# Clinical and research data integration: the i2b2 FSM experience

Laboratory of Biomedical Informatics for Clinical Research Fondazione Salvatore Maugeri - FSM - Hospital, Pavia, italy

Laboratory of Biomedical Informatics "Mario Stefanelli" University of Pavia, Pavia, Italy

Daniele Segagni, MS, Biomedical Engineer daniele.segagni@fsm.it



## **FSM** institutes

#### 18 FSM institutes in Italy

3 i2b2 project:

- Pavia: Oncology data & Cardiology data
- Puglia: Administrative data
- Sources and CRC in Pavia



## **I2b2** instances

#### Oncology

Active since	Patients	Visits	Observations	Genetic data	NLP	Plug- ins	Cells
2010	28.838	142.464	2.341.771	Y	у	6	1

#### Sub-project: biobank

 Active since	Patients	Visits	Observations	Genetic data	NLP	Plug- ins	Cells
2011	786	-	6.855	-	У	1	0

#### Cardiology

Active since	Patients	Visits	Observations	Genetic	NLP	Plug- ins	Cells
2009	6.334	15.094	205.418	у	n	5	1

#### **Adminstration**

Active since	Patients	Visits	Observations	Genetic	NLP	Plug- ins	Cells
2011	5.611	7.726	23.175	n	n	2	0

## I2b2 development



Program Visualization Tools Help

Waikato Environment for Knowledge Analysis

۰ 🕈

Version 3.6.2

(c) 1999 - 2010

The University of Waikato

Hamilton, New Zealand

#### 7 web plug-in



Weka GUI Chooser

of Waikato

#### 2 new cells in development

Applications

Explorer

Experimenter KnowledgeFlow

Simple CLI

(~)

Clinical Measurement Chart i2b2 plug-in Chart maker for clinical measurement

#### ExportXLS





Retrive biobank samples information for a defined patient-set

GATE



Monitor the amount of concepts and/or modifier in your data warehouse

general architecture

for text engineering

FSM - iNaviGraphMaker View similar patient using this graph maker plug-in

FSM - Mutation Monitor Mutation monitor M&M

FSM - Health Activity Tracker Graphical analysis of FSM health activity

#### NHLBI Exome Sequencing Project (ESP)

Exome Variant Server

1 NLP module

Many ETL jobs





Pentaho Data Integration Previously Kettle

## i2b2 FSM data flow



## **ONCO-i2b2: System Architecture**

#### 1. Pathology Unit

2. Biobank "Bruno Boerci"



**Research Data** 

**Clinical Data** 

"-omics" data [in progress]

### Integration process: clinical data

#### 3. FSM HIS



Integration in the i2b2 data warehouse of medical concept related to the day by day clinical practice, for example:

- Anamnesis
- Therapies
- Diagnosis (ICD9)
- Procedures (ICD9)
- Haematochemical blood analysis

## Integration process: research data

2. Biobank "Bruno Boerci"

1. Pathology Unit



Natural language processing (NLP) identifies various concept types in the istological textual records that are associated with each patient for each medical record.

Automated tracking for samples coming from PU to biobank

## Integration process: research data

2. Biobank "Bruno Boerci"

1. Pathology Unit



#### **Biobank sample anonymization**

Once a patient is hospitalized he/she may sign an informed consent to donate samples, specimens and data collected for clinical reasons to research

#### **NLP** in action

#### Patology Medical Report In italian

Tipo esame : Istologico N° Esame : 11-I-03276 Data accettazione : 08/06/2011

Reparto di provenienza : U.O. Chirurgia Gen. Senologia

tissue type 🛛 🔪 biospecimen id

Materiale inviato A LINFONDDO SENTINELLA CAVO ASCELLARE DESTRO INVIATO IL 08 06 2011 PER ESAME ESTEMPORANEO B OUADRANTE SUPERIORE MAMMELLA DESTRA

#### Descrizione Macroscopica tissue description

 A) Due linfonodi rispettivamente di cm. 1.6 e cm. 1.1 esminati completamente con sezioni multiple seriate.
 B) Tessuto mammario di cm. 7 x 6 x 3, orientabile (peso gr. 73 dopo fissazione). Al taglio, prossima al margine infero esterno (distanza minima cm. 1.2), neoplasia a margini infiltrativi di cm. 1.

