Federated Data Sharing, Natural Language Processing and Deep Phenotyping to Advance Precision Medicine

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

Department of Biomedical Informatics
Institute for Personalized Medicine
University of Pittsburgh Cancer Institute

Precision Medicine Initiative Working Group Final Report

"Identifying specific clinical phenotypes from EHR data require use 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

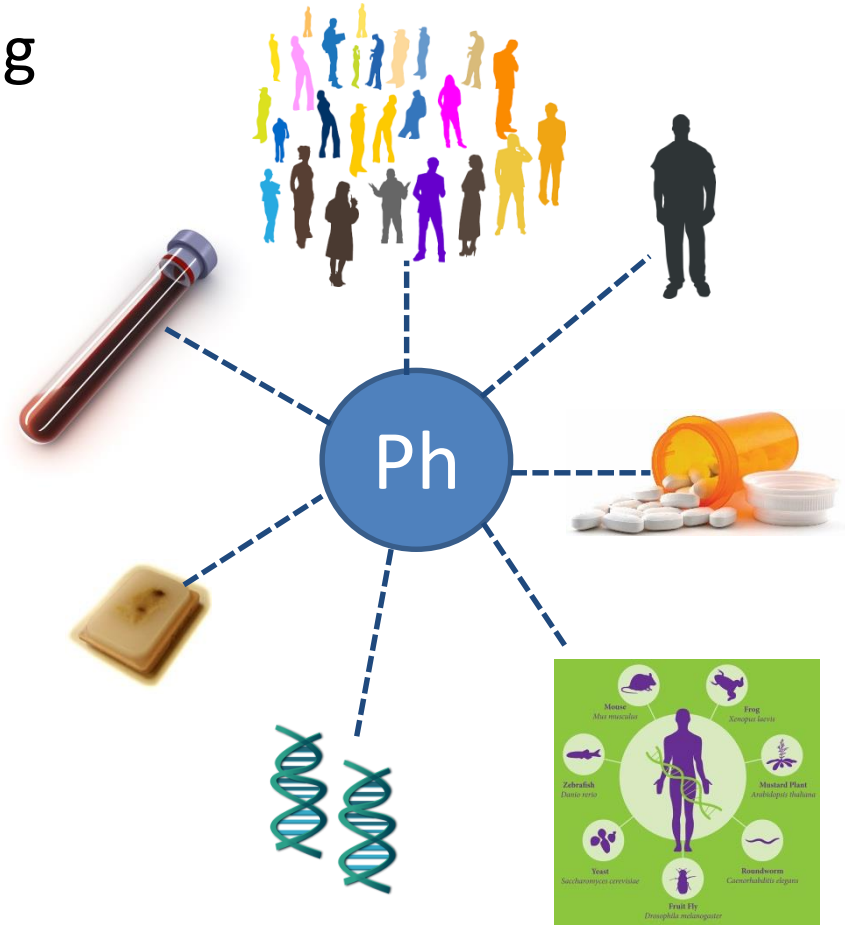
Nature, November 5 2015

FACT SHEET: Investing in the National Cancer Moonshot

"Data sharing can break down barriers between institutions, including those in the public and private sectors, to enable maximum knowledge gained and patients helped."

Phenotyping Use Cases

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

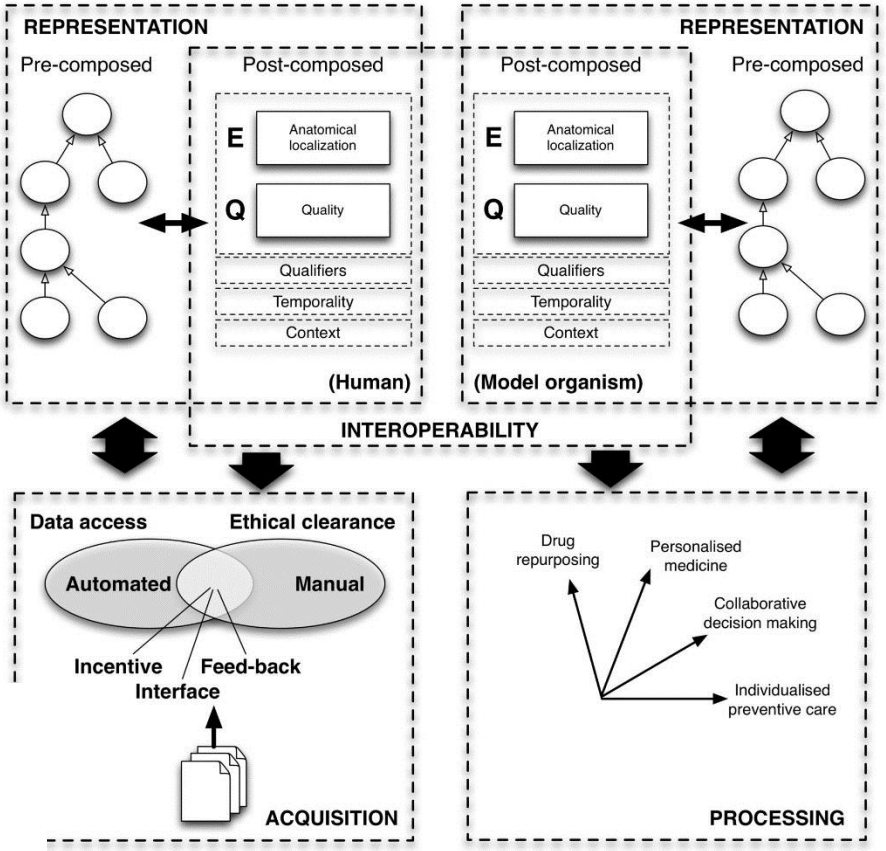


What's different now?

- Speed and depth in which we can interrogate the human genome, cancer genome, microbiome
- Widespread adoption of EMRs, availability of data with positives and negatives
- Increasingly challenging technical and regulatory environment
- Advancing science of Ontology and NLP, model organisms, more widespread use
- Ability to aggregate data across organizations makes it possible to identify and study rare phenotypes
- 'Direct to consumer' phenotyping (e.g. 23andMe)

Cancer Phenotyping

- Less about defining a specific cohort with the disease (except for familial risk)
- More about identifying specific subpopulations with different behaviors that can drive forward molecular classification, systems biology, inform treatment decisions
- Annotation of cancer specimens (retrospective and prospective) will be a critical factor in success



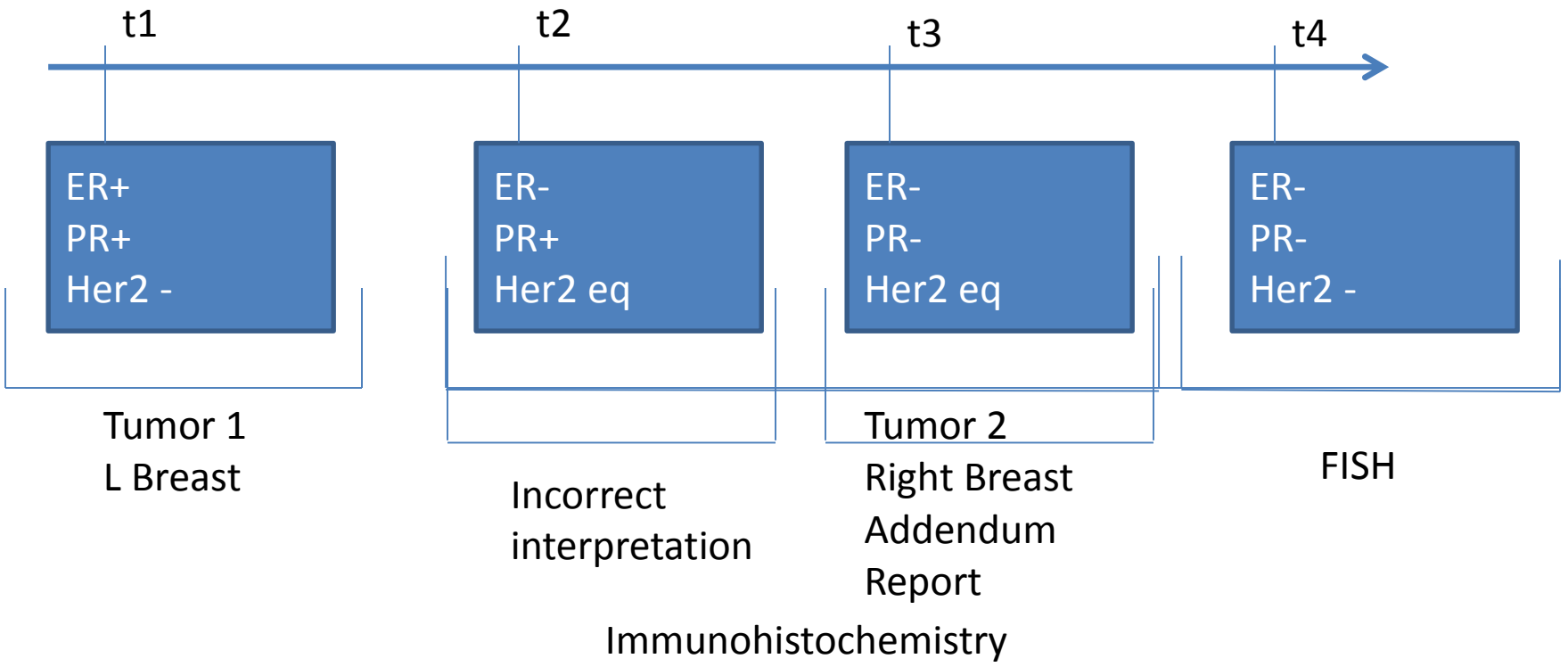
The digital revolution in phenotyping

Anika Oellrich*, Nigel Collier*, Tudor Groza*, Dietrich Rebholz-Schuhmann*, Nigam Shah*, Olivier Bodenreider, Mary Regina Boland, Ivo Georgiev, Hongfang Liu, Kevin Livingston, Augustin Luna, Ann-Marie Mallon, Prashanti Manda, Peter N. Robinson, Gabriella Rustici, Michelle Simon, Liqin Wang, Rainer Winnenburg and Michel Dumontier

Corresponding author. Anika Oellrich, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, United Kingdom. E-mail: anika.oellrich@kcl.ac.uk
* These authors contributed equally to this work.

Anika Oellrich et al. Brief Bioinform 2015

The Acquisition Problem: What is the ER, PR, Her2 status of this patient?



