Cancer Deep Phenotype Extraction from Electronic Medical Records (DeepPhe) – Year 2

Guergana K. Savova, PhD
Guergana.Savova@childrens.harvard.edu

Associate Professor
Boston Children's Hospital
Harvard Medical School

Rebecca Crowley Jacobson, MD, MS
rebeccaj@pitt.edu

Professor
Department of Biomedical Informatics
University of Pittsburgh Cancer Institute

Precision Medicine Initiative Working Group Final Report

"Identifying specific clinical phenotypes from EHR data require use of algorithms incorporating demographic data, diagnostic and procedure codes, lab values, medications, and natural language processing (NLP) of text documents."

"Such 'deep phenotyping', as it is known, gathers details about disease manifestations in a more individual and finer-grained way, and uses sophisticated algorithms to integrate the resulting wealth of data with other…information.

DEEP PHENOTYPING

The details of disease

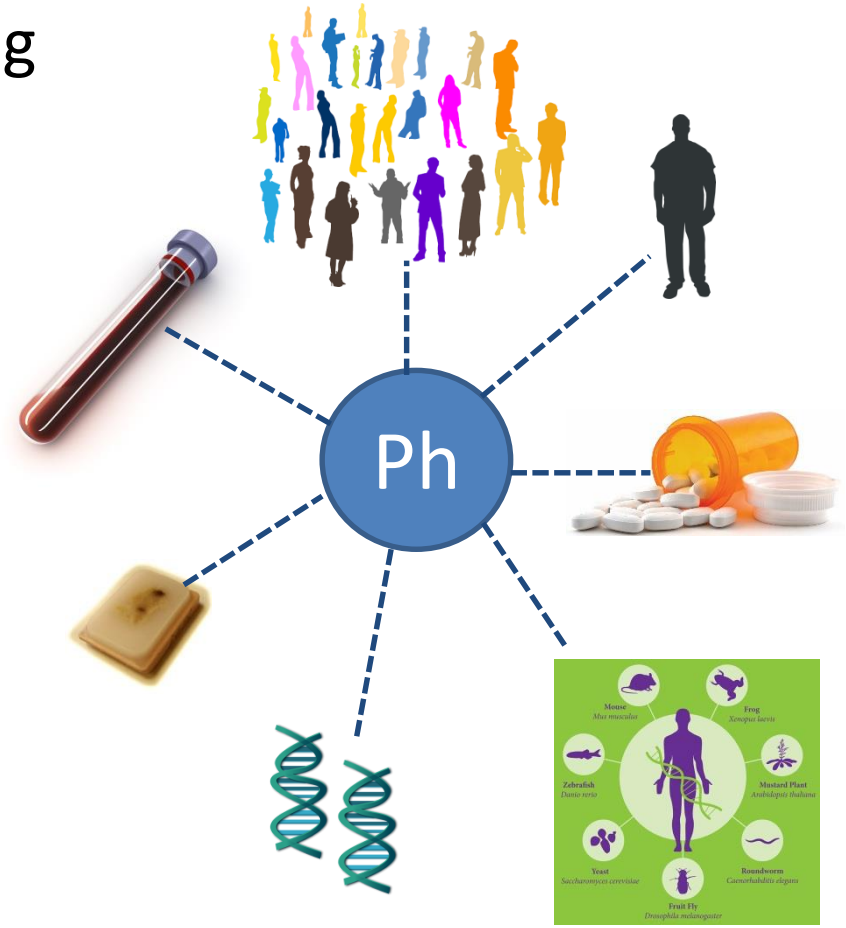
Nature, November 5 2015

FACT SHEET: Investing in the National Cancer Moonshot

"…will encourage data sharing and support the development of new tools to leverage knowledge about genomic abnormalities, as well as the response to treatment and long-term outcomes.