#### Diagnosi intraoperatoria diagnosis

A) Due linfonodi indenni da metastasi. Reperti e Conclusioni B) Carcinoma duttale infiltrante della mammella, G2 SBR, a crescita di tipo infiltrativo, con scarsa reazione linfoplasmacellulare e minima componente (10%) peritumorale di carcinoma intraduttale di grado nucleare intermedio ed alto (DIN 2-DIN 3), prevalentemente di tipo solido e cribroso con focale necrosi.

#### Valutazione assetto recettoriale, frazione proliferante e c-erbB2:

Recettori estrogeni: 90% Recettori progesterone: 70% KI67: 15% c-erb B2:punteggio DAKO Hercept]

hormone receptors values

c-erb B2:punteggio DAKO HerceptTest: 3\* (positivo); intensa e completa colorazione delle membrane citoplasmatiche nell'80% delle cellule neoplastiche.



#### NLP output - XML format

<?xml version="1.0" encoding="UTF-8"?>

cdocument>

## NLP Module

<ul> <li>Gocumente</li> </ul>
<date>08/06/2011</date>
<type>lstologico</type>
<number>11-I-03276</number>
<analysis></analysis>
<estrogens receptors="">90%</estrogens>
<progesterone_receptors>70%</progesterone_receptors>
<ki67>15%</ki67>
<c-erb b2="">3+</c-erb>
<grade>G 2</grade>
<state>pT1bÅ-pN0(sn)Å-pM</state>
<snomed codes=""></snomed>
<snomed consistent="true"></snomed>
<code1>T-D8100</code1>
<code2>M-09410</code2>
<nome1>Axilla structure (body structure)</nome1>
<nome2>No evidence of neoplasm (finding)</nome2>
<snomed consistent="true"></snomed>
<code1>T-04000</code1>
<code2>M-85003</code2>
<name1>Breast structure (body structure)</name1>
<name2>Infiltrating duct carcinoma (morphologic abnormality)</name2>
[]
<snomed consistent="true"></snomed>
<code1>T-04020</code1>
<code2>M-09410</code2>
<name1>Right breast structure (body structure)</name1>
<name2>No evidence of neoplasm (finding)</name2>
<snomed consistent="true"></snomed>
<code1>T-04000</code1>
<code2>M-09410</code2>
<name1>Breast structure (body structure)</name1>
<name2>No evidence of neoplasm (finding)</name2>

#### Integration process: -omics data

#### 5.-OMICS DATA



Data coming from High-Throughput Screening (HTS) experiments

The results of these experiments provide high amount of data that will be analyzed with the purpose of verifying the existence of genetic mutations Ion Proton<sup>™</sup> Sequencer

Future developments: I2b2-noSQL database integration





Clinical Measurement Chart i2b2 plug-in Chart maker for clinical measurement

- Shows trend for clinical concepts and modifiers (ecg, holter values or therapy dose) for each patient present in the patient set
- Used in oncology and cardiology





#### ExportXLS

This plugin tabulates unidentified patient data, and applicable diagnoses



• Export patient data in Excel Format V3

This is the home of the Related Project - ExportXLS space.