Biomarker Phenotyping Rules: Her2/Neu Values

- Her2/Neu phenotype preferentially obtained from Pathology Report
- Lab interpretation (e.g. “positive”) should be extracted in preference over raw scores (e.g. “2+”)
- In the absence of explicit interpretation statement:
  - IHC 3+ or greater is positive
  - IHC 0 or 1+ is negative
  - IHC 2+ is equivocal
- FISH/CISH interpretation of
  - “amplified” = positive
  - “not amplified” = negative
- Value from FISH or CISH takes precedence over IHC except when IHC is done after an equivocal FISH test

Importance of NLP/IE

- Despite progress, most important clinical data for phenotyping still mainly free text (especially pathology and radiology)
- One of many seminal lessons from eMERGE is that NLP is important to accurate EMR-based phenotyping
- Challenges abound: temporality, summarization, coreference resolution, cross-document, task-specific
- Going from mentions (and/or EHR result) to phenotypes for a wide range of purposes will require advances in the science of phenotyping

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

Example phenotypic features

- Histopathologic type
- TNM Stage
- Comorbidities
- Medications
- Biomarkers
- Clinical mutation testing (e.g. BRAF)
- Local recurrence
- Distant recurrence

Apache cTAKES: Sample Pipeline

The patient underwent a CT scan in April which did not reveal lesions in his liver.

Boundary Detection

…	The patient underwent a CT scan in April which did not reveal lesions in his liver.	…

Tokenization

The	patient	underwent	a	CT	scan	in	April	which	did	not	reveal	lesions	in	his	liver	.

Normalization

-	-	undergo	-	-	-	-	-	-	do	-	-	lesion	-	-	-	

Part-of-speech Tagger

DT	NN	VBD	DT	NN	NN	IN	NNP	WDT	VBD	RB	VB	NNS	IN	PRP$	NN	.

Entity Recognition

CT scan	Lesion	Liver
Procedure	Disease / Disorder	Anatomy
UMLS ID: C0040405	UMLS ID: C0022198	UMLS ID: C0023884

Biomedical End-Use

Chunking

NP	VP	NP	PP	NP	VP	NP

Constituency Parsing

S — NP — DT — NN — VP … — …

Dependency Parsing

…

SRL

undergo.01 (A1.patient; A2.scan; AM-TEMP.in)
reveal.01 (A0.scan; R-A0.which; AM-NEG.not; A1.lesions; AM-LOC.in)

Entity Properties

CT scan	Lesion	Liver
Negated: no	Negated: yes	Negated: no
Subject: patient	Subject: patient	--

UMLS Relation

locationOf (lesions, liver)

Biomedical End-Use

Event, Temp. Expr.

CT scan	April	Reveal	Lesions

Temporal Relation

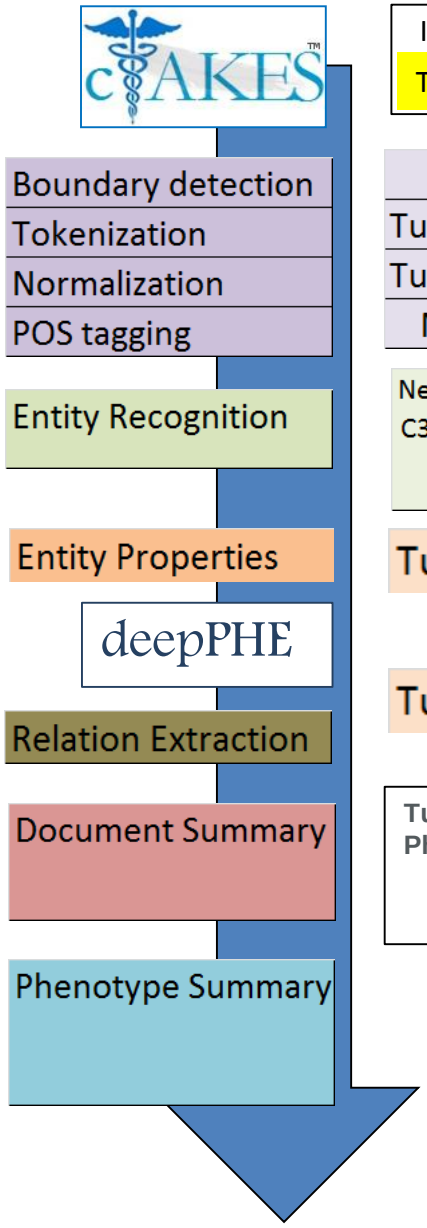
April CONTAINS CT scan
CT scan CONTAINS lesions

Coreference

identity (the patient, his)

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

DeepPhe NLP Pipeline



Invasive Ductal Carcinoma. 4.4 cm

Tumor is ER-, PR-, Her2-.

Path

Tumor is ER -, PR -, HER2 -.									
Tumor	is	ER	-	,	PR	-	,	HER2	-	.
Tumor	is	ER	neg	,	PR	neg	,	HER2	neg	.
NN	VBZ	NNP	JJ		NNP	JJ		NNP	JJ	

Neoplasm C3273930 | Estrogen Receptor C0034804 | Progesterone Receptor C0034833 | erbB2 protein C0069515

Tumor — ER — PR — Her2

Tumor — ER Neg — PR Neg — Her2 Neg

Tumor Phenotype: ER Receptor negative | PR Receptor negative | Her2 receptor negative

58 yo F presents to the ER with slurred speech.

Patient has triple negative breast cancer.

ER

Patient has triple negative breast cancer.					
Patient	has	triple	negative	breast	cancer	.
Patient	has	triple	negative	breast	cancer	.
NN	VBZ	JJ	JJ	NN	NN	

Patient C0030705 | Malignant Neoplasm of Breast C0006142

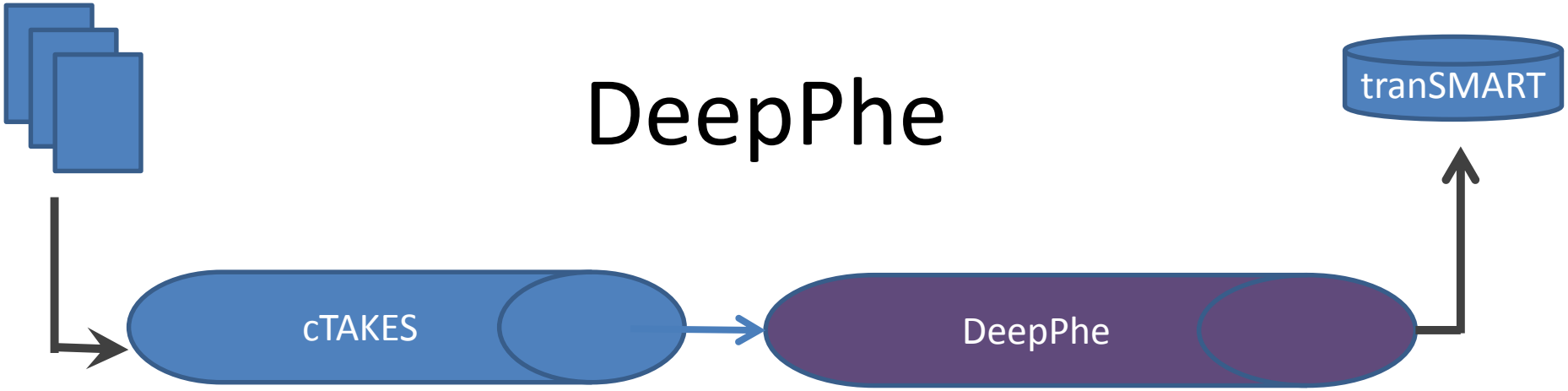
Triple-Negative — Tumor

ER Neg — PR Neg — Her2 Neg — Tumor

Tumor Phenotype: ER Receptor negative | PR Receptor negative | Her2 receptor negative

Tumor Phenotype: ER Receptor negative | PR Receptor negative | Her2 receptor negative

DeepPhe



Detailed NLP annotation (information extraction)

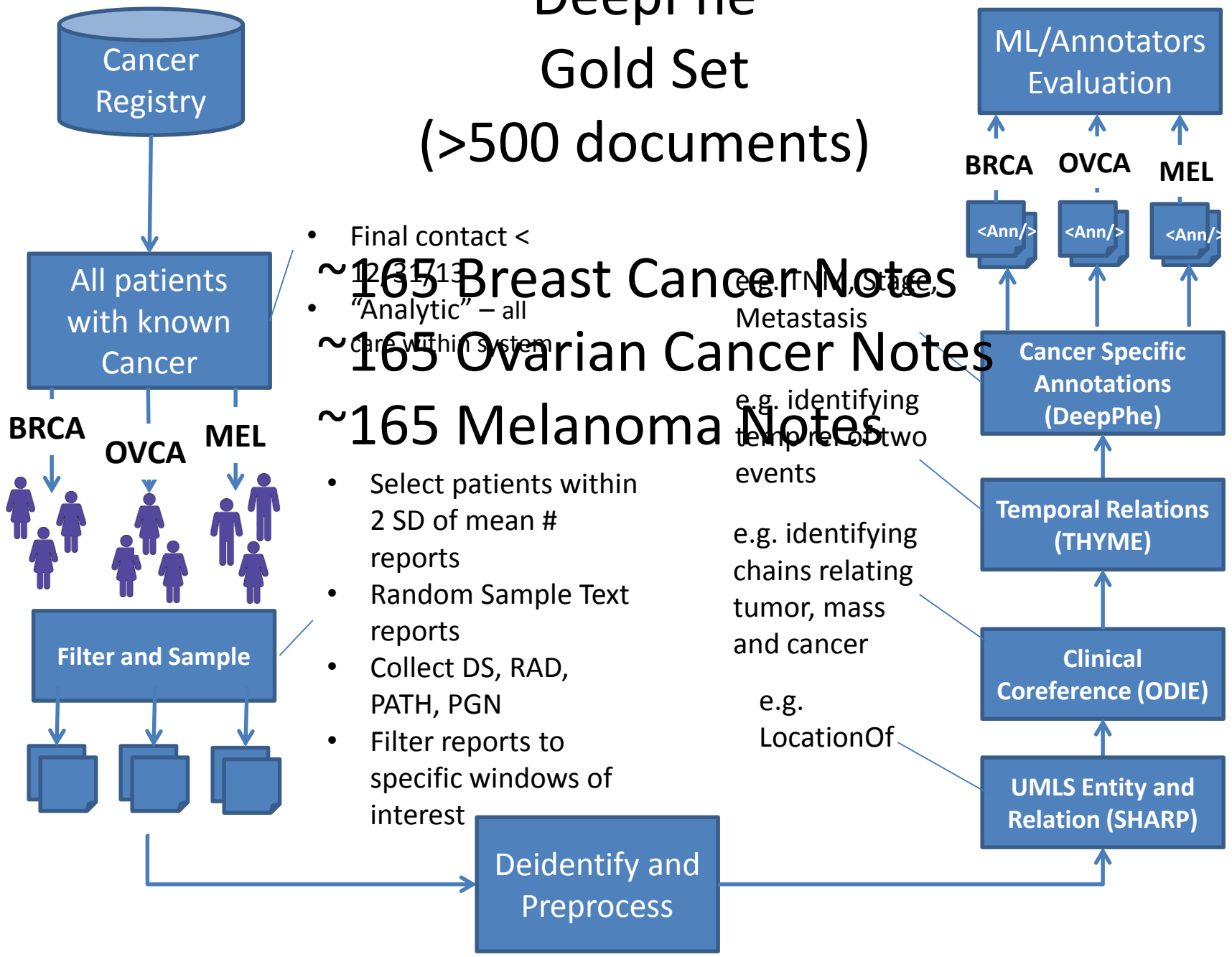
- Concepts (e.g. Dis/Disord, AL, procedure, findings, medication statements)
- Negation
- Coreference chains
- Temporal expressions and relations
- Relation extraction (e.g. attribute-value pairs)

