

ODM to i2b2

Ward Blondé, 2015/03/13, Utrecht
BioMedBridges, TraIT, VUmc

Overview

- Tools and standards
 - Electronic Data Capture systems
 - OpenClinica
 - ODM/XML
 - i2b2
- ODM to i2b2
 - 3 tabular files
 - REDCap to i2b2

Electronic Data Capture systems


- Systematic collection of clinical data
- Use ODM: OpenClinica, REDCap
- Items are the smallest data unit:
 - Numerical values
 - Categorical values: yes/no/maybe
 - Free text

Electronic Data Capture systems

- Events: intake, doctor's visit, treatment, questionnaire
- For each event type: 1 or more Case Report Forms (CRFs)
- Partonomic organization of the data:
 - Event (root)
 - CRF (branch)
 - Item group (branch)
 - Item (leaf)

OpenClinica










← https://sandbox.openclinica.nl/OpenClinica/ListStudySubjects

 Silicon digestion (Silicon digestion) | Change Study/Site wblonde (Data Manager) en | Log Out




Home | Subject Matrix | Notes & Discrepancies | Study Audit Log | Tasks ▾ Report Issue | Support | Study Subject ID Go

Alerts & Messages ▾
Instructions ▾
Info ▾
Icon Key ▾

Statuses










-  Not Started
-  Scheduled
-  Data Entry Started
-  Stopped
-  Skipped
-  Completed
-  signed
-  Locked
-  Invalid

Actions

-  View
-  Edit
-  Remove

Subject Matrix for Silicon digestion ?

◀ ◀ ▶ ▶ 15 Show More Select An Event ▾ Add New Subject

Study Subject ID	NonRep Event	Repeating Event	doktersbezoek	Actions
				Apply Filter Clear Filter
SIL_10_10001_0		 x2	 x3	  
SIL_10_10002_1		 x2		

Results 1 - 2 of 2.

Subject: SIL_10_10001_0 X

Event: doktersbezoek Add Another Occurrence 1|2|3

Occurrence#1 of 3	Occurrence#2 of 3	Occurrence#3 of 3
15-Jul-2014	02-Jul-2014	20-Jul-2015
Status: data entry started	Status: data entry started	Status: data entry started

Click for more options

ODM/XML

- Operational Data Model
- An XML standard for EDC systems
- Export format of OpenClinica and RedCap
- Allows some OpenClinica-specific extensions