- Project developed by Mauro Bucalo, University of Pavia and Wayne Chan from the BMI-Core, University of Massachusetts Medical School, Worcester
- https://community.i2b2.org/wiki/display/ExportXLS

E	kpor	tXLS												
9	peci	fy Data Vie	w Results	Plugin Hel	р									
F		·	L						_					
(	Click	the [Excel Exp	oort] button to	o download	the fo	llowin	g table into	an Ex	cel file.	xcel Export				
Γ									Pati	ont Informati	on			
					for 1	Patie	nt Set ' <i>Tr</i> .	ipli	TX - Pr@16	:22:40 [10-17	-2012] [dani	ele] [P)	ATIEN1	SET_981]'
		PATIENT_ID	Vital Status	Birth Year	Sex	Age	Language	Race	Religion	Marital Status	State\City	Income	Age	Procedures
	1	284	N	1965	F	40	0	0	0	0			40	
	2	705	N	1952	F	59	0	0	0	0			59	ALTRA TERAPIA FISICA
	3	3637	N	1963	F	43	0	0	0	0			43	Altra riparazione o ricostruzione del capezzolo
	4	8470	N	1974	F	32	0	0	0	0			31	Impianto di protesi monolaterale
	5	9684	N	1955	F	52	0	0	0	0			12	Altro innesto di cute su altre sedi
	6	10246	N	1951	F	60	0	0	0	0			2	Quadrantectomia della mammella
	7	14103	N	1975	F	36	0	0	0	0			36	ALTRA TERAPIA FISICA
	8	16680	N	1959	F	52	0	0	0	0			52	IMPIANTO DI PROTESI MONOLATERALE
	9	16752	N	1947	F	62	0	0	0	0			62	Visita generale
	10	17214	¥	1953	F	57	0	0	0	0			57	INIEZIONE O INFUSIONE DI SOSTANZE CHEMIOTERAPICHE PER TUMORE
	11	17925	N	1957	F	54	0	0	0	0			54	Quadrantectomia della mammella
	12	19128	N	1952	F	59	0	0	0	0			59	
	13	27664	N	1954	F	56	0	0	0	0			56	INIEZIONE O INFUSIONE DI SOSTANZE CHEMIOTERAPICHE PER TUMORE
	14	28268	N	1959	F	51	0	0	0	0			51	INIEZIONE O INFUSIONE DI SOSTANZE CHEMIOTERAPICHE PER TUMORE



#### FSM - Biobank Info

Retrive biobank samples information for a defined patient-set

- Retrive bio-speciments information (tube position, tissue desription, blood aliquots) for a selected patient set
- Used in oncology

obank Info									
- Tipolog	gia: Sangue 🛛 —								
👻 P	aziente: 17012	Totale: 2 campi	oni						
	Barcode	Posizione	Box	Vassoio	Rack	Cassetto	Frigo		
	694_1_1	21	Sangue Mammella 5	С	Sangue-Plasma Senologia	CASSETTO2	Frigo 1		
	694_1_2	22	Sangue Mammella 5	С	Sangue-Plasma Senologia	CASSETTO2	Frigo 1		
						Seleziona t	utte le voo	si: 🔲	
Tingle									
- Tipolog	gia: Tessuto Tumo	rale							
- Tipolog P	gia: Tessuto Tumo 'aziente: 17012	rale	oni						
- Tipolog - P	gia: Tessuto Tumo <b>'aziente: 17012</b>	rale Totale: 3 campi	oni	Vassain	Dack	Corrotto	Frigs		
- Tipolog - P	gia: Tessuto Tumo aziente: 17012 Barcode	rale Totale: 3 campi Posizione	oni Box	Vassoio	Rack	Cassetto	Frigo		
- Tipolog	gia: Tessuto Tumo aziente: 17012 Barcode 694_3_M_1	rale Totale: 3 campi Posizione 31	oni Box Mammella 6	Vassoio B	Rack Tessuti Senologia	Cassetto CASSETTO2	Frigo Frigo 1		
- Tipolog	gia: Tessuto Tumo aziente: 17012 Barcode 694_3_M_1 694_3_M_2	rale Totale: 3 campi Posizione 31 32	oni Box Mammella 6 Mammella 6	Vassoio B B	Rack Tessuti Senologia Tessuti Senologia	Cassetto CASSETTO2 CASSETTO2	Frigo Frigo 1 Frigo 1		
- Tipolog	gia: Tessuto Tumo aziente: 17012 Barcode 694_3_M_1 694_3_M_2 694_3_M_3	rale Totale: 3 campi Posizione 31 32 33	oni Box Mammella 6 Mammella 6 Mammella 6	Vassoio B B B B	Rack Tessuti Senologia Tessuti Senologia Tessuti Senologia	Cassetto CASSETTO2 CASSETTO2 CASSETTO2	Frigo Frigo 1 Frigo 1 Frigo 1 Frigo 1		
- Tipolog	gia: Tessuto Tumo aziente: 17012 Barcode 694_3_M_1 694_3_M_2 694_3_M_3	rale Totale: 3 campi Posizione 31 32 33	oni Box Mammella 6 Mammella 6 Mammella 6	Vassoio B B B	Rack Tessuti Senologia Tessuti Senologia Tessuti Senologia	Cassetto CASSETTO2 CASSETTO2 CASSETTO2 Seleziona t	Frigo Frigo 1 Frigo 1 Frigo 1 sutte le voo		
- Tipolog	gia: Tessuto Tumo l <mark>aziente: 17012</mark> Barcode 694_3_M_1 694_3_M_2 694_3_M_3	rale Totale: 3 campi Posizione 31 32 33	oni Box Mammella 6 Mammella 6 Mammella 6	Vassoio B B B	<b>Rack</b> Tessuti Senologia Tessuti Senologia Tessuti Senologia	Cassetto CASSETTO2 CASSETTO2 CASSETTO2 Seleziona t	Frigo Frigo 1 Frigo 1 Frigo 1 sutte le voo		