Phenotyping Use Cases

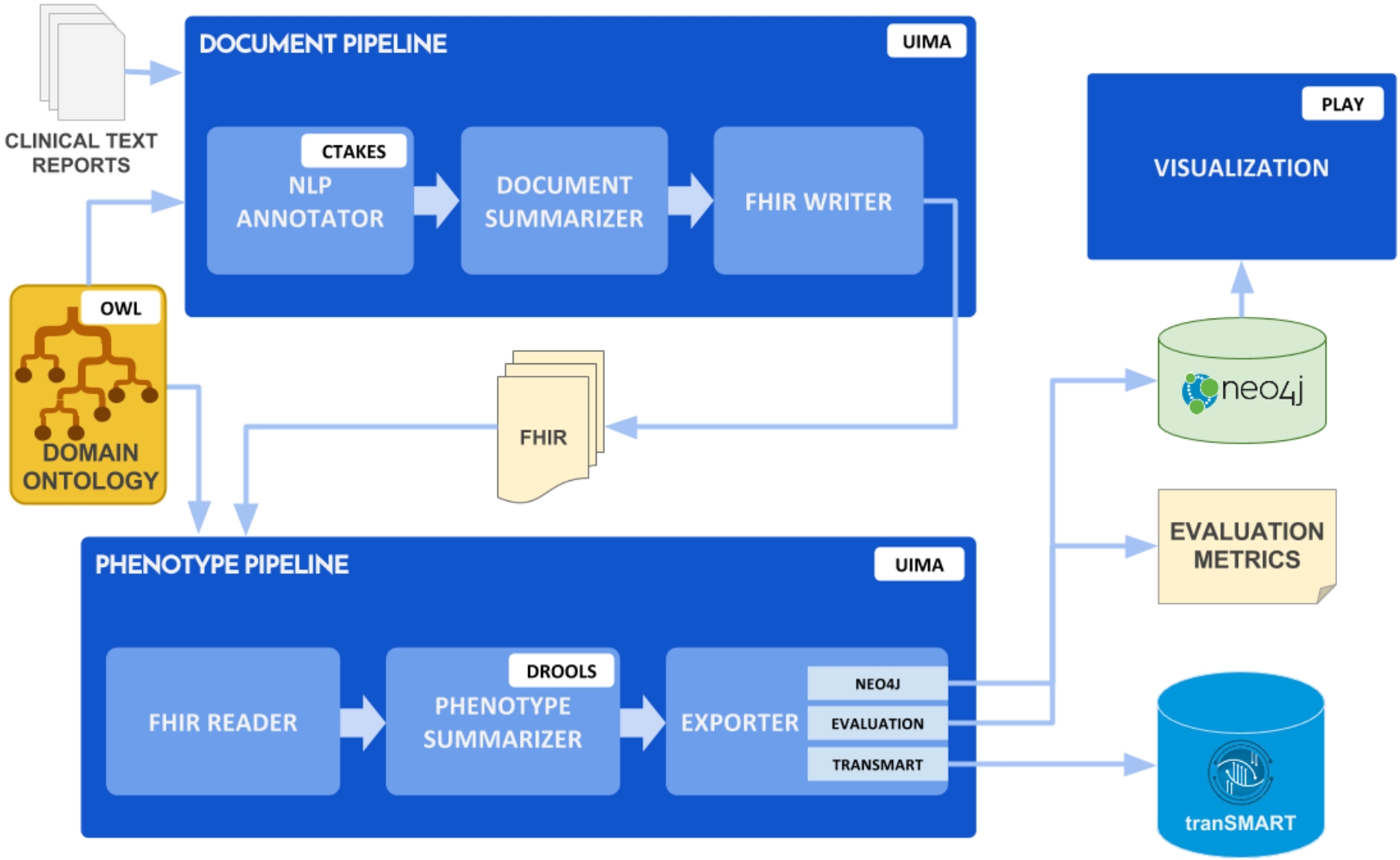
- Cohort discovery supporting translational science
- Targeted Therapeutics and Personalized Medicine
- Biomarker Discovery and Validation
- Pharmacogenomics
- Pharmacovigilence
- Disease Surveillance
- Drug repurposing
- Point of care
- ….