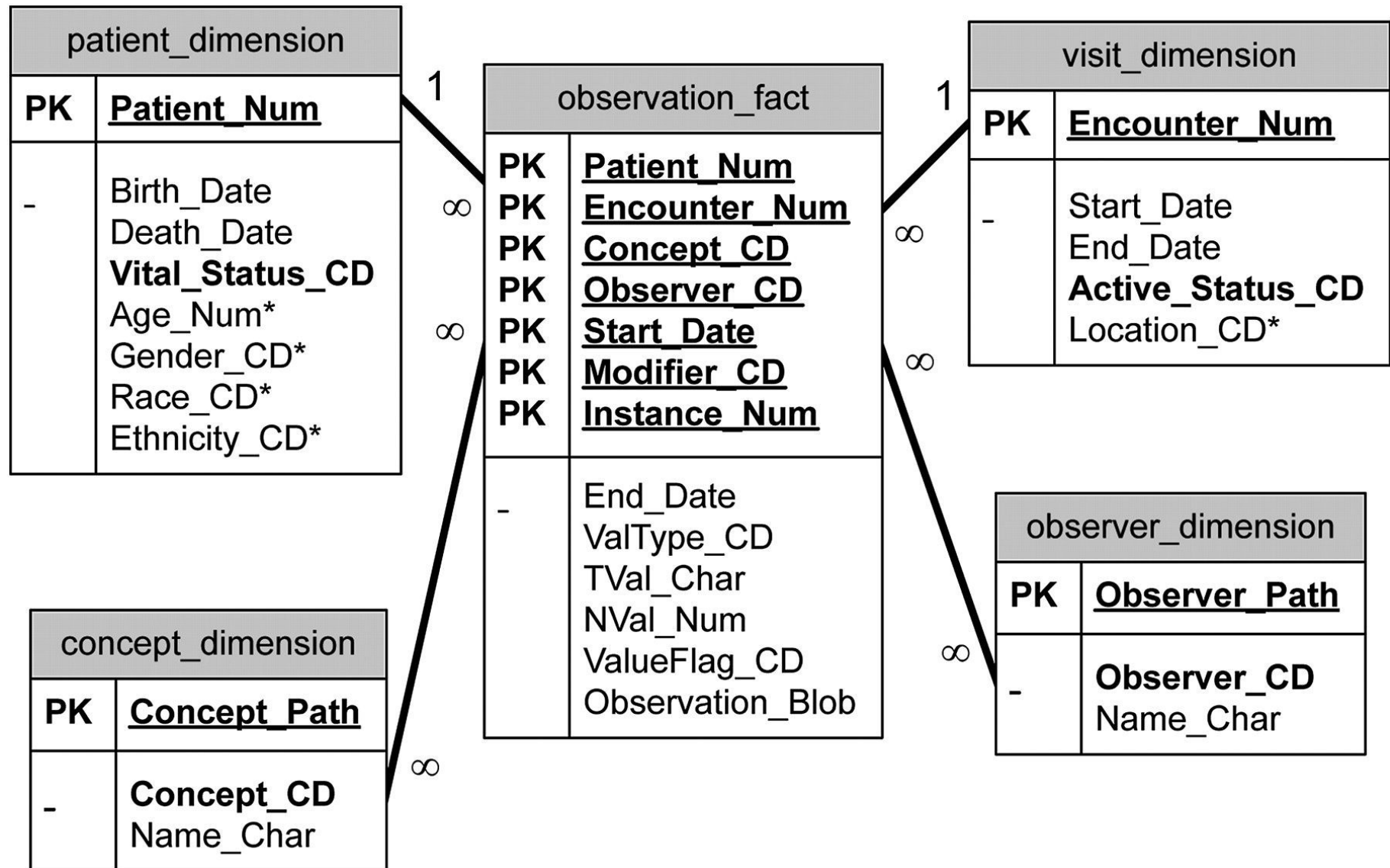
ODM/XML

```
1 <?xml version="1.0" encoding="US-ASCII"?><ODM xmlns="http://www.cdisc.org/ns/odm/v1.3" xmlns:OpenClinica="
http://www.openclinica.org/ns/odm_ext_v130/v3.1" xmlns:OpenClinicaRules="http://www.openclinica.org/ns/rules/v3.1"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" FileOID="Dataset_for_TransMART_D20140501144126+0200"
Description="the CRFs for subjects n001, n081 and n2222 are completed by Naomi" CreationDateTime=
"2014-05-01T14:41:26+02:00" FileType="Snapshot" ODMVersion="1.3" xsi:schemaLocation="
http://www.cdisc.org/ns/odm/v1.3 OpenClinica-ODM1-3-0-OC2-0.xsd">
2   <Study OID="S_POSEIDON">
3     <GlobalVariables>
4       <StudyName>Poseidon</StudyName>
5       <StudyDescription>
6         Poseidon
7       </StudyDescription>
8       <ProtocolName>Poseidon</ProtocolName>
9     </GlobalVariables>
10    <MetaDataVersion OID="v1.0.0" Name="MetaDataVersion_v1.0.0">
11      <Protocol>
12        <StudyEventRef StudyEventOID="SE_SELECTIONCRITERIA" OrderNumber="1" Mandatory="Yes"/>
13        <StudyEventRef StudyEventOID="SE_HOSPITALS" OrderNumber="2" Mandatory="Yes"/>
14        <StudyEventRef StudyEventOID="SE_DIAGNOSIS" OrderNumber="3" Mandatory="Yes"/>
15        <StudyEventRef StudyEventOID="SE_TREATMENTS" OrderNumber="4" Mandatory="Yes"/>
16        <StudyEventRef StudyEventOID="SE_DISEASECOURSE" OrderNumber="5" Mandatory="Yes"/>
17        <StudyEventRef StudyEventOID="SE_RESOURCEUSE" OrderNumber="6" Mandatory="Yes"/>
18        <StudyEventRef StudyEventOID="SE_BIOMARKERS" OrderNumber="7" Mandatory="Yes"/>
19      </Protocol>
20      <StudyEventDef OID="SE_SELECTIONCRITERIA" Name="Selection Criteria" Repeating="No" Type="Scheduled">
21        <FormRef FormOID="F_POSSELECTION_V02" Mandatory="Yes"/>
22        <OpenClinica:EventDefinitionDetails StudyEventOID="SE_SELECTIONCRITERIA">
23          <OpenClinica:Description>Selection Criteria</OpenClinica:Description>
24        </OpenClinica:EventDefinitionDetails>
```

i2b2

- Informatics for integrating **biology** & the **bedside**
- Platform for analyzing data in an EDC system
- Data is organized in a star schema:
 - Central dimension: observation
 - Peripheral dimensions:
 - Patient
 - Concept
 - Event
 - Observer
 - Modifier

i2b2 Star Schema



i2b2 Workbench

The screenshot displays the i2b2 Workbench interface, which is divided into several panes. The top-left pane shows a navigation tree with categories like Custom Metadata, Demographics, and Diagnoses. The top-right pane is the Query Tool, where a query named 'Female-Digesti@04:56:38' is being built. It features three groups for defining criteria, with Group 1 containing 'Female' and Group 2 containing 'Digestive system [Principal Diagnosis]'. The bottom-right pane shows the Analysis View, which includes a bar chart titled 'Number of patients for "Female-Digesti@04:56:38"'. The chart shows a single bar with a value of 36. The bottom-left pane shows a search bar and a list of results, including 'Female-Digesti@04:56:38 [02-11-2015] [demo]'.