#### FSM - Concept Visual Monitor

Monitor the amount of concepts and/or modifier in your data warehouse

 Monitor the status of the i2b2 data warehouse by creating charts that represent the number of observations and the number of patients related to each concept or modifier selected

nserisci i dati BarChart PieChart Air	ito			
Concetto		Nume	ero Osservazioni	Numero Pazient
Clinicopathological Features\Histolog (CF_HR)	ical Report		626	468
Totale:			626	468
Concetto-Modifier	Numero Oss	servazioni	Numero Pazienti	
Histological Report\ER (HR_E)	61	3	468	
	61	2	468	
Histological Report\HER2 (HR_H)				
Histological Report\HER2 (HR_H) Histological Report\Ki67 (HR_K)	61	3	464	
Histological Report\HER2 (HR_H) Histological Report\Ki67 (HR_K) Histological Report\PgR (HR_P)	61	3	464 468	





#### FSM - iNaviGraphMaker

View similar patient using this graph maker plug-in

- Using a CBR procedure, researchers are able to enhance the patient selection with a process based on the medical concept space related to a patient set to identify a group of similar patients
- Analysis based on binary, nominal and continuous variables
  - Binary variables: presence/absence of diseases or signs/symptoms
  - Nominal and continuous variables: discrete/continuous values of clinical observations



- Concept: Concept Unique Identifier (CUI) in UMLS Metathesaurus
- Distance between cases : semantic similarity between concepts in the UMLS ontology



#### FSM - Mutation Monitor Mutation monitor M&M

- View the genetic information and search the available publications on Exome Variant Server <u>http://evs.gs.washington.edu/EVS/</u>
- Used in cardiology
- Still in development

Mutat	tionMonitor									
Inse	risci i dati Aiut	to								
Tras	Trascina un set di Paziente per individuarne le mutazioni e i corrispondenti effetti									
	Pazienti: B Qtc-St El-Pq@10:41:38 [10-19-2012] [daniele] [PATIEN]									
Gei	ne and Codine	g Effect					0			
	Gene	Coding Effect	First Aminoacid	Last Aminoacid	Position	Result				
	SCN5A	K1500 del								
	SCN5A	E764K	Glu	Lys	764					
	SCN5A	R1193Q	Arg	Gln	1193					
	SCN5A	R568H	Arg	His	568					
	SCN5A	L1222S.E.								
	SCN5A	S940N	Ser	Asn	940					
	SCN5A	F1293S	Phe	Ser	1293					
	SCN5A	R975W	Arg	Тгр	975					
	SCN5A	11 02 N	lle	Asn	102					
	SCN5A	E1385X		[A see ]						
	SCN5A	V1323G	Val	Gly	1323					
م ا	Trova 🌩 Aggic	orna 🛨 Go!	ान २न	Pagina 💶 di 1 🔛 🖬		Visualizzati 1 - 1	1 di 11			