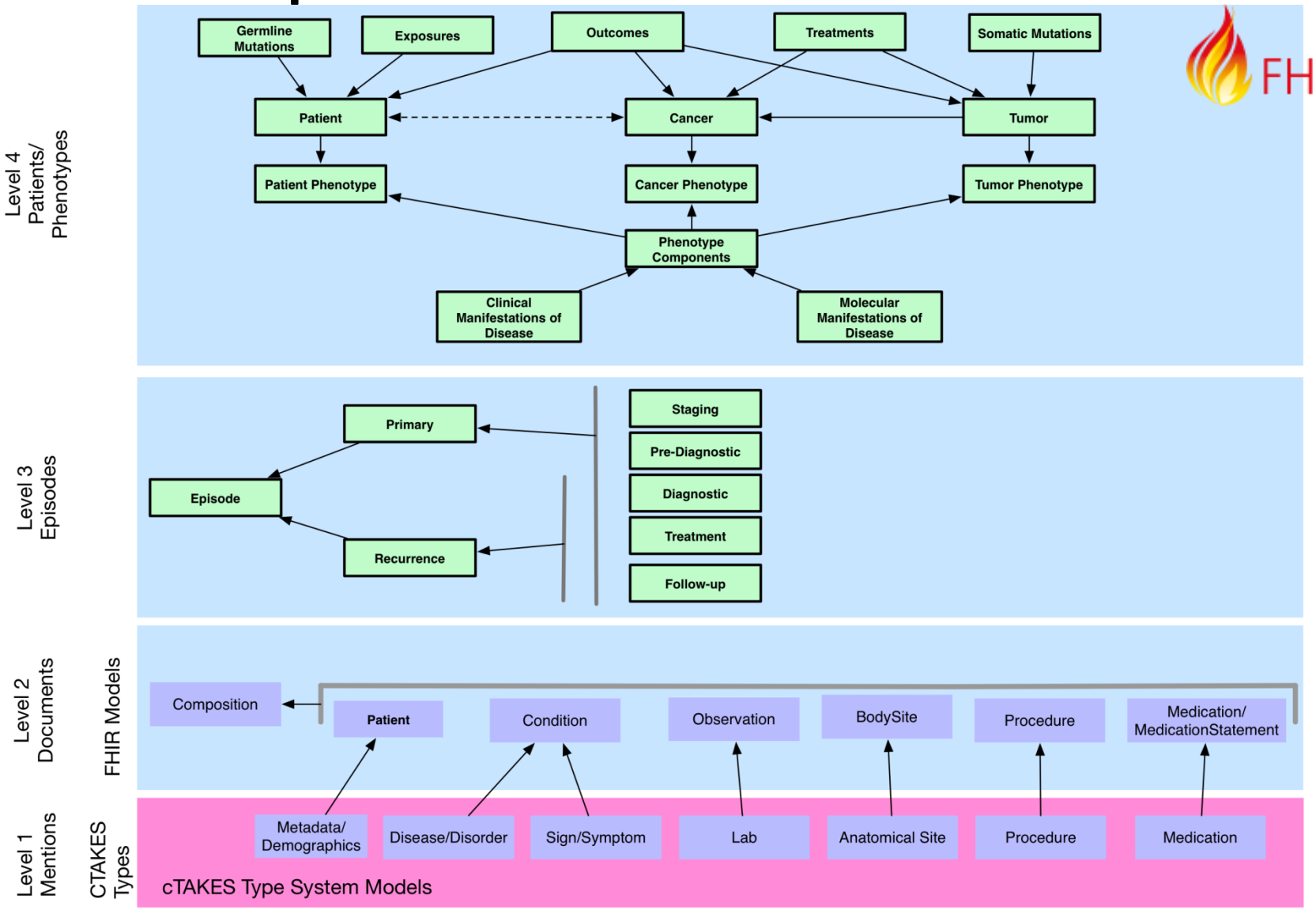
DeepPhe Project http://cancer.healthnlp.org