Document Summarization

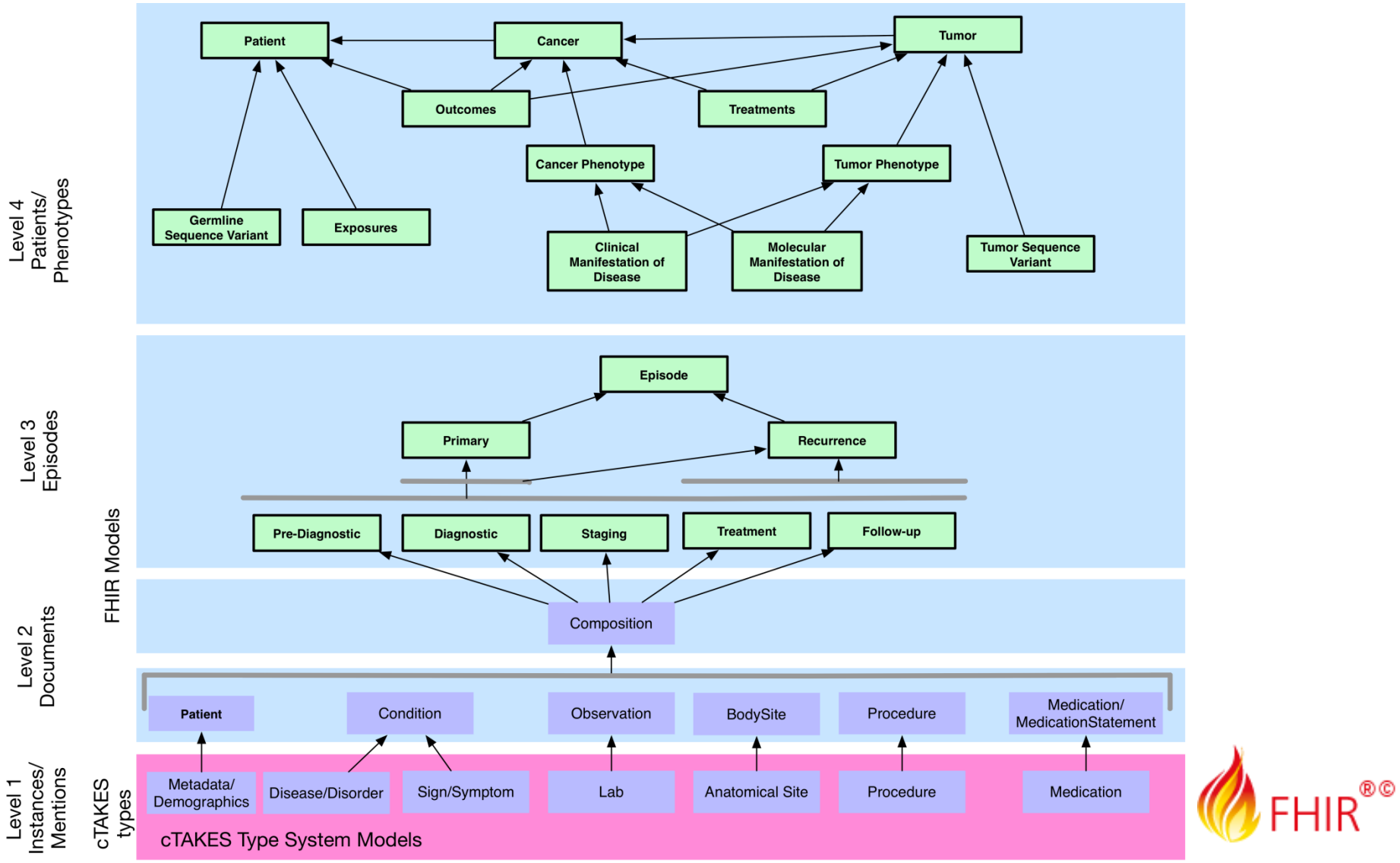
- Classification from multiple mentions
- Relationship of mentions to higher level entities (e.g. tumor)

Phenotype Summarization

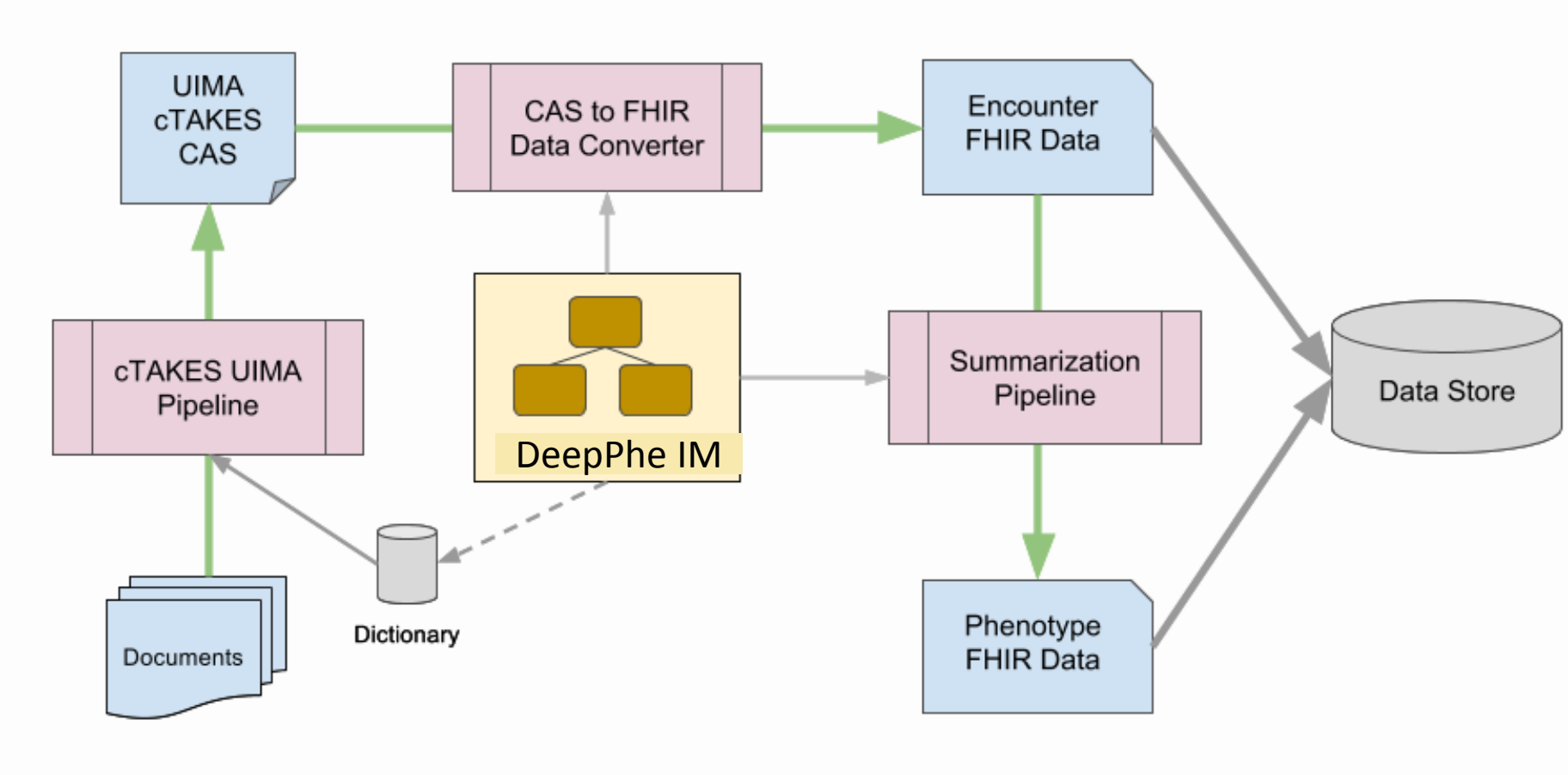
- Summarization across documents of different genre
- Incorporation of structured data with unstructured data


DeepPhe
Gold Set
(>500 documents)
~165 Breast Cancer Notes
~165 Ovarian Cancer Notes
~165 Melanoma Notes
Cancer Registry
All patients with known Cancer
BRCA
OVCA
MEL
Filter and Sample
Final contact < 12/31/13
"Analytic" – all care within system
Select patients within 2 SD of mean # reports
Random Sample Text reports
Collect DS, RAD, PATH, PGN
Filter reports to specific windows of interest
Deidentify and Preprocess
e.g. TNM, Stage, Metastasis
e.g. identifying temporal of two events
e.g. identifying chains relating tumor, mass and cancer
e.g. LocationOf
UMLS Entity and Relation (SHARP)
Clinical Coreference (ODIE)
Temporal Relations (THYME)
Cancer Specific Annotations (DeepPhe)
BRCA
OVCA
MEL
<Ann/>
<Ann/>
<Ann/>
ML/Annotators Evaluation

DeepPhe Information Model



Data Flow in DeepPhe



Initial Results

4 breast cancer patients ; 48 documents with gold annotations

14 mentions of TNM

"…. T STAGE, PATHOLOGIC: pT2; N STAGE, PATHOLOGIC: pN0; M STAGE: Not applicable…."

"… a clinical stage IIIA (T3 N2 Mx) triple negative infiltrating…."