FSM - Health Activity Tracker Graphical analysis of FSM health activity

- Developed to support the generation of reports that explain the performance of the hospital medical units
- Used only in administration

Plug	ginPatient									Р	luginPat	tient							
Sp	ecify Data	View Resu	ults Plugin Help							5	Specify D	Data V	iew Result	s Plugin I	Help				
Se	lezionare il rep	parto	L					-		5	Selezion	are il repa	arto						
C	ardiologia Rial	bilitazione	e 🌲	PluginPat	tient						Cardiol	ogia Riab	ilitazione		*	PluginPatient	]		
	Patologie	Prevale	enza delle Patologie S	Specify [	Data	View Resu	Its Plug	in Help			Patolo	ogie	Prevalen	za delle Pa	tologie Stra	Specify Data	View Results	Plugin Help	]
	Numerosità d	delle Pato	ologie per DRG In	Selezior	Selezionare il reparto						Nume	rosità de	elle Patol	ogie per DF	RG Incid	Selezionare il	l reparto		
			•	Cardio	ogia Ri	abilitazione		* *				Preva	lenza de	lle patolo	gie per inte	Cardiologia	Riabilitazione		
	ICD9_CM:42	28.22		Patol	ogie	Prevale	nza delle	Patologi	e Stratificate pe		1.00					Patologie	Prevalenz	a delle Patolog	ie Stratificate per Sesso
	ICD9_CM:42	e 29.4 Fi	unctional disturbances f	Nume	rosità	delle Pato	logie per	DRG	Incidenza dell	e	0.75 -					Numerosi	tà delle Patolo	gie per DRG	Incidenza delle Patologie
	ICD9_CM:42	24.1 A	ortic valve disorders			Provalon	za dollo	natolog	io nor sosso		0.50 -					12	N	Iomogramma di I	Barber
	ICD9_CM:42	25.4 O	ther primary cardiomyo	1.00 -		revalen	za uene	μαισιου	ie pei sesso	-	0.25 -		1			4.4			Ir = 30 TU = 75%
	ICD9_CM:42	28.1 Le	iastolic heart failure, ch	0.75 -						_	0.00					9	0		Area di accettabilita Cardiologia Riabilitazione
	ICD9_CM:42	28.22 S	ystolic heart failure, chr	0.50 -							0.00 -	ICDS	ICDS	ICDS	ICDS	DM 6	0		Ortopedia Riabilitazione
	Export Data	a										CW:	_CW:7	CW:	CW:7	2.2		Otopedia Riabilitazione	
				0.25 -	_				_			129,4	128.1	125.4	124.1	1.1 3	0 - Cardiologia F	tiavilitazione	
				0.00 -	C	ō	ī		<u></u>	1	Sav	/e Image							
					09_CM	09_CM	D9_C1	09_01	09_01	09_C						0.0	0.0000 1.6667	3.3333 5.0000	6.6667 8.3333 10.0000
					M:429.	M:428.	M:425.	M:424.	1:428.	N:428.							145.	127.	135.
				Sa	/e Imad	ie P	4	4	22	32						Save Im	nage	0	0 0