- Collaboration between DBMI and BCH
- Goal is to develop next generation cancer deep phenotyping methods
- Addresses information extraction but also representation and visualization
- Support **high throughput approach** – process and annotate all data at multiple levels (from mention to phenotype) and across time
- Combine IE with structured data (cancer registry)
- Develop phenotyping rules/reasoners/classifiers
- Driven by translational research scientific goals

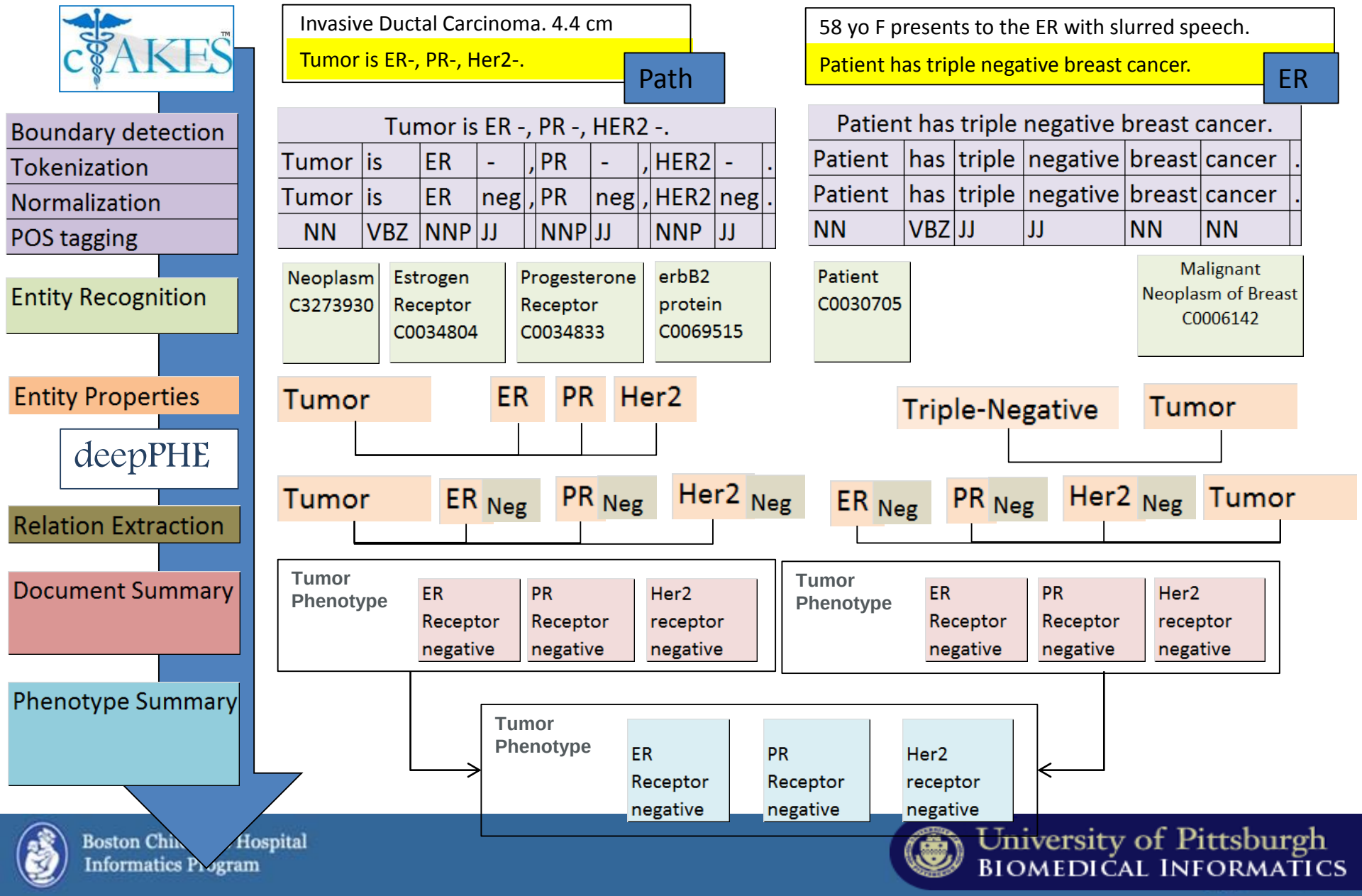
Architecture



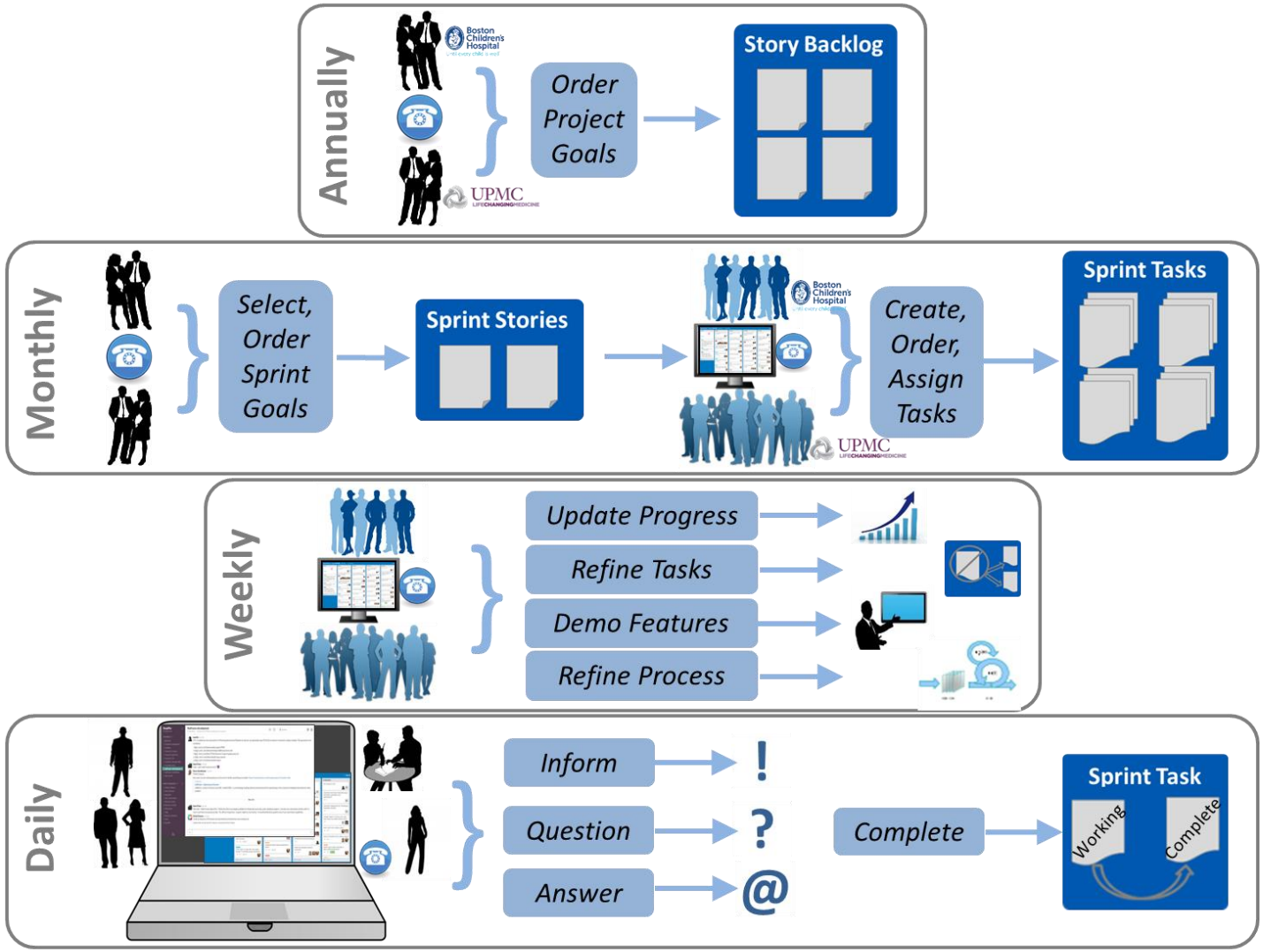
DeepPhe Information Model



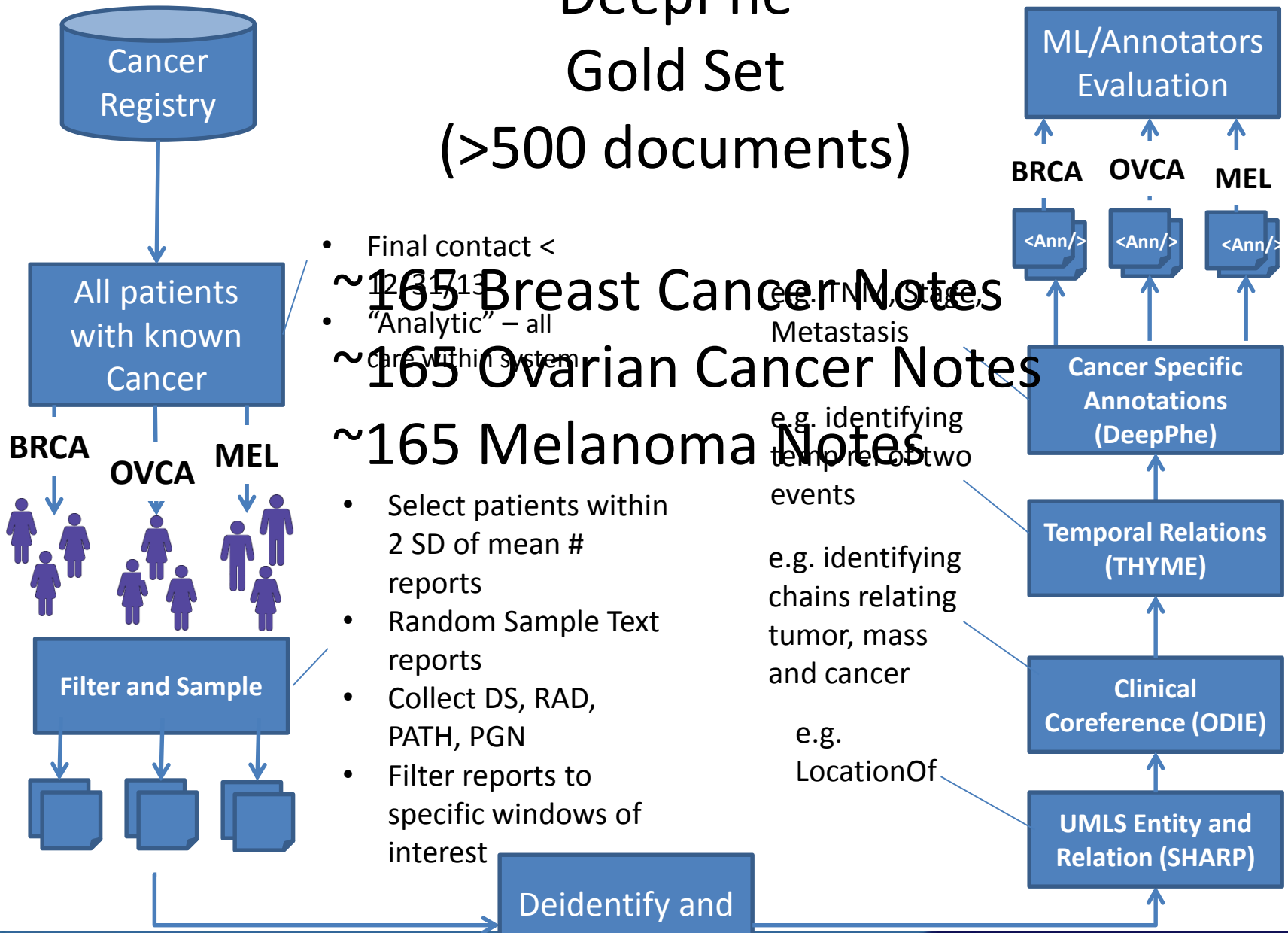
DeepPhe NLP Pipeline



Software Development Process



DeepPhe Gold Set (>500 documents)