6 mentions of Stage

"…The patient has stage 4 breast carcinoma…"

"…-Sister: Breast cancer (Stage I)

39 mentions of Receptors

"…. ER: Positive 270; PR: Positive 23…"

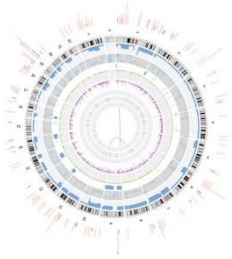
"… triple negative" breast carcinoma…"

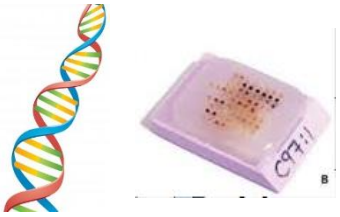
"… REPORTED TO BE NEGATIVE FOR ER, PR, AND HER-2/NEU…"

Type	*Precision*	*Recall*	*F1-score*
TNM	0.87	0.93	0.90
Stage	1	0.83	0.91
Receptors	1	0.90	0.95

How we share data and biospecimens today

Institution A





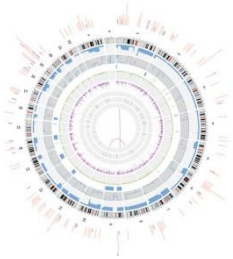
IRB OOR OGC

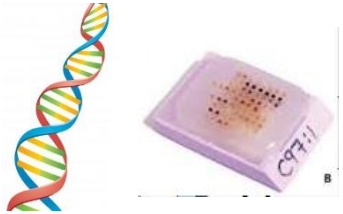
DUA
(specific to dataset)

MTA
(specific to specimen set)

OGC OOR IRB

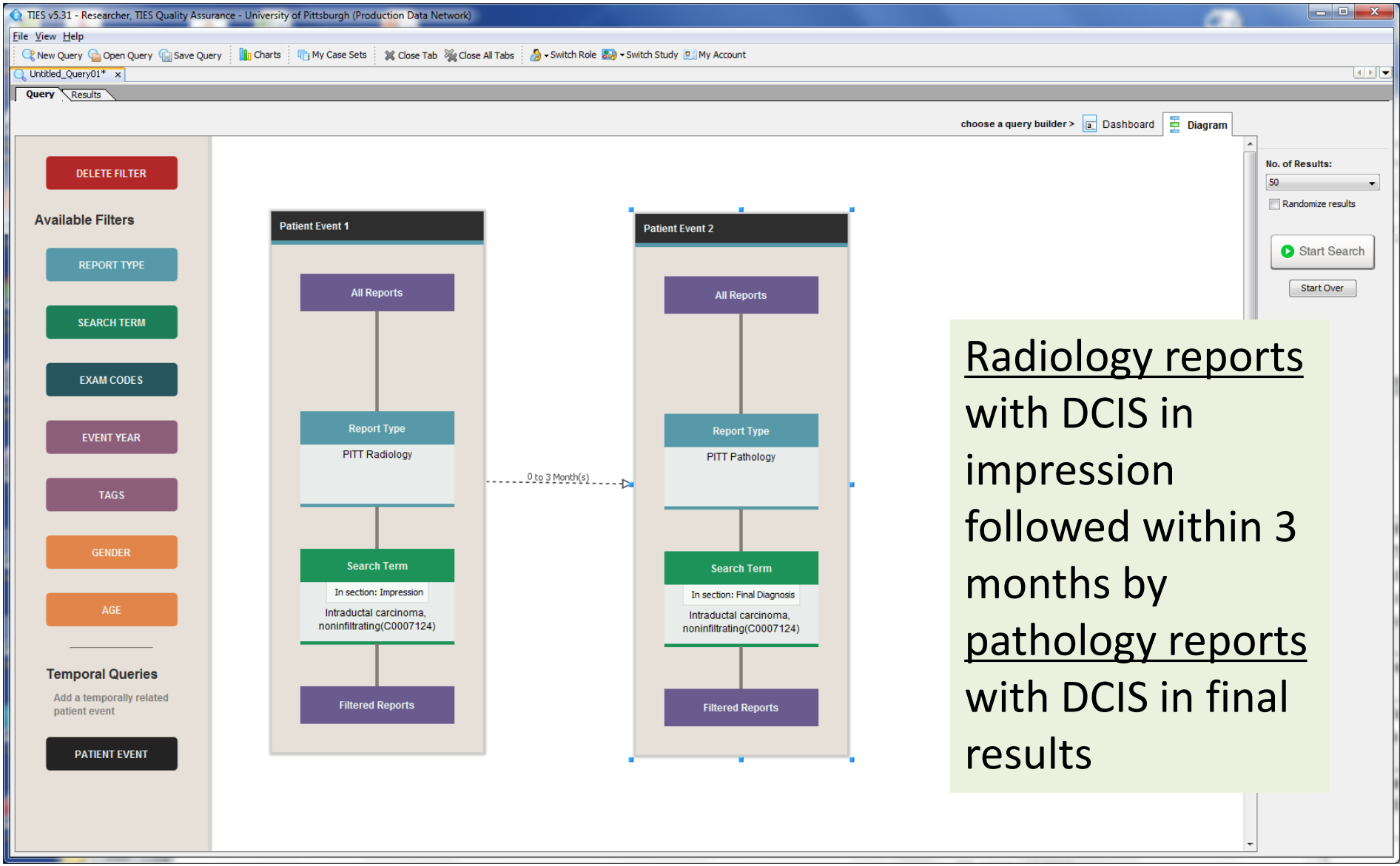
Institution B



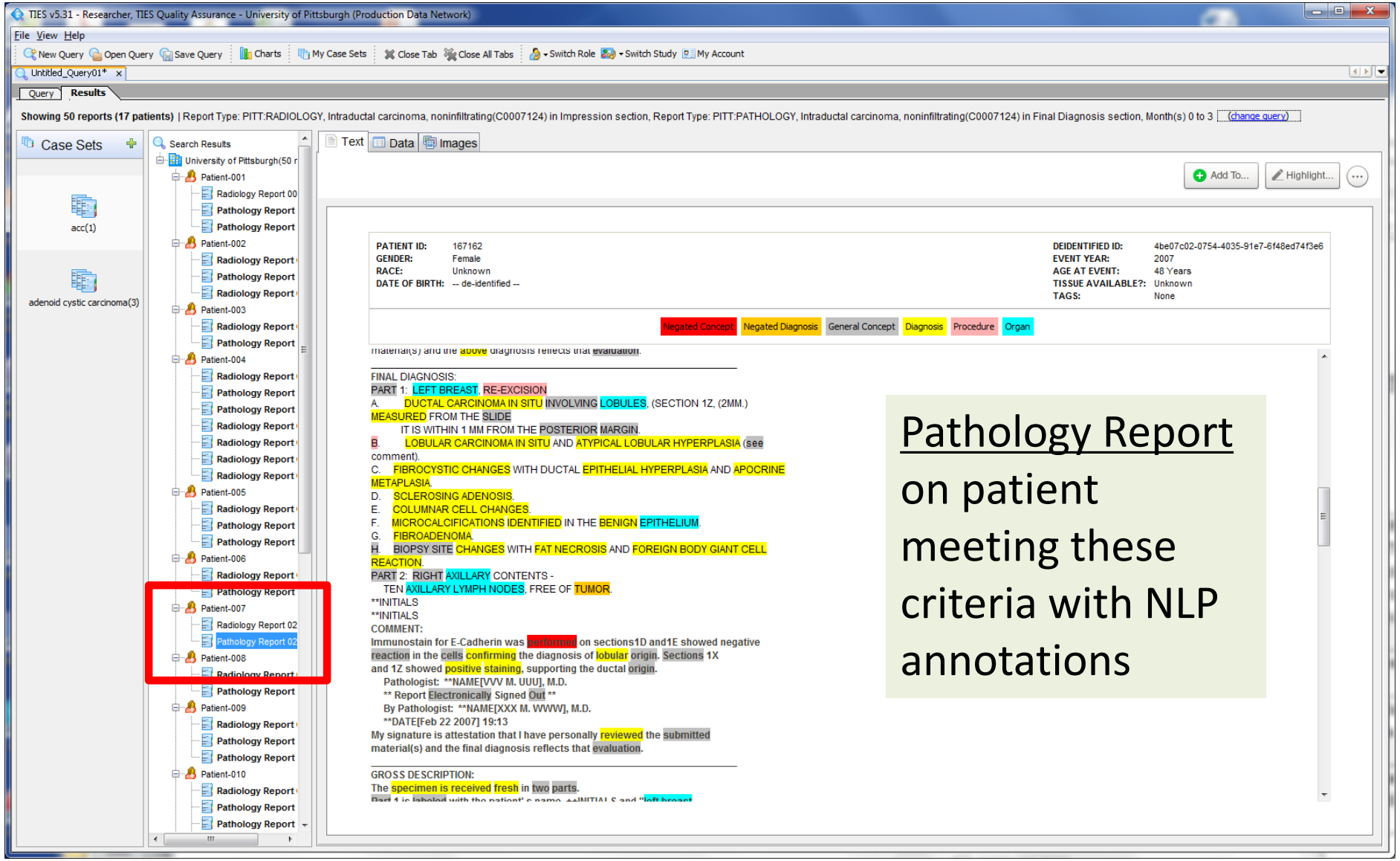


What is TIES? http://ties.pitt.edu

- An **NLP pipeline** for de-identifying, annotating and storing clinical documents
- A system for **indexing research resources** (FFPE, FF, images) with document annotations
- A system **for querying** large repository of annotated clinical documents and obtaining resources locally, using an honest broker model
- An open source platform to support **federated data and biospecimen sharing** among networks of cancer centers and other institutions