## Server side developed

- Integated R statistical computing environment to perform survial analysis on demand by selecting
  - the patient set of interest
  - the event related to the survival
  - the stratification variable
- Used in cardiology
  - running on I2b2 version 1.5
  - currently in test for version 1.6



aplanN	leierSta	st 📃							đ	
Drop a	Patient	Set into t	he input b	iaxes bela	w, select one e	vent and then click the "	run Kaplan Meie	r" bu	ation to retrieve information about the analysis.	^
Patien	Set:	Brugad	JaDiagnos	sis_0-70 y	rears@11:30:10	(7-22-2010) (daniele) (1	PATIENTSET_28	3]		
Events										
Synco	pe									
Cardia	nc Arre	st								
Appro	priate l	Discharg	le 🗌							
Run K	aplan N	Aeier Ana	alysis	Kapla	n Meier analysi:	done, please wait for r	esults			
Kaplar	Meier	graph:								
View K	aplan N	Aeier sur	vival curve	18						
Data d	escript	ion:								
			Gr	oup: Fem	ale					
time 46	n.risk	n.event	SUIVIVAI 0.0036	std.err	lower 95% CI	upper 95% CI			Kala Mia Onial One	
52	117	1	0.9851	0.0106	0.9646	0.9646			Kapian Meier Survival Curves	-
58	71	1	0.9712	0.0173	0.9380	0.9380		0		
time	n.risk	n.event	G	std.err	e lower 95% CI	upper 95% CI		-		
15	1131	1	0.9991	0.0009	0.9974	0.9974			Tothe state	
17	1119	1	0.9982	0.0013	0.9958	0.9958		0.8		
18	1114	1	0.9973	0.0015	0.9943	0.9943				
29	984	1	0.9954	0.0021	0.9913	0.9913	i k			
31	941	1	0.9943	0.0023	0.9898	0.9898	abil	0.6	-	
32	927	2	0.9922	0.0028	0.9868	0.9868	prot			
33	895	1	0.9911	0.0030	0.9853	0.9853	val	-		
36	808	5	0.9849	0.0040	0.9771	0.9771	-unc	0.4	1	
37	726	1	0.9837	0.0042	0.9755	0.9755				
40	672	1	0.9795	0.0048	0.9701	0.9701		0		
41	635	4	0.9734	0.0057	0.9623	0.9623		0	1	
42	588	1	0.9717	0.0059	0.9602	0.9602				
43	552	2	0.9682	0.0064	0.9557	0.9557		0	- Female	
45	476	2	0.9641	0.0070	0.9505	0.9505		0		
40	441	4	0.9554	0.0082	0.9395	0.9395			0 20 40 60 80	
48	374	2	0.9411	0.0099	0.9218	0.9218			Time (Veare)	
49	351	2	0.9357	0.0106	0.9152	0.9152			(reas)	
51	284	5	0.9193	0.0127	0.8947	0.8947				
52	251	1	0.9156	0.0132	0.8902	0.8902				
54	210	2	0.9069	0.0144	0.8791	0.8791				
58	194	1	0.9022	0.0151	0.8731	0.8731				
59	119	1	0.8884	0.0178	0.8542	0.8542				
61	94	1	0.8789	0.0199	0.8407	0.8407				
63	74	1	0.8670	0.0229	0.8232	0.8232				
64	63	1	0.8533	0.0264	0.8031	0.8031				
68	23	1	0.8162	0.0442	0.7340	0.7340				
Log-ra	nk test	:								
Group	s	N	Ob	served	Expected					
group	Femal	e 35-	4 3		15.041					
group	male	11:	59 54		41.959					
Chisq:	13.24	1 on 1 de	grees of f	freedom, p	000.0 =					

## 12b2 server side development



◎ pentaho™ open source business intelligence\*

> Pentaho Data Integration Previously Kettle

- Insert the Pentaho Data Mining Community Edition (CE) also known as Weka (original project of university of Waikato, NZ) <u>http://weka.pentaho.com/</u>
- tools for machine learning and data mining in a server side i2b2 cell
  - Suite of classification, regression, association rules and clustering algorithms
  - Design a dedicated plug-in for temporal abstraction



- Insert the **NHLBI Exome Sequencing Project batch query** program in a dedicated i2b2 cells, in order to retrive:
  - information about mutations
  - pubblications related to mutations
- Batch query program:
  - Based on Java 6
  - java -jar YOUR\_DOWNLOADED\_EVS\_CLIENT\_JAR\_FILE -h