- Final contact < 12/31/13
- "Analytic" – all care within system

~165 Breast Cancer Notes

~165 Ovarian Cancer Notes

~165 Melanoma Notes

- Select patients within 2 SD of mean # reports
- Random Sample Text reports
- Collect DS, RAD, PATH, PGN
- Filter reports to specific windows of interest

e.g. TNM, Stage, Metastasis

e.g. identifying temporal of two events

e.g. identifying chains relating tumor, mass and cancer

e.g. LocationOf

Evaluation Results

Template Instance Distribution: BrCa		
	#instances in train split	#instances in development split
stage	6	5
tnm	14	30
receptors	62	50
metastasis	54	58
train split: 4 patients, 48 documents		
development split: 2 patients, 42 documents		

Results on BrCa development split: Template Instance System vs. Gold (results in brackets are for the Inter-annotator agreement)				
	stage	**tnm**	**receptors**	**metastasis**
overlapping span of template anchor (mention instance)				
Precision/PPV	1 (1)	1 (1)	1 (1)	0.94 (1)
Recall/Sensitivity	1 (1)	0.97 (1)	0.68 (0.81)	0.78 (0.3)
F1	1 (1)	0.98 (1)	0.81 (0.89)	0.85 (0.46)
attribute accuracy				
*conditional	1 (1)	n/a	1 (1)	1 (1)
*uncertainty	1 (1)	1 (1)	1 (1)	0.2 (1)
*negation	1 (1)	n/a	1 (1)	0.9 (0.66)
*subject	1 (1)	n/a	1 (1)	1 (1)
*generic	1 (1)	n/a	1 (1)	0.7 (1)
associated neoplasm (span)	1 (1)	0.62 (0.79)	0.5 (0.86)	0.56 (0.64)
concept unique identifier (CUI)	1 (1)	1 (1)	0.98 (1)	-
body location (span)	n/a	n/a	n/a	0.76 (1)
test method (CUI)	n/a	n/a	0.92 (0.78)	n/a
* indicates weighted accuracy per SemEval 2015 to take into account default value prevelance rates				

Evaluation Results

Cancer Template Distribution: BrCa	
	#instances in corpus
cancer	6
body location	14
body location side	8
clinical stage	3
clinical T value	2
clinical N value	2
clinical M value	2
pathologic T value	3
pathologic N value	3
pathologic M value	1
corpus: 6 patients, 90 documents	

Results on BrCa Train & Development: Phenotype System vs. Gold (results in parentheses are for inter-annotator agreement)			
	Precision/PPV	**Recall/Sensitivity**	**F1 measure**
cancer	0.81 (1)	0.81 (1)	0.81 (1)
body location	0.52 (1)	1 (1)	0.69 (1)
body location side	0.5 (n/a)	1 (n/a)	0.67 (n/a)
clinical stage	1 (0.80)	1 (1)	1 (0.89)
clinical T value	0.40 (0.89)	1 (1)	0.57 (0.94)
clinical N value	1 (0.89)	1 (1)	1 (0.94)
clinical M value	1 (0.89)	1 (1)	1 (0.94)
pathologic T value	0.75 (0.89)	1 (1)	0.86 (0.94)
pathologic N value	0.75 (0.78)	1 (0.88)	0.86 (0.82)
pathologic M value	1 (0.62)	1 (1)	1 (0.77)

Evaluation Results

Tumor Template Distribution: BrCa	
	#instances in corpus
tumor	15
body location	15
body location side	11
body clockface	6
body quadrant	5
diagnosis	14
tumor type	15
er interpretation	8
er method	5
pr interpretation	8
pr method	5
her2 interpretation	7
her2 method	5
corpus: 6 patients, 90 documents	