Query Tool

Query Name: Female-Digesti@04:56:38

Group 1	Group 2	Group 3
<input type="checkbox"/> Dates <input type="checkbox"/> Occurs > 0x <input type="checkbox"/> Exclude	<input type="checkbox"/> Dates <input type="checkbox"/> Occurs > 0x <input type="checkbox"/> Exclude	<input type="checkbox"/> Dates <input type="checkbox"/> Occurs > 0x <input type="checkbox"/> Exclude
Treat Independently	Treat Independently	Treat Independently
Female	Digestive system [Principal Diagnosis]	

Analysis Types

- ☐ Patient set
- ☐ Encounter set
- ☒ Number of patients
- ☐ Gender patient breakdown
- ☐ Vital Status patient breakdown
- ☐ Race patient breakdown
- ☐ Age patient breakdown
- ☒ Timeline

Query Timing

- ☒ Treat all groups independent
- ☐ Selected groups occur in the
- ☐ Items instance will be same

Get Everyone **Run Query Above** Patient(s) returned: 36

Timeline View

Graphic Analyses

Female-Digesti@04:56:38 [02-11-2015] [demo]

Number of patients for "Female-Digesti@04:56:38"

37,5
35,0
32,5
30,0

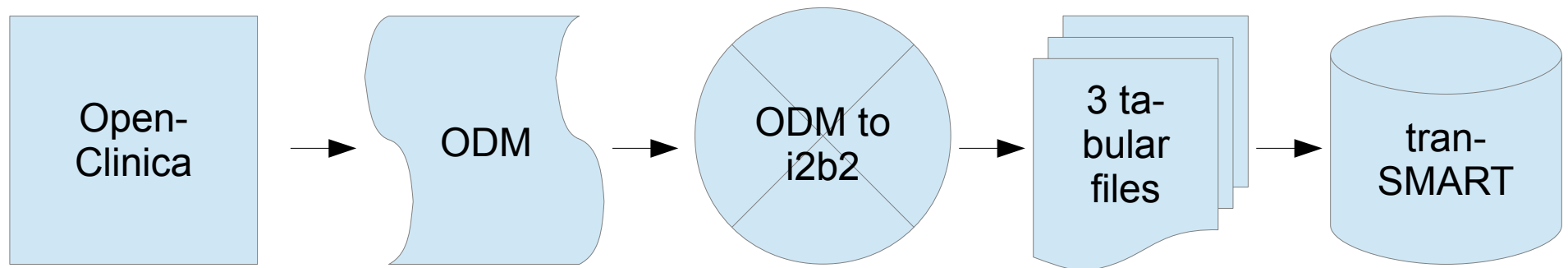
36

Overview

- Tools and standards
 - Electronic Data Capture systems
 - OpenClinica
 - ODM/XML
 - i2b2
- ODM to i2b2
 - 3 tabular files
 - REDCap to i2b2

ODM to i2b2

- Actually OpenClinica to tranSMART
- tranSMART has defined an import format consisting of 3 tabular files
- Shell scripts loads the 3 tabular files in the i2b2-database below tranSMART



3 tabular files

- Clinical data file:
 - Contains all the data
- Columns file:
 - Contains metadata for each column in the clinical data file
- Wordmap file:
 - Maps categorical word values to integers in order to have a shorter representation in the clinical data file

Clinical data file

- 5 columns form a primary key together:
 - Subject ID = Patient_num in i2b2
 - Event type ID = Encounter_num in i2b2
 - Event repeat key
 - Item group type ID
 - Item group repeat key = Instance_num in i2b2
- The rows are either patients, event repeats, or item group repeats
- The other columns represent concepts

Clinical data file

	A	B	C	D	E	F	G	H	I	J
1	Patient_num	Encounter_num	Encounter_name	Encounter_repeat_key	Item_group_id	Item_group_name	Instance_num	The p	This	The
2	SS_N2222							1	1	1
3	SS_N2222	SE_HOSPITALS	Hospitals		IG_POSHO_HOSPITALS	Hospitals	1			
4	SS_N2222	SE_DISEASECOURSE	Disease course		IG_POSDI_PROGRESSION	progression	1			
5	SS_N2222	SE_RESOURCEUSE	Resource use	1						
6	SS_N2222	SE_RESOURCEUSE	Resource use	1	IG_POSRE_OUTPATIENTV	OutpatientV	1			
7	SS_N2222	SE_RESOURCEUSE	Resource use	1	IG_POSRE_OUTPATIENTV	OutpatientV	2			
8	SS_N2222	SE_RESOURCEUSE	Resource use	1	IG_POSRE_OUTPATIENTV	OutpatientV	3			
9	SS_N2222	SE_RESOURCEUSE	Resource use