# nternationalization

### **International ontologies**

#### International Ontologies:

- ICD9-CM ٠
- **SNOMED** •
- TNM ٠

<b>~</b>		E-C Certain conditions originating in the perinatal period
BioPortal Browse Search Mappings Recommende	r Annotator Resource Inc	🕀 🗇 Complications of pregnancy, childbirth, and the puerperium
		🗊 🛱 Congenital anomalies
International Classification of Dia		🗉 📴 Diseases of the blood and blood-forming organs
International Classification of Dise	eases Terms	🕀 🔂 Diseases of the circulatory system
		🕀 🔂 Diseases of the digestive system
Jump To:	Details Visualization Note	🕀 🔂 Diseases of the genitourinary system
		🕀 🔂 Diseases of the musculoskeletal system and connective tissue
	Preferred Name	🕀 🔂 Diseases of the nervous system and sense organs
	ID	🕀 🔂 Diseases of the respiratory system
		🕀 🔂 Diseases of the skin and subcutaneous tissue
DISEASES OF THE BLOOD AND BLOOD-FORMING ORGANS	Full Id	🕀 📴 Endocrine, nutritional and metabolic diseases, and immunity disorders
B <sup>™</sup> DISEASES OF THE CIRCULATORY SYSTEM	CHD	🕀 🔁 Infectious and parasitic diseases
⊕ DISEASES OF THE DIGESTIVE SYSTEM		🕀 🔁 Injury and poisoning
IDISEASES OF THE GENITOURINARY SYSTEM	ICA	🕀 🔁 Mental disorders
DISEASES OF THE MUSCULOSKELETAL SYSTEM AND CONNECTIVE TISSUE	Company's Trans	🖃 🔁 Neoplasms
DISEASES OF THE NERVOUS SYSTEM AND SENSE ORGANS	Semantic_Type	Enign neoplasms
DISEASES OF THE RESPIRATORY SYSTEM	SOS	E Carcinoma in situ
DISEASES OF THE SKIN AND SUBCUTANEOUS TISSUE		🖃 💭 Malignant neoplasm of bone, connective tissue, skin, and breast
DISORDERS		🗄 🔁 Kaposi's sarcoma
INFECTIOUS AND PARASITIC DISEASES		🖽 🖸 Malignant melanoma of skin
INJURY AND POISONING	ти	H C Malignant neoplasm of bone and articular cartilage
MENTAL, BEHAVIORAL AND NEURODEVELOPMENTAL DISORDERS	1	in the soft tissue
🖨 NEOPLASMS	I UMLS_CUI	🖃 🖸 Malignant neoplasm of female preast
🖮 BENIGN NEOPLASAAS		E Manghant noophaon of axiliary tail of iemale breast
GARCINOMA IN SITU     CARCINOMA IN		Dimensional and the second sec
ALIGNANT NEOPLASM OF BONE, CONNECTIVE TISSUE, SKIN, AND		Dimaiignant neoplasm of central portion of female breast
BREAST		D Malignant neoplasm of lower-inner quadrant of female breast
tar Kaposis sarcoma		Malignant neoplasm of lower-outer quadrant of ternale breast
Malignant metalloma of some and articular cartilage		Malignant neoplasm of other encoded alter of female breast
Maignant neoplasm of connective and other and tissue		Malignant neoplasm of upper inper guedrant of female breast
Malignant neoplasm of female breast		Malignant neoplasm of upper-inner quadrant of lemale breas     Malignant neoplasm of upper-inner quadrant of female breas
Matignant neoplasm of mate preast		Manghant neoplasm of male breast
Other and unspecified malignant neoplasm of skin		The main main and the plasm of male bleast
MALIGNANT NEOPLASM OF DIGESTIVE ORGANS AND PERITONEUM		Malignant reconless of directive organs and neritoneum
MALIGNANT NEOPLASM OF GENITOURINARY ORGANS		Malignant neoplasm of depitourinary organs
Image: Malignant Neoplasm of Lip, Oral Cavity, and Pharynx		Malignant neoplasm of generating organs
MALIGNANT NEOPLASM OF LYMPHATIC AND HEMATOPOIETIC TISSUE		Malignant neoplasm of lymphatic and hematopoietic tissue
🕸 - MALIGNANT NEOPLASM OF OTHER AND UNSPECIFIED SITES 💦 🥃		The Malignant neoplasm of other and unspecified sites
		Malignant neoplasm of respiratory and intrathoracic organs
		Beoplasms of uncertain behavior
		• Reoplasms of unspecified nature