Results on BrCa Train & Development: Phenotype System vs. Gold (results in parentheses are for inter-annotator agreement)			
	Precision/PPV	**Recall/Sensitivity**	**F1 measure**
tumor	0.37 (0.79)	0.69 (0.88)	0.48 (0.84)
*body location	1 (n/a)	1 (n/a)	1 (n/a)
*body location side	1 (n/a)	1 (n/a)	1 (n/a)
body clockface	0.67 (0.89)	0.40 (0.73)	0.50 (0.80)
body quadrant	1 (0.73)	0.2 (0.80)	0.33 (0.76)
diagnosis	0.47 (0.93)	0.88 (0.93)	0.61 (0.93)
tumor type	1 (1)	1 (1)	1 (1)
er interpretation	0.75 (1)	0.60 (1)	0.67 (1)
er method	1 (1)	0.25 (1)	0.4 (1)
pr interpretation	0.75 (1)	0.5 (1)	0.67 (1)
pr method	NaN (1)	0 (1)	NaN (1)
her2 interpretation	0.67 (1)	0.5 (1)	0.57 (1)
her2 method	0.5 (0.83)	0.25 (0.83)	0.33 (0.83)
*attribute used to align system and gold annotations			

Publications and Collaborations

- Towards Portable Entity-Centric Clinical Coreference Resolution (submitted to the Journal of the Medical Informatics Association)
- An Information Model for Cancer Phenotypes (submitted to BMC Medical Informatics and Decision Making)
- Improving Temporal Relation Extraction with Training Instance Augmentation (submitted to the BioNLP workshop at the Association for Computational Linguistics conference)
- ITCR Supplement to build tools for TCGA clinical data and metadata with Mayo caCDE QA (see our poster)
- Supplement to work with SEER to extend DeepPhe methods to cancer surveillance
- Collaboration with THYME (thyme.healthnlp.org)

Goals for Next Year

- IE methods
  - Coreference
  - Temporal relations
  - Template filling improvement
- Additional templates for Procedures, Medications, Clinical Genomics, Tumor size
- New model for Ovarian Cancer
- Merging information from structured and unstructured EMR
- Visualization of patient timelines
- Evaluation of system with breast cancer clinical research questions (using EMR data from Pitt TCGA patients)

DeepPhe

deepphe.boston

Guergana Savova, MPI
Sean Finan
Timothy Miller
Dmitriy Dligach
Chen Lin
David Harris

deepphe.pgh

Rebecca Jacobson, MPI
Harry Hochheiser
Girish Chavan
Eugene Tseytlin
Olga Medvedeva
Melissa Castine
Mike Davis
Adrian Lee
John Kirkwood
Francesmary Modugno

Funding

NCI U24 CA132672 Cancer Deep Phenotyping from Electronic Medical Records (Savova and Jacobson, MPIs)

Demo

https://youtu.be/61gelUfD3VU

EXTRA SLIDES

cTAKES Component or Function	Score	Score Type
Sentence boundary [2]	0.949	Accuracy
Context sensitive tokenizer [2]	0.949	Accuracy
Part-of-speech tagging [2] [10]	0.936 – 0.943	Accuracy
Shallow parser [2]	0.952 ; 0.924	Accuracy ; F1
Entity recognition [2]	0.715 / 0.824	F1 [1]
Concept mapping (SNOMED CT and RxNORM) [2]	0.957 / 0.580	Accuracy [1]
Negation NegEx [11] [2]	0.943 / 0.939	Accuracy [1]
Uncertainty, modified NegEx [11] [2]	0.859 / 0.839	Accuracy [1]
Constituency parsing [12]	0.810	F1
Dependency parsing [10]	0.854 / 0.833	F1 [2]
Semantic role labeling [10]	0.881 / 0.799	F1 [3]
Coreference resolution, within-document [12]	0.352 ; 0.690 ; 0.486 ; 0.596	MUC ; B^3 ; CEAF ; BLANC
Relation discovery [13]	0.740-0.908 / 0.905-0.929	F1 [4]
Events (publication in preparation)	0.850	F1
Temporal expression identification [14]	0.750	F1
Temporal relations: event to note creation time [15]	0.834	F1
Temporal relations: on i2b2 challenge data [15]	0.695	F1