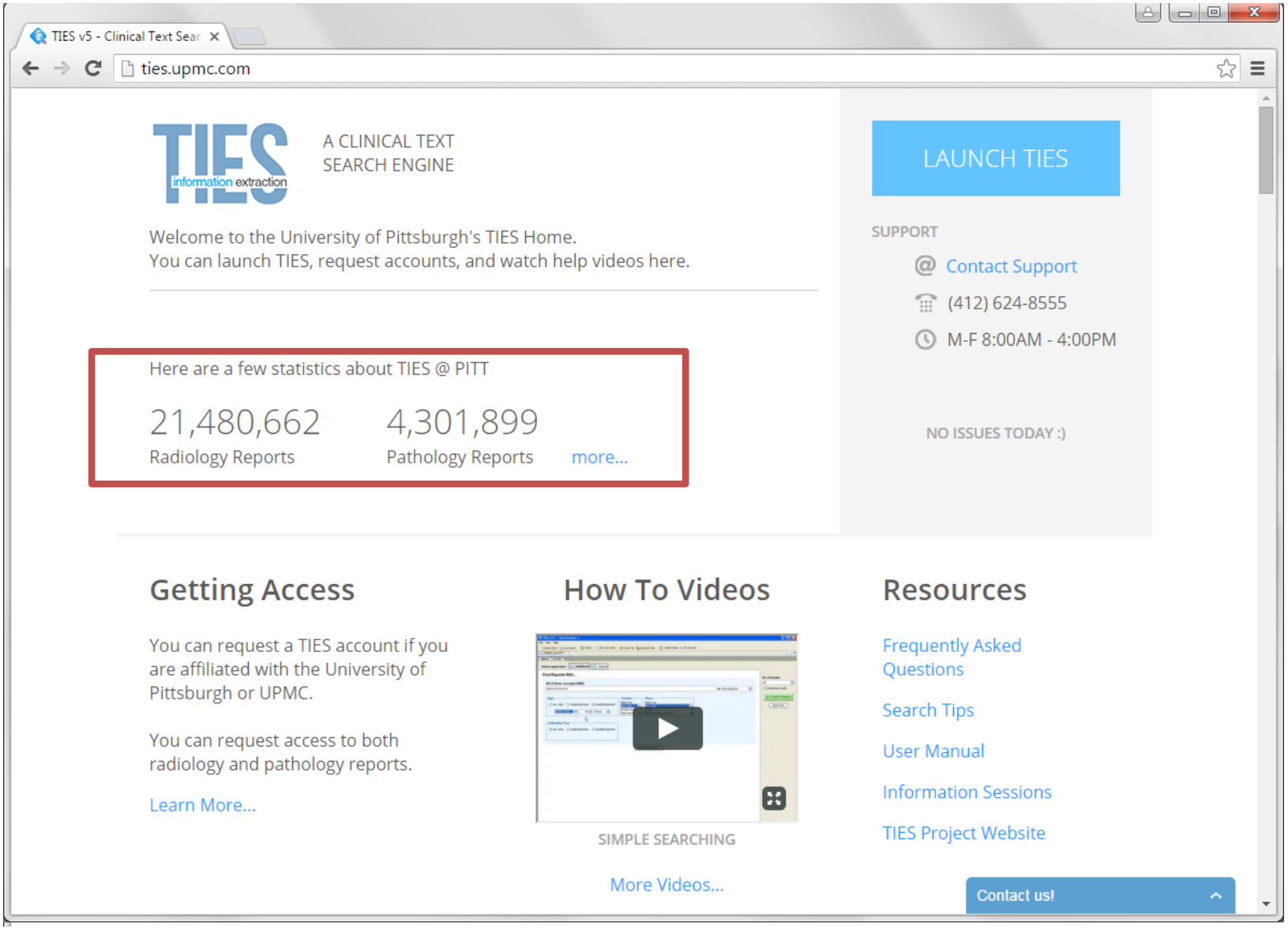
Radiology reports with DCIS in impression followed within 3 months by pathology reports with DCIS in final results


TIES v5.31 - Researcher, TIES Quality Assurance - University of Pittsburgh (Production Data Network)
File View Help
New Query Open Query Save Query Charts My Case Sets Close Tab Close All Tabs Switch Role Switch Study My Account
Untitled_Query01*
Query Results
Showing 50 reports (17 patients) | Report Type: PITT:RADIOLOGY, Intraductal carcinoma, noninfiltrating(C0007124) in Impression section, Report Type: PITT:PATHOLOGY, Intraductal carcinoma, noninfiltrating(C0007124) in Final Diagnosis section, Month(s) 0 to 3 (change query)
Case Sets
acc(1)
adenoid cystic carcinoma(3)
Search Results
University of Pittsburgh(50 r
Patient-001
Radiology Report 00
Pathology Report
Pathology Report
Patient-002
Radiology Report
Pathology Report
Radiology Report
Patient-003
Radiology Report
Pathology Report
Patient-004
Radiology Report
Pathology Report
Pathology Report
Radiology Report
Radiology Report
Radiology Report
Patient-005
Radiology Report
Pathology Report
Pathology Report
Patient-006
Radiology Report
Pathology Report
Patient-007
Radiology Report 02
Pathology Report 02
Patient-008
Radiology Report
Pathology Report
Patient-009
Radiology Report
Pathology Report
Pathology Report
Patient-010
Radiology Report
Pathology Report
Pathology Report
Text Data Images
Add To... Highlight...
PATIENT ID: 167162
GENDER: Female
RACE: Unknown
DATE OF BIRTH: -- de-identified --
DEIDENTIFIED ID: 4be07c02-0754-4035-91e7-6f48ed74f3e6
EVENT YEAR: 2007
AGE AT EVENT: 48 Years
TISSUE AVAILABLE?: Unknown
TAGS: None
Negated Concept Negated Diagnosis General Concept Diagnosis Procedure Organ
material(s) and the above diagnosis reflects that evaluation.
FINAL DIAGNOSIS:
PART 1: LEFT BREAST, RE-EXCISION
A. DUCTAL CARCINOMA IN SITU INVOLVING LOBULES, (SECTION 1Z, (2MM.) MEASURED FROM THE SLIDE
IT IS WITHIN 1 MM FROM THE POSTERIOR MARGIN.
B. LOBULAR CARCINOMA IN SITU AND ATYPICAL LOBULAR HYPERPLASIA (see comment).
C. FIBROCYSTIC CHANGES WITH DUCTAL EPITHELIAL HYPERPLASIA AND APOCRINE METAPLASIA.
D. SCLEROSING ADENOSIS.
E. COLUMNAR CELL CHANGES.
F. MICROCALCIFICATIONS IDENTIFIED IN THE BENIGN EPITHELIUM.
G. FIBROADENOMA.
H. BIOPSY SITE CHANGES WITH FAT NECROSIS AND FOREIGN BODY GIANT CELL REACTION.
PART 2: RIGHT AXILLARY CONTENTS -
TEN AXILLARY LYMPH NODES, FREE OF TUMOR.
**INITIALS
**INITIALS
COMMENT:
Immunostain for E-Cadherin was performed on sections1D and1E showed negative reaction in the cells confirming the diagnosis of lobular origin. Sections 1X and 1Z showed positive staining, supporting the ductal origin.
Pathologist: **NAME[VVV M. UUU], M.D.
** Report Electronically Signed Out **
By Pathologist: **NAME[XXX M. WWW], M.D.
**DATE[Feb 22 2007] 19:13
My signature is attestation that I have personally reviewed the submitted material(s) and the final diagnosis reflects that evaluation.
GROSS DESCRIPTION:
The specimen is received fresh in two parts.
Part 1 is labeled with the patient's name, **INITIALS and "left breast
Pathology Report on patient meeting these criteria with NLP annotations

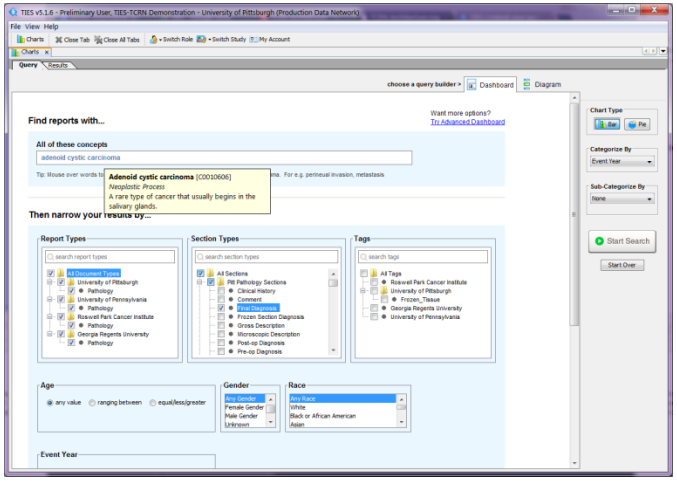
caTIES: a grid based system for coding and retrieval of surgical pathology reports and tissue specimens in support of translational research

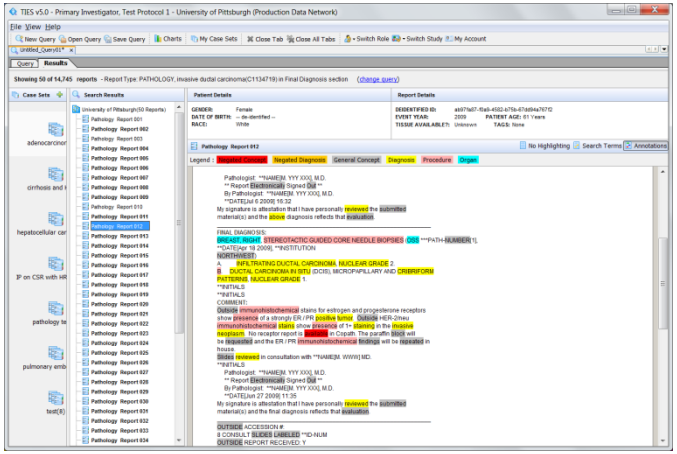
Rebecca S Crowley,[1,2,3] Melissa Castine,[1] Kevin Mitchell,[1] Girish Chavan,[1] Tara McSherry,[4] Michael Feldman[4]

#	Complexity	Query	Response time over three retrievals					Performance metrics					
			Number Reports Retrieved	*Mean time to first results (sec)*	*SD*	*Mean time to all results (sec)*	*SD*	*Number of Reports or Report Sets (complex) Classified*	*Agreement*	*TP*	*FP*	*Precision*	
1	Low	Men, 60-80 with prostatic adenocarcinoma on prostatectomy	1792	1.08	0.62	4.63	1.92	50	0.98	49	1	0.98	
2	Low	Women, 30-50 with atypical endometrial hyperplasia	792	0.70	0.19	0.70	0.19	33	1.00	33	0	1.00	
3	Low	Patients, 20-50 with phaeochromocytoma	54	0.95	0.31	0.96	0.31	50	0.98	49	1	0.98	
4	Low	Patients with hemangiosarcoma of scalp	17	0.49	0.13	0.49	0.13	17	1.00	17	0	1.00	
5	Low	Patients 10-30, with cystosarcoma phylloides	18	0.59	0.07	0.59	0.07	18	0.94	16	2	0.89	
6	Low	Patients with superficial spreading melanoma, metastatic	5	0.46	0.08	0.46	0.08	5	1.00	5	0	1.00	
7	Low	Patients with medullary carcinoma in thyroid gland	27	0.59	0.26	0.60	0.26	27	0.96	26	1	0.96	
8	Low	Patients with adenocarcinoma in brain	156	0.65	0.33	0.89	0.44	50	1.00	50	0	1.00	
9	Low	Men with invasive ductal carcinoma of breast	29	0.53	0.15	0.53	0.15	29	1.00	29	0	1.00	
10	Low	Patients, >60 with Hodgkins disease	549	0.64	0.17	0.84	0.22	50	0.94	34	16	0.68	
		All Low Complexity Queries	**3439**	**0.67**	**0.20**	**1.07**	**1.26**	**329**	**0.98**	**308**	**21**	**0.94**	