#### i2b2 Query & Analysis Tool

🗄 🔂 Clinicopathological Features

Find Terms

Navigate Terms

🗄 🔂 Biobanks

🗄 🔂 Demographics 

- artilage
- er soft tissue
- nale breast unspecified
- f female breast
- drant of female breast
- drant of female breast a of female breast
- sites of female breast
- drant of female breast
- drant of female breast
- eritoneum
- rynx
- poietic tissue
- sites
- racic organs

## Results

- i2b2 personalization, both on the client side and on the server side
- NLP modules to enrich the clinical information and provide in this way the possibility to query textual reports
- ETL procedures for automatic population and update of the i2b2 CRC
- Implement no/off line data analysis tools for analyze data mart instances

The IT architecture created at FSM is a concrete example of how this type of integration can be implemented and made available for increasing the quality of clinical practice as well as improving the scientific results

## Future ideas: i2b2 and noSQL

- Manage noSQL databases for storage of BIG DATA coming from
  - Continuous patient monitoring systems (Cardiac Holter, Glycaemic Holter,...)
  - Next Generation Sequencing experiments
  - Exposome: mapping the environment to understand the risk of diseases

The idea is to integrate this type of non relational data source with i2b2 using server side cells and/or web plug-ins.



## Future ideas: Shrine

#### Shared Health Research Information Network (SHRINE)

- Developed by Harverd Catalyst: the Harvard Clinical and Translational Science Center
- Compiling large groups of well-characterized patients
- Determine the aggregate total number of patients at partecipating hospitals who meet a given set of inclusion and exclusion criteria
- Because counts are aggregate, patients privacy is protected

#### http://catalyst.harvard.edu/spotlights/shrine.html



- Connections between FSM institutes in the north part of Italy (e.g. biobank network: a prototype could be ready in the coming months)
- Connections between all FSM institutes
- Open to other national and international research institutes

# Clinical and research data integration: the i2b2 FSM experience

## Questions? THANK YOU

Daniele Segagni, MS, Biomedical Engineer daniele.segagni@fsm.it



## Phenotypes from clinical notes

#### **Medical Report**





Kohane IS. Using electronic health records to drive discovery in disease genomics. Nat Rev Genet. 2011 Jun;12(6):417-28. Epub 2011 May 18. Review. PubMed PMID: 21587298.

## I2b2 data for clinical issues

- i2b2- ontology relying NCBO BioPortal <u>http://bioportal.bioontology.org</u>
- Analysis based on ICD9-CM codes related to
  - diagnosis
  - procedures
- Selection of patients with diagnosis of
  - malignant neoplasm of female breast (ICD9-CM code 174.0-174.9)
  - personal history of malignant neoplasm (ICD9-CM V10.3)
- Starting from 28.838 patients we actually focused our attention on 8.969

## I2b2 data for clinical issues

#### **Oncology Medical Unit**



## I2b2 data for clinical issues

Which is the actual patient care flow? Which is the adherence with guidelines?

- Off line analysis on clinical procedures and dayhospital/inpatient treatment using
  - Temporal data mining
  - Process mining
- Extract the common care flows in FSM oncology division
  - the most common pathway is hospitalization in breast surgery division followed by day-hospital treatements
- Extraction of most frequent sequences of procedures
  - The most common patterns extracted matched with the implementation of userguide
  - Example: pattern describing control ECGs that are performed due to the potential cardiotoxicity of some chemotherapies