

# How pathology registries can be integrated with the i2b2 framework

Valentina Tibollo

5<sup>th</sup> i2b2 Meeting 2017

# Scientific Clinical Institute Maugeri - ICSM

## Hospital network

- 18 centers
- 6 Italian regions
- Rehabilitative medicine field



# i2b2

## Since 2010

### Data sources:

- Hospital Information System (HIS)
- EHR
- Biobank

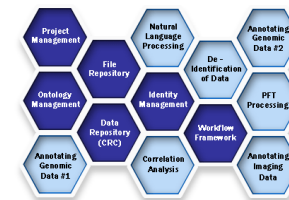
...

### Fields of interest:

- Type 2 diabetes
- Cardiology
- Oncology
- Nephrology



integration



# REDCap

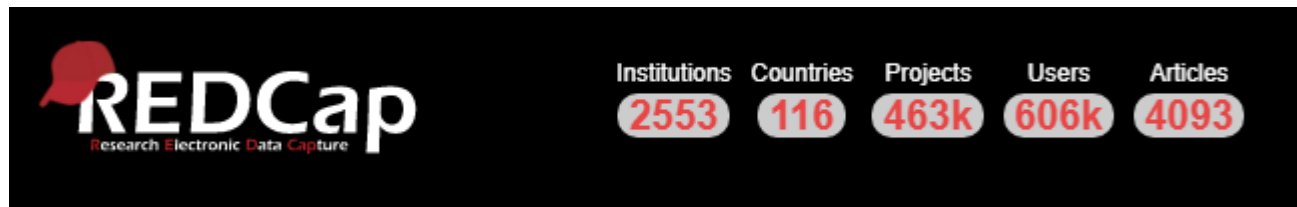
Since 2016, new dispositions for pathology registries by the Italian health ministry

## Main benefits:

- Important elements to organize clinical trials
- Help physicians to update standards for pathology management
- More focused attention and regular follow-ups

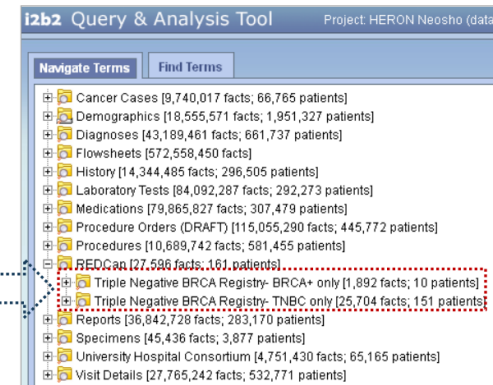
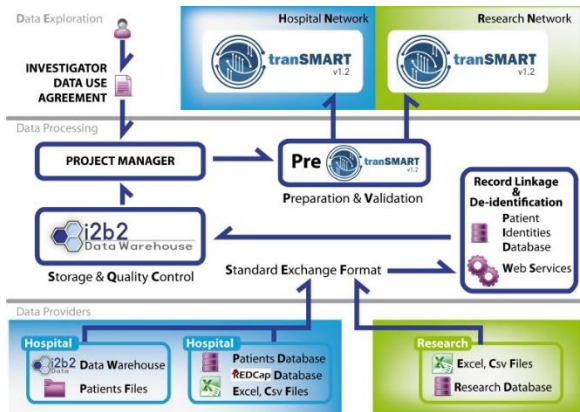
## ICSM multi-centers Projects:

- Heart failure
- Stroke
- Respiratory diseases
- Parkinson syndrome



# Previous works

- Waitman LR, Warren JJ, Manos EL, Connolly DW. Expressing Observations from Electronic Medical Record Flowsheets in an i2b2 based Clinical Data Repository to Support Research and Quality Improvement. AMIA Annu Symp Proc. 2011;2011:1454-63. Epub 2011 Oct 22
- James Gregoric, REDCap-to-i2b2 Loader, March 7, 2012, i2b2 Community Wiki
- Pittman, Christine A., and Amrendra S. Miranpuri. "Neurosurgery clinical registry data collection utilizing Informatics for Integrating Biology and the Bedside and electronic health records at the University of Rochester." *Neurosurgical focus* 39.6 (2015): E16.
- Rance, Bastien, et al. "Integrating heterogeneous biomedical data for cancer research: the CARPEM infrastructure." *Applied clinical informatics* 7.2 (2016): 260.
- Mauro Bucalo, i2b2: integration with REDCap and multi-project management, 4th i2b2 Meeting 2016