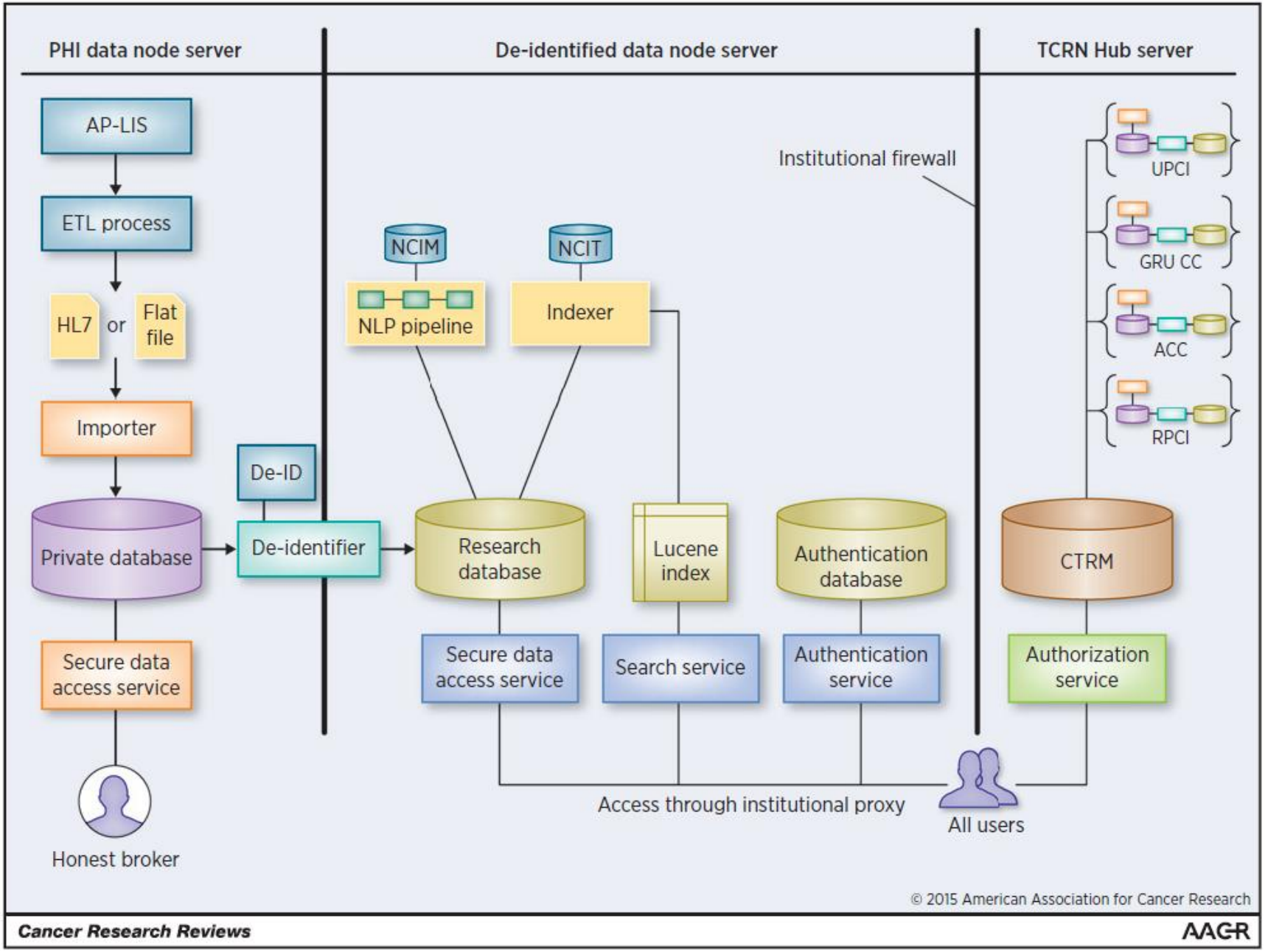

TIES v5 - Clinical Text Sear
ties.upmc.com
TIES
information extraction
A CLINICAL TEXT
SEARCH ENGINE
Welcome to the University of Pittsburgh's TIES Home.
You can launch TIES, request accounts, and watch help videos here.
LAUNCH TIES
SUPPORT
Contact Support
(412) 624-8555
M-F 8:00AM - 4:00PM
NO ISSUES TODAY :)
Here are a few statistics about TIES @ PITT
21,480,662
Radiology Reports
4,301,899
Pathology Reports
more...
Getting Access
You can request a TIES account if you are affiliated with the University of Pittsburgh or UPMC.
You can request access to both radiology and pathology reports.
Learn More...
How To Videos
SIMPLE SEARCHING
More Videos...
Resources
Frequently Asked Questions
Search Tips
User Manual
Information Sessions
TIES Project Website
Contact us!

TIES Functionality



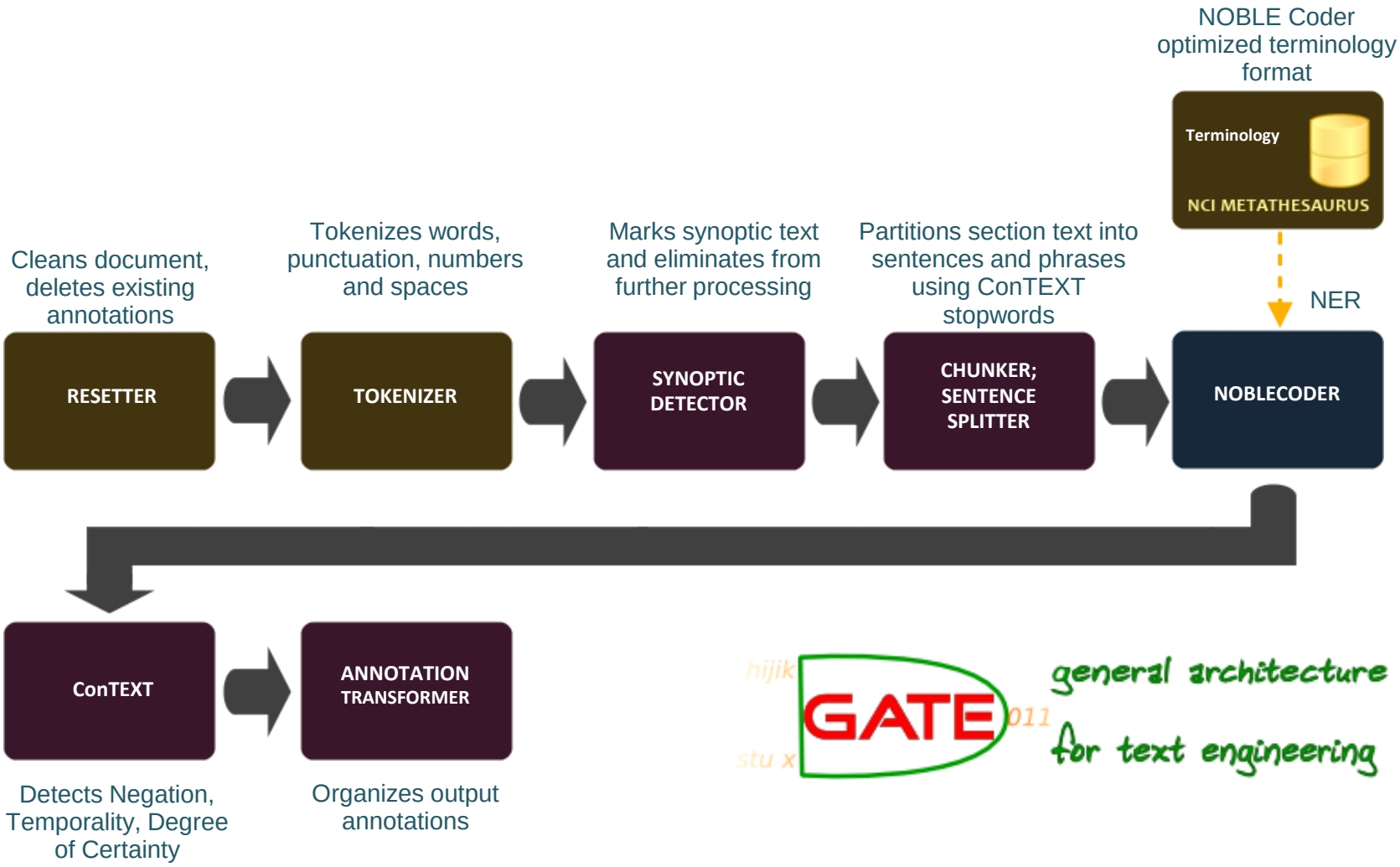


- NLP; Concept annotation with NCIM; ontology indexing with NCIT using Lucene
- Infrastructure to code and recode; parallelize coders
- De-identification, encryption, separation of PHI, auditing, X.509, quarantining
- Honest Broker model built in to software. HBs see identifiers when working with investigator
- Workflow to request FFPE, Frozen Tissue, Radiology Images
- Virtual Slides
- Other datatypes (e.g. Cancer Registry data)



Jacobson et al, Cancer Research 2015

Current TIES NLP Pipeline



BMC Medical Informatics and Decision Making

BioMed Central

Research article

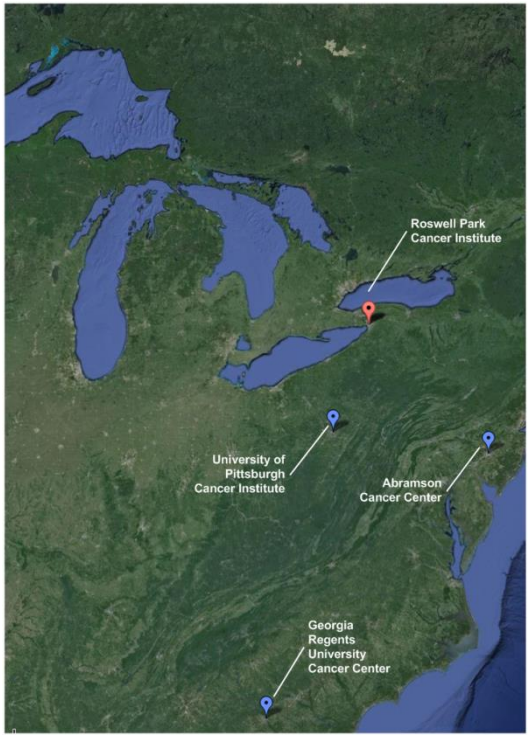
Open Access

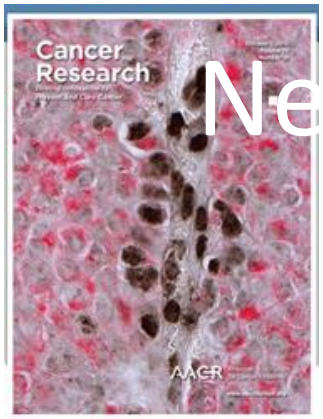
Security and privacy requirements for a multi-institutional cancer research data grid: an interview-based study

