" BEXXAR rituximab cytoxan "R CH	HOP" anthracycline "autologous stem cell transplant" "diffuse large cell lymphoma" "salvage therapy More
Overview         Summaries         Patel, Joshua 1000000047         Comment         Tag         Patient 2 of 12         D	Document 4 of 10 Page Top Print Page Bottom
ADMITTING DIAGNOSIS: Relapsed Diffuse large cell lymphoma	
TRANSPLANT TYPE: Autologous	
PROTOCOL: CTN 0401 (Bexxar - BEAM)	
TRANSPLANT DATE: **DATE	
VIRAL TITERS: CMV positive, HSV positive, VZV positive	
HISTORY OF PRESENT ILLNESS: Mr. **NAME is a **AGE year old male who is being admitted for an autologous stem cell transplant per BMT CTN ; randomized to Bexxar-BEAM therapy. Mr. **NAME is currently admitted to undergo the BEAM portion of his conditioning regimen. Autologous stem cell transplant is scheduled for **DATE.	
PAST MEDICAL HISTORY: The patient presented in **DATE, with proptosis of the right eye and sinus congestion. A CT scan at that time revealed a 3.4 cm x 5.8 cm right orbital mass, and 6.7 cm x 3.7 cm right maxillary mass. The patient had developed significant exophthalmos of the right eye with visual changes along with coexistent bilateral cervical adenopathy. A biopsy of the mass was obtained on **DATE, initially consistent with a CD20 positive, GDIS positive, MELTZ positive, marginal zone lymphoma. The patient was subsequently treated with 5 cycles of R-CHOP between **DATE, and **DATE.	
A second-look surgery performed on **DATE, of the right ethmoid and right maxillary sinus mass revealed that the lesion had transformed to a diffuse large cell lymphoma, again CD20 positive, CD45 positive, BCL-2 positive. The patient subsequently received	
* additonal cycles of <mark>R-CHOP</mark> between **DATE, through **DATED, YYYY>. He has now received a total <mark>anthracyc</mark> m. A CT scan obtained at the end of induction therapy revealed a CR status. No XRT was given.	cline dose of 240 mg/sq.
In **DATE, the patient developed recurrent nasal congestion with the lacrimal duct obstruction. A CT scan on **DATED, YYYY>, revealed a new left maxillary mass, 4 diameter. The mass extended from the midline of the oropharynx laterally to the point of the lateral pterygoid plate on the left side. Thus, the patient recurred contralateral to the initial side of disease. There was no other adenopathy in the chest or abdomen on screen at that time. The patient started salvage therapy with RICE **DATE. He has received 2 cycles of RICE to date, the 2nd cycle given **DATE. He enrolled on CTN in **DATE, receiving <u>Situxinab</u> , and <u>Cytoxan</u> 4 gm/m2 . He underwent autologous stem cell collection **DATED>-**DATE, with 4.4 x 10(6) CD34 cells/	4.5 cm in maximal

Patients All (2,351,932)	EMERSE -
Dates All Dates: 01/01/1900 through 05/18/2018	
Terms emesis	
Quick Terms Term Bundles Advanced Terms	
Name/Description         Manage Terms         View Terms         Sharing         Clear/Delete	
The current settings will enable you to search through All Patien Click individual terms to highlight or de-highlight Vertical View *	
Terms with different colors will be searched using the AND oper Patients identified will be those that have at least one document	
With a Quick Terms search, EMERSE assigns the colors to the te barf barfed barfer barfing barfs blowing chunks chunder chundered chundering chunders dry heaved dry heaved dry heaves	
The current screen looks similar to a <b>Term Bundle</b> screen, but be adjust the colors of the terms yourself.	
For more control over the colors assigned to terms, use a Term B induce vomiting induced vomiting induces vomiting induces vomiting inducing vomiting N & V N and V N&V N/V nausea nausea & vomiting nausea & vomiting	
Search Options nausea / vomiting nausea and vomiting nausea or vomiting nausea vomiting nausea/vomiting nauseated nauseated nauseated nauseated nauseating	
nauseous nauseousness NBNB emesis non-bloody, non-bilious emesis non-bloody, nonbilious emesis nonbloody, non-bilious emesis	
Terms to include nonbloody, nonbilious emesis NV post-anesthesia emesis post-anesthesia vomiting post-anesthetic emesis post-anesthetic vomiting	
Click on the term to edit	
emesis Synonyms Synonyms post-tussive emesis postanesthesia emesis postanesthesia vomiting postanesthetic emesis postanesthetic vomiting postanestheti	
puked puker pukes puking purge purged purger purges purging queasiness queasy ralph ralphs regurgitated regurgitated	
regurgitates regurgitating retch retched retches retching reverse peristalsis threw it all up threw it up threw up throw it all up	
throw it up throwing throwing chunks throwing it all up throwing it up throwing up throws it all up throws it up throws up tossing cookies	
up-chuck up-chucked up-chucking up-chucks upchuck upchuck upchucking upchucks vomit vomited vomiter vomiters vomiting	
vomitive vomitory vomits vomitus	
Spelling Alternatives (9)	
emeisis emeisis emeisis emsis metic metics nause and vomitting queaziness	
Highlight All De-Highlight All Add Highlighted Terms	

Patients     Demo List (ne       Dates     All Dates: 02/       Terms     emesis       Overview	w) (12 Patients) 15/2008 through 09/28/2011				User: er	nerse EME	RSE -
Overview Sorted By: Insert Order	Ascending			Numbers Grayscale Mosaic			
MRN	Patient Name	Careweb	Radiology	Pathology			
100000049	Bloom, Harrison			-			
100000047	Patel, Joshua						
100000036	SCOUTTEN, MARILYN						
100000073	Errazuriz, Alberto						
100000048	Fay, Pat						
100000040	Chen, John						
100000035	LUCCHESSI, VINCENZO						

Patients Demo List (ne	ew) (12 Patients)				User: emerse EMERSE +
Terms barf barfe	barfer barfing barfs "blowing chunks" "dr	y heave" "dry heaved" "dry hea	aves" "dry heaving" emetic	emetics emetogenic emetogen	icity heave heaved heaves heaving hurl hurled h More
Overview					
Overview				Number Original News	
Sorted By: Insert Order	Ascending \$			<sup>1</sup> 2 <sub>3</sub> Mosaic	
MRN	Patient Name	Careweb	Radiology	Pathology	
100000049	Bloom, Harrison				
100000047	Patel, Joshua				
100000036	SCOUTTEN, MARILYN	<b></b>			
100000073	Errazuriz, Alberto			<b></b>	
100000048	Fay, Pat	2.4			
100000040	Chen, John	2.50			
100000035	LUCCHESSI, VINCENZO				

## WHERE IS EMERSE USED?





"In our clinical research studies, EMERSE allowed us to mine the data within the EHR to identify adverse events (for the phase 1/2 clinical trial), scores of acute graft-versus-host disease, body mass indices (weight and height), molecular genes (presence of FLT3+), as well as other hematopoietic cell transplant related complications."

"Our laboratories focus on translational research targeted at understanding the molecular mechanisms involved in the development of adrenocortical cancer. Using EMERSE we extracted many clinical characteristics from the patient charts including tumor staging, the reason the biopsy was done, complications resulting from the biopsy, and other salient details."

"I am a laboratory investigator with research focus on using proteomics to develop biomarkers to better predict the onset and severity of graft versus host disease (GVHD). Because of the translational nature of my work, it is necessary to integrate my lab-based findings with what our patients actually experience clinically...to link my biomarker data to patient phenotypic data, treatment specifics, and detailed clinical outcomes. It would have been nearly impossible to have carried out this work in an efficient and accurate manner without the use of EMERSE."

"EMERSE was used to search through all patients with head and neck cancer over a 13-year period to find women who were pregnant at the time of the cancer. The four patients identified (all with tongue cancer), were then analyzed further for biomarkers and HPV status. EMERSE helped tremendously by rapidly narrowing this rare patient cohort from among thousands..."

## **PROJECT PLANS**

- Securely network sites for cross-institution searches
- Further develop detailed documentation
- Improve local handling of synonyms
- Incorporate de-identified documents
- Feature and usability enhancements
- User evaluation
- Dissemination

## WHAT WE OFFER

- Installation files and/or source code
- Virtual machine to "kick the tires" (or load your own data)
- Instructions for installation, data integration, and operation (under development)
- User guide (under development)
- Synonyms file (license from U of Michigan)
- Technical support

#### COLLABORATORS



## ACKNOWLEDGEMENTS









Health Information Technology & Services

ITCR (NCI): U24CA204863 U of Michigan Rogel Cancer Center (NCI): P30CA046592 MICHR (CTSA): UL1TR000433 and UL1TR002240

#### WEBSITE

# http://project-emerse.org