# Basic assumption

## REDCap:

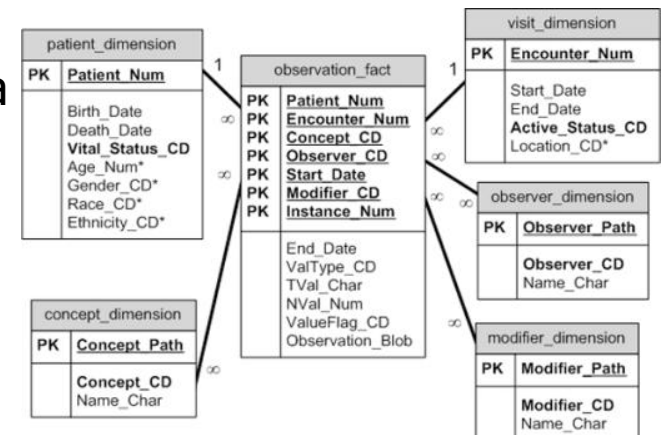
- Metadata table that defines all the study fields with data types
- Data grouped in events
- Entity Attribute Value (EAV) model

REDCAP_DATA		REDCAP_METADATA	
EVENT_ID		FORM_NAME	
FIELD_NAME		FIELD_NAME	
PROJECT_ID		PROJECT_ID	
RECORD		ELEMENT_TYPE	
VALUE		ELEMENT_VALIDATION_TYPE	

	Enrollment	Month 1	Month 2	Event	End of study
<b>Record ID</b>	Basic Demography Form	Visit Data	Visit Data	Adverse Events Page	Completion Data
2 (Bianchi,Lucia)	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
5 (Viola,Rita)	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6 (Riva,Anna)	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## i2b2:

- Concept dimension analogous to REDCap metadata
- Data grouped in encounters/visits
- Central observation fact table (EAV model)



➔ A REDCap event corresponds to an i2b2 visit



# Main differences

## REDCap:

- Data stored in studies, forms, events, fields
- Different types of fields: text, check-box, multiple choice
- Fields can be organized in tables and repeating fields collections

Editing existing RECORD ID 9-1

Event Name: **Visita 1**

RECORD ID 9-1

CODIFICA ATC	Nome Commerciale	Classe	Data inizio	Data fine	Dosaggio tot giornaliero
formoterol and be... R03AK08	foster	LABA-ICS	19-04-2017 Today D-M-Y		100/6
salbutamol R03AC02	ventolin	SABA	01-01-1997 Today D-M-Y		100
montelukast R03DC03	montelukast	ALTRO	19-04-2017 Today D-M-Y		10
methylprednisolone H02AB04	deltacortene	ALTRO	19-04-2017 Today D-M-Y	23-04-2017 Today D-M-Y	25
methylprednisolone H02AB04	deltacortene	ALTRO	24-04-2017 Today D-M-Y	28-04-2017 Today D-M-Y	12.5
methylprednisolone H02AB04	deltacortene	ALTRO	29-04-2017 Today D-M-Y	03-05-2017 Today D-M-Y	6.25

## i2b2:

- Data stored according to an ontology of terms
- Each i2b2 observation requires a start date