Frank J Manion*[1], Robert J Robbins[2], William A Weems[3] and Rebecca S Crowley[4]

U Pittsburgh Cancer Institute
Abramson Cancer Center (Penn)
Roswell Park Cancer Institute
GRU Cancer Center
……And others are preparing to join

http://ties.pitt.edu/tcrn





Ne

Resource

Cancer Research

A Federated Network for Translational Cancer Research Using Clinical Data and Biospecimens

Rebecca S. Jacobson[1], Michael J. Becich[1], Roni J. Bollag[2], Girish Chavan[1], Julia Corrigan[1], Rajiv Dhir[1], Michael D. Feldman[3], Carmelo Gaudioso[4], Elizabeth Legowski[1], Nita J. Maihle[2], Kevin Mitchell[1], Monica Murphy[4], Mayurapriyan Sakthivel[4], Eugene Tseytlin[1], and JoEllen Weaver[3]

Network Trust Agreements

- Instrument of Adherence
- IRBs agree that use of data for investigators is NHSR, no additional IRB protocol even for record level de-id data
- Establishes governing body

Policies and Processes

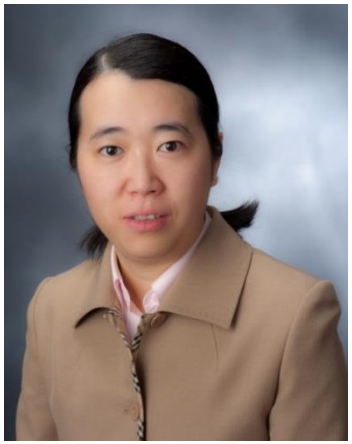
- QA and validation
- User authorization
- Auditing
- Incident Reporting
- Joining of new members
- Governance

Table 2. TCRN case statistics for numbers of patients and cases (A) and the number of cases of rare tumors (B) and common cancer categories (C) based on final diagnosis

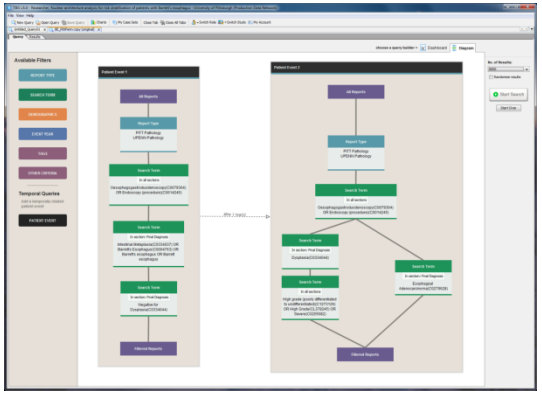
	GRU	RPCI	ACC	UPCI	Total
A. Case statistics					
Patients	76,404	72,376	465,717	1,840,156	2,454,653
Pathology cases	157,316	156,555	857,681	4,588,017	5,759,569
B. Rare tumors					
Adenoid cystic carcinoma	41	88	404	509	1,042
Adrenocortical carcinoma	5	20	59	63	147
Alveolar soft part sarcoma	3	15	10	25	53
Angioimmunoblastic lymphadenopathy	12	35	58	84	189
Chordoma	5	14	124	245	388
Follicular dendritic cell sarcoma	2	2	8	13	25
Merkel cell carcinoma	9	72	165	196	442
Ovarian granulosa cell tumor	4	10	23	34	71
Phaeochromocytoma	15	38	272	164	489
Pleomorphic xanthoastrocytoma	2	5	12	53	72
Pseudomyxoma peritonei	6	36	46	129	217
Rhabdomyosarcoma	34	70	86	270	460
Sebaceous adenocarcinoma	13	33	26	94	166
Sinonasal undifferentiated carcinoma	2	6	31	27	66
Thymoma	13	45	433	210	701
C. Common cancer categories					
Bladder carcinoma	345	1,618	3,873	6,711	12,547
Breast carcinoma	1,143	9,605	28,262	37,691	76,701
Colorectal carcinoma	465	2,530	6,898	11,608	21,501
Endometrial carcinoma	394	1,815	3,707	7,706	13,622
Esophageal carcinoma	63	1,477	2,452	3,514	7,506
Hepatic carcinoma	153	633	2,912	5,720	9,418
Lung carcinoma	820	4,264	10,208	17,955	33,247
Lymphoma	1,387	6,795	10,605	15,689	34,476
Malignant glial neoplasm	242	292	2,198	4,943	7,675
Malignant melanoma	335	2,675	5,180	7,068	15,258
Ovarian carcinoma	503	2,872	4,659	6,446	14,480
Pancreatic carcinoma	162	740	1,866	3,622	6,390
Prostate carcinoma	903	3,612	18,867	19,445	42,827
Renal cell carcinoma	364	1,319	3,183	10,950	15,816
Thyroid carcinoma	474	1,236	7,681	12,387	21,778

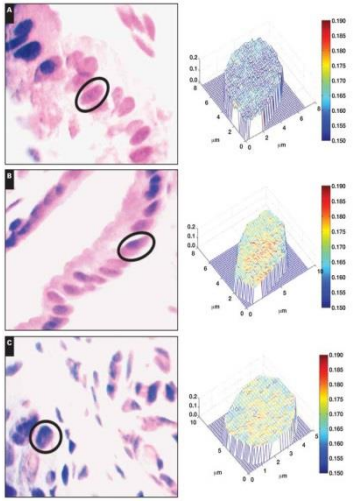
Jacobson et al, Cancer Research 2015

Example of TCRN Pilot Project



UPCI Investigator
Yang Liu, PhD





Published OnlineFirst September 17, 2015; DOI: 10.1158/0008-5472.CAN-15-1274

Integrated Systems and Technologies

Cancer Research

Early Prediction of Cancer Progression by Depth-Resolved Nanoscale Mapping of Nuclear Architecture from Unstained Tissue Specimens

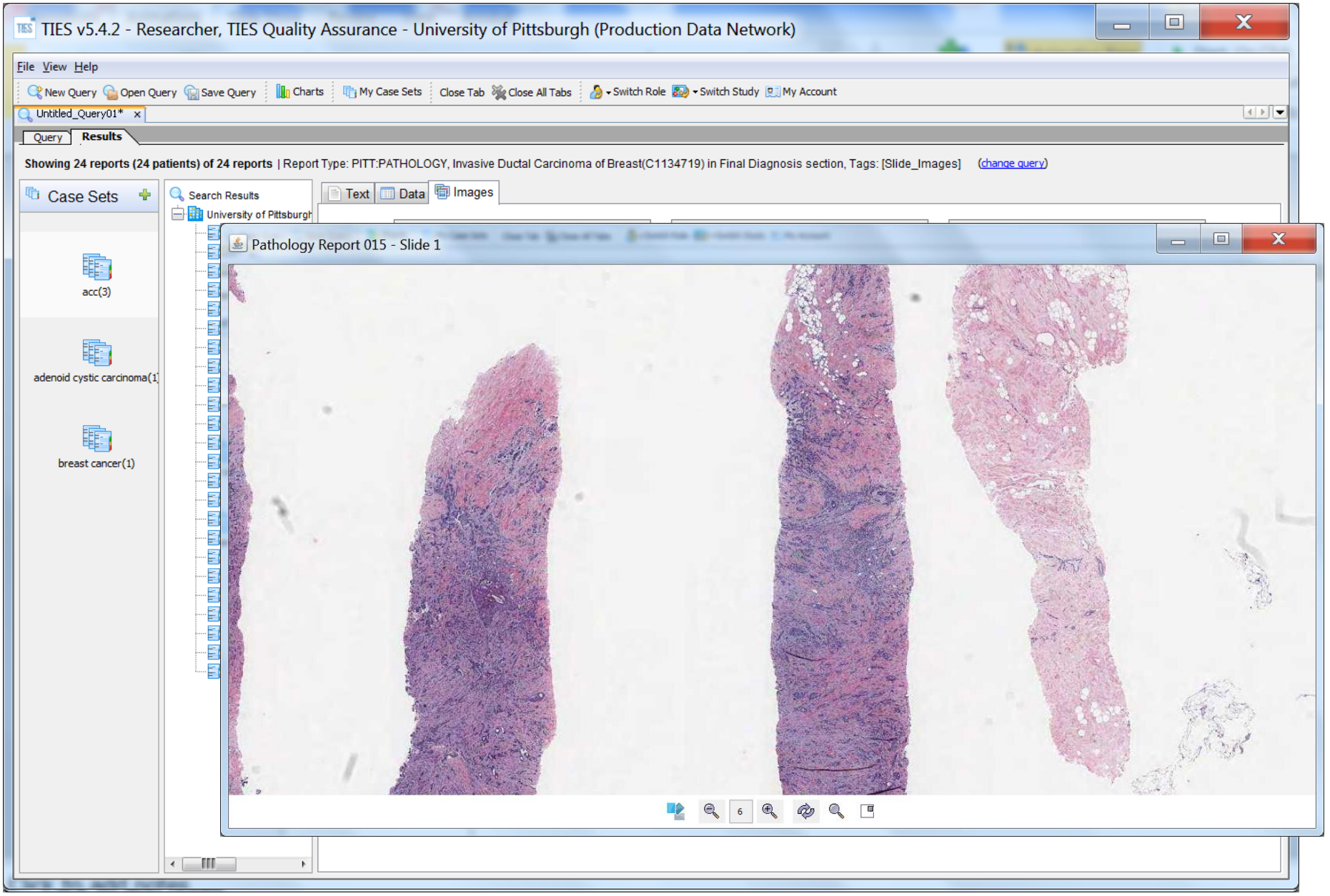
Shikhar Uttam[1], Hoa V. Pham[1], Justin LaFace[1], Brian Leibowitz[2,3], Jian Yu[2,3], Randall E. Brand[4], Douglas J. Hartman[2], and Yang Liu[1,3,4]

Abstract

Early cancer detection currently relies on screening the entire at-risk population, as with colonoscopy and mammography. Therefore, frequent, invasive surveillance of patients at risk for developing cancer carries financial, physical, and emotional burdens because clinicians lack tools to accurately predict which patients will actually progress into malignancy. Here, we present a new method to predict cancer progression risk via nanoscale nuclear architecture mapping (nanoNAM) of unstained tissue sections based on the intrinsic density alteration of nuclear structure rather than the amount of stain uptake. We demonstrate that nanoNAM detects a gradual increase in the density alteration of nuclear architecture during malignant transformation in animal models of colon carcinogenesis and in human patients with ulcerative colitis, even in tissue that appears histologically normal according to pathologists. We evaluated the ability of nanoNAM to predict "future" cancer progression in patients with ulcerative colitis who did and did not develop colon cancer up to 13 years after their initial colonoscopy. NanoNAM of the initial biopsies correctly classified 12 of 15 patients who eventually developed colon cancer and 15 of 18 who did not, with an overall accuracy of 85%. Taken together, our findings demonstrate great potential for nanoNAM in predicting cancer progression risk and suggest that further validation in a multicenter study with larger cohorts may eventually advance this method to become a routine clinical test. *Cancer Res; 75(22); 4718–27. ©2015 AACR.*

Doubled
Study N using TCRN

UPMC: 46
Penn: 44

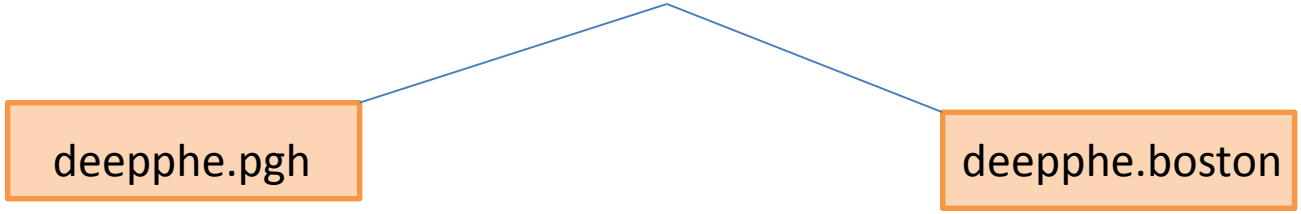

TIES v5.4.2 - Researcher, TIES Quality Assurance - University of Pittsburgh (Production Data Network)
File View Help
New Query Open Query Save Query Charts My Case Sets Close Tab Close All Tabs Switch Role Switch Study My Account
Untitled_Query01*
Query Results
Showing 24 reports (24 patients) of 24 reports | Report Type: PITT:PATHOLOGY, Invasive Ductal Carcinoma of Breast(C1134719) in Final Diagnosis section, Tags: [Slide_Images] (change query)
Case Sets
acc(3)
adenoid cystic carcinoma(1)
breast cancer(1)
Search Results
University of Pittsburgh
Text Data Images
Pathology Report 015 - Slide 1
6

The Future of Cancer Phenotyping

- Deep phenotyping, high throughput phenotyping
- Representing and extracting phentotypes
- Combining text and structured data more effectively
- Sharing of unstructured data, phenotypes, images, even biospecimens across federated networks

- Large scale centralized genotype-phenotype knowledge bases
- Local integration with clinically derived data and data warehouses
- Rapid development and sharing of validated phenotypes across networks
- National data and biospecimen sharing networks

DeepPhe



Harry Hochheiser
Girish Chavan
Eugene Tseytlin
Kevin Mitchell
Melissa Castine
Mike Davis

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

Funding

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

TIES and the TIES Cancer Research Network

TIES Team

Girish Chavan
Eugene Tseytlin
Kevin Mitchell
Julia Corrigan
Liz Legowski

Adi Nemlekar
Yining Zhao
Vanessa Benkovich
Liron Pantanowitz
Rajiv Dhir

<u>Roswell Park</u>
Carmelo Gaudioso
Monica Murphy
Mayurapriyan Sakthivel
Amanda Rundell

<u>GRU</u>
Roni Bollag
Samir Khleif
Jennifer Carrick
Nita Maihle
And more…..

<u>Penn</u>
Michael Feldman
Nate DiGiorgio
Tara McSherry
Joellen Weaver

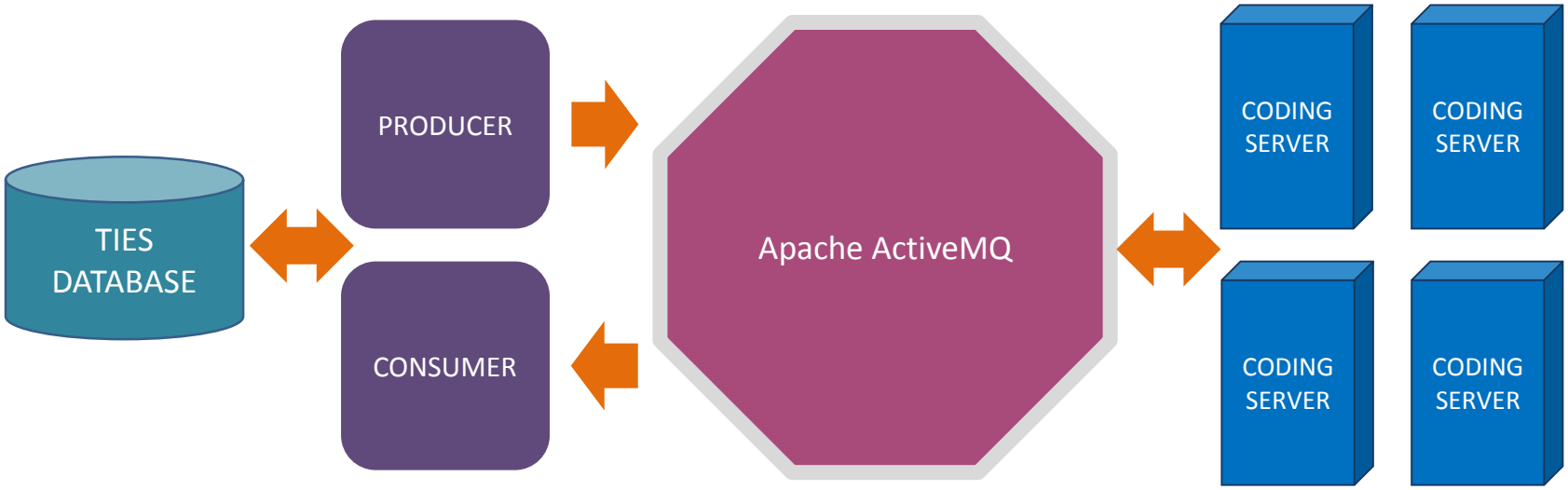
Funding

NCI U24 CA180921 Enhanced Development of TIES

Thank You

High throughput coding using Java Messaging Service(JMS)

- Each TIES coding service can be configured to run multiple processes internally to utilize multi-core CPUs effectively
- Additionally, TIES can use Java Messaging to utilize multiple servers for coding a large dataset. This reduces the load on the database server by using a JMS provider like ActiveMQ to act as intermediary

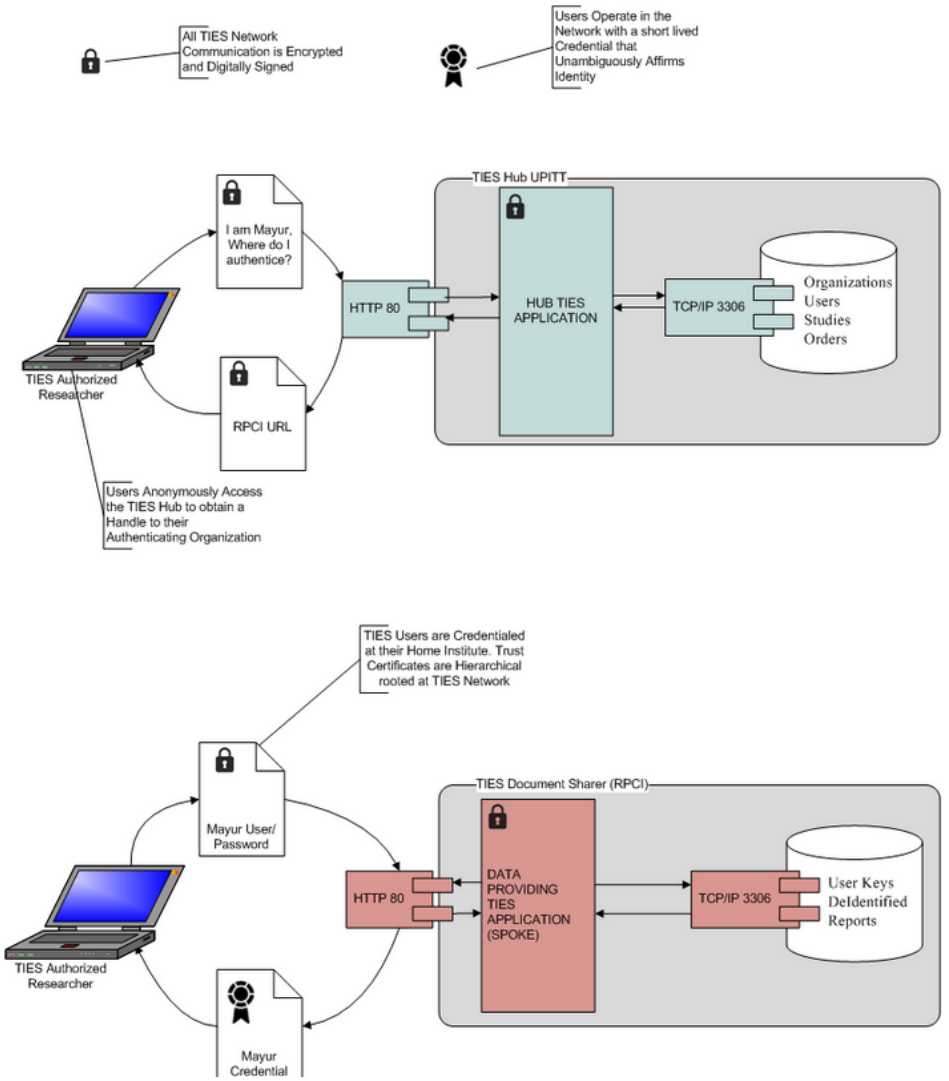


Multi-layered approach to data security

- TIES separates the PHI and de-identified data into separate databases that can be hosted on different servers for additional protection
- OGSA-DAI grid services encrypt all communication between the client and servers using RSA-1024 encryption
- Role based access control allows for data access granularity at three different levels
- Users can quarantine any reports containing PHI, which immediately hides that report from all users until an QA admin reviews it
- All queries and document views are logged by user and study. Auditing view lets you easily retrieve past activity for auditing purposes

Authentication and Authorization

- Authentication happens at user's institution
- Authorization happens at Hub server for the network
- After successful authentication, X.509 proxy certificates with a 12 hour validity are generated and used to communicate with any nodes in the network
- Services are further secured using gridmaps that only allow specific individuals to access them



Structured Data Support