Navigate Terms Find Terms

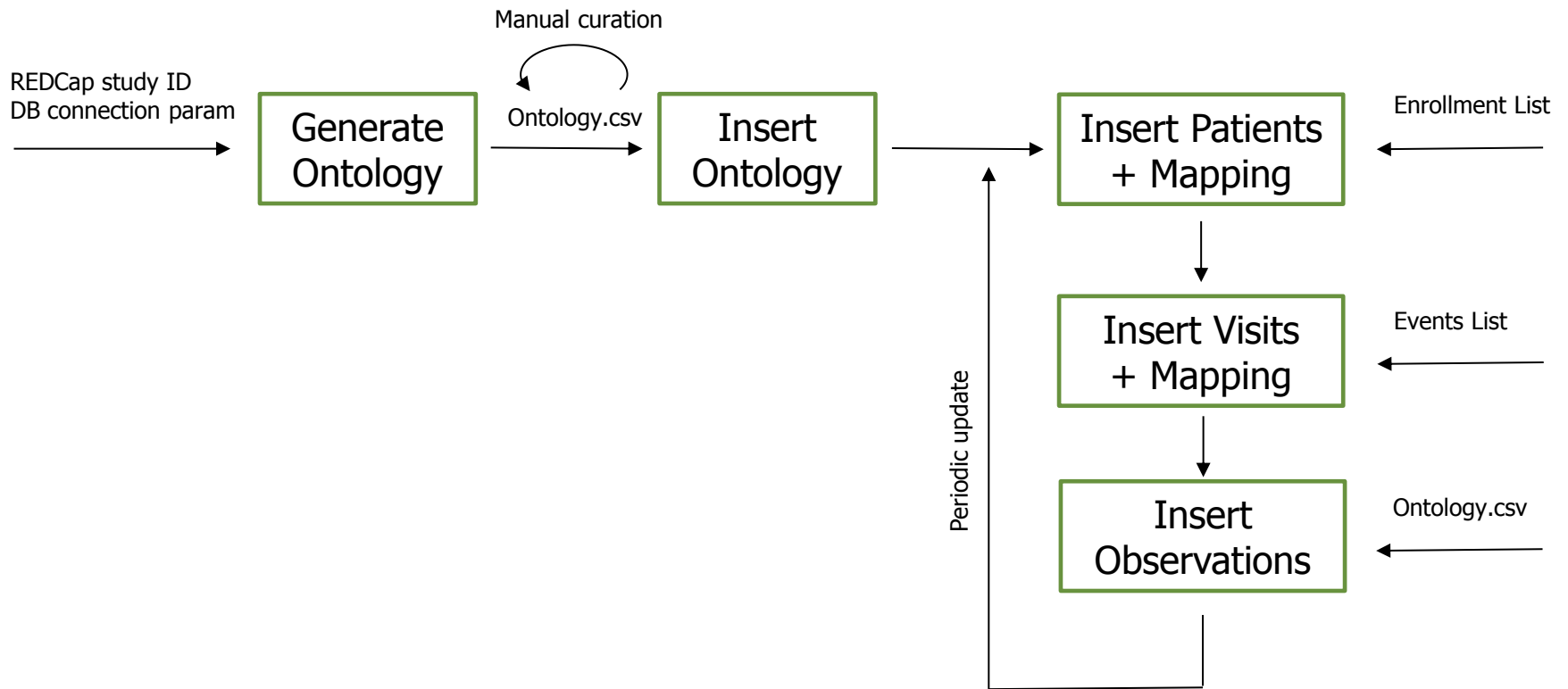
- Clinical Trials
- Demographics
  - Age
  - Gender
  - Ethnicity
  - Marital Status
  - Race
  - Radial Distance
  - Vital Status
- Diagnoses
  - Primary Diagnosis
  - Secondary Diagnosis
  - Diseases And Injuries
  - Supplementary Classification Of External Causes Of Injur
  - Supplementary Classification Of Factors Influencing Healt
- Laboratory Tests
- Medications
- Procedures (ICD-9)
- Vital Signs

observation_fact	
PK	Patient_Num
PK	Encounter_Num
PK	Concept_CD
PK	Observer_CD
PK	Start_Date
PK	Modifier_CD
PK	Instance_Num
End_Date	
ValType_CD	
TVal_Char	
NVal_Num	
ValueFlag_CD	
Observation_Blob	



# Method Overview

General ETL procedure, developed using KETTLE software (Pentaho suite)





# Ontology Generation I

Ontology.csv file automatically generated:

- Study name as main folder
- A folder for each form
- Text box fields → ontology leaf
- Multiple/single choice fields → parent folder containing leaves for corresponding choices
- Repeating fields → single ontology leaf
- Date fields → ignored, used for observation creation

Manual curation step:

- Delete unwanted fields (headers, descriptive fields, fields created for branching logic purposes, ...)
- For each observation, manually specify the related start/end date fields and the default start/end date field.

The image shows two screenshots from the REDCap interface. The left screenshot displays the 'Data Collection' section for a project named 'Contatto 1'. Under 'Data Collection Instruments', a list of forms is shown with radio buttons next to them. The 'Contatto' form is selected. A red dashed box highlights the list of instruments, and a red arrow points from this box to the right screenshot. The right screenshot shows the 'i2b2 BIOMEDIS Query' interface, which displays a hierarchical tree of ontology terms. The 'Scompenso Cardiaco' folder is expanded, and a red dashed box highlights the list of terms under it, including 'Anagrafica', 'Contatto', 'Cartella Infermieristica', 'Scale di valutazione', 'Comorbidità', 'Fattori di rischio', 'Complicanze', 'Esami strumentali', 'Esami laboratorio', 'Diagnosi e procedure', and 'Terapie'. The 'Contatto' term is highlighted in blue in both screenshots, indicating the mapping between the form and the ontology term.



# Ontology Generation II

## Contatto

Re-assign this record to another Data Access Group? Pavia

Editing existing RECORD ID 1 1211

Event Name: **Contatto 1**

RECORD ID: 1

Numero SDO  \* must provide value

Regime degenza: Ordinario (dropdown) \* must provide value

Data ricovero:  Today D-M-Y \* must provide value

Data dimissione:  Today D-M-Y

Modalità dimissione: ORDINARIA AL DOMICILIO (dropdown) \* must provide value

Provenienza: Trasferito da ospedale pubblico (dropdown)

Note Anamnestiche:

Ricoveri per SC nei 6 mesi precedenti:  Yes  No

Classe NYHA:  valori ammissibili: 1, 2, 3, 4 \* must provide value

Fibrillazione atriale:  Yes  No

Form Status

Complete?: Incomplete (dropdown)

Lock this record for this form?  Lock

If locked, no user will be able to edit this record on this form until someone with Lock/Unlock privileges unlocks it.

Save Record  
Save and Continue  
Save and go to Next Form

Navigate Terms Find Terms

- Clinical Trials
  - ScompensoCardiaco
    - 001 - Anagrafica
    - 002 - Contatto
      - 001 - Numero SDO
      - 002 - Regime degenza
        - 001 - Ordinario
        - 002 - Day Hospital
        - 003 - Accompagnatore
        - 004 - Nessuna corrispondenza
        - 005 - Sconosciuto
      - 003 - Modalità dimissione
      - 004 - Provenienza
      - 005 - Note Anamnestiche
      - 006 - Ricoveri per SC nei 6 mesi precedenti
      - 007 - Classe NYHA
      - 008 - Fibrillazione atriale
      - 009 - Form Status
    - 003 - Cartella Infermieristica
    - 004 - Scale di valutazione
    - 005 - Comorbidità
    - 006 - Fattori di rischio
      - 001 - Fumo
      - 002 - Fumo - quantità
      - 003 - Abuso di alcool
      - 004 - Familiarità
      - 005 - Totale fattori di rischio
      - 006 - Form Status
    - 007 - Complicanze
    - 008 - Esami strumentali
    - 009 - Esami Laboratorio
    - 010 - Diagnosi e procedure
    - 011 - Terapia



# Patients and visits

## Features:

- ICSM REDCap studies are de-identified.
- Same procedure used during i2b2 patients insertion → REDCap patient identifier = i2b2 patient\_id
- One visit for each event specified in the Event List file
- Each event has a set of possible REDCap date fields to specify start/end date



# Observations

## Features:

- Different strategies for
  - Numeric or calculated fields
  - Select, radio and check-box fields
  - Simple text fields
  - Text area fields
- Repeating fields: use of instance\_num
- Start date
  - Mapped in the ontology file. If mapped, but not compiled, default date field is used (corresponding to the visit start date)
  - If not mapped, 1<sup>st</sup> date field of the form
  - If no date field, from the REDCap LOG table
  - Automatic procedure to get date for repeating fields



# Results

## REDCap Heart Failure Registry

- 1600 patients since 2013 from 2 ICSM centers

## I2b2 instance

- 56 thousands patients from hospitalizations, weekly updated
- 20 thousands patients with disease of the circulatory system

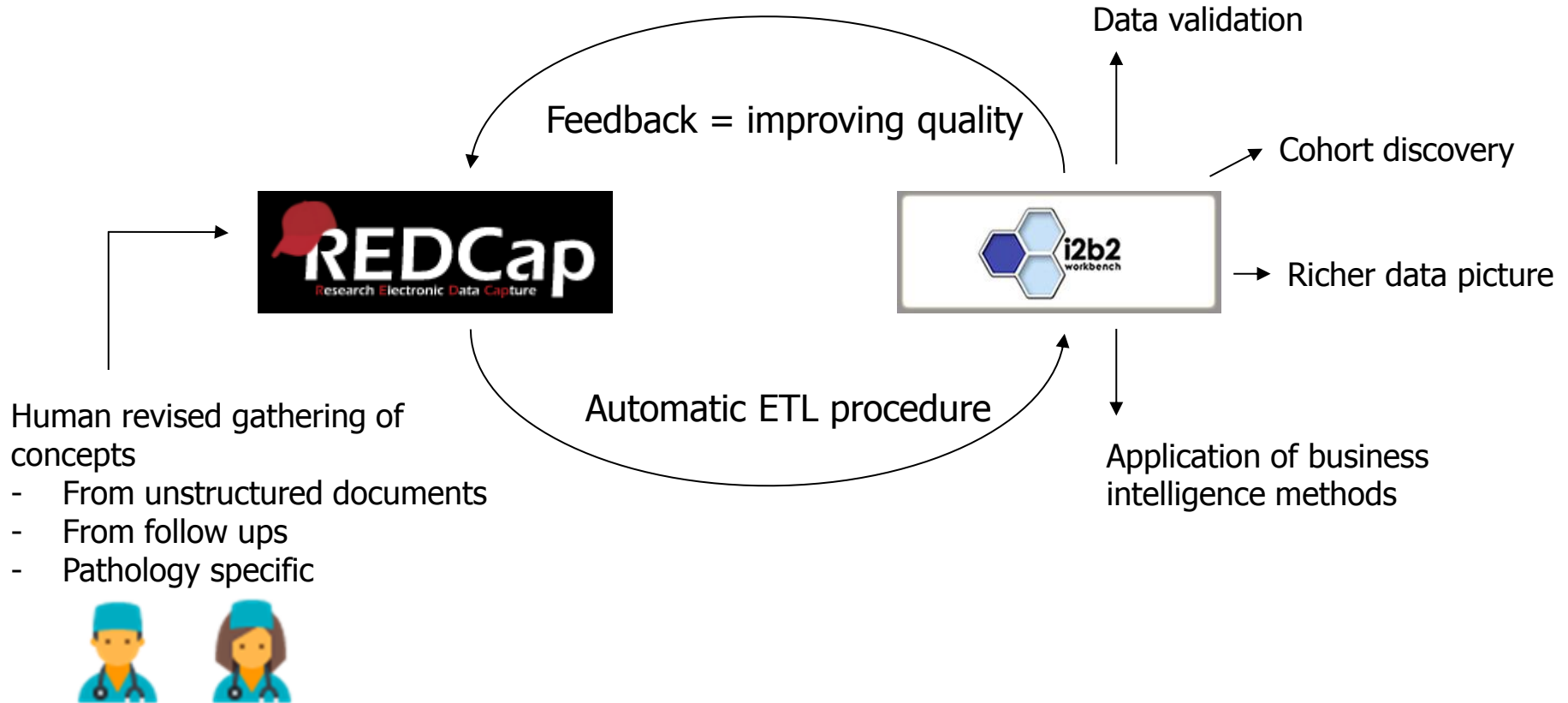


More than 70 thousands observations created

- Demographics
- Hospitalization data
- Laboratory tests
- Risk factors
- Comorbidity
- Rehabilitation data
- Therapy prescriptions during hospitalization



# REDCap-i2b2 integration



# Thanks

Valentina Tibollo  
LISRC Labs, Scientific Clinical Institute  
Maugeri, Pavia  
[valentina.tibollo@icsmaugeri.it](mailto:valentina.tibollo@icsmaugeri.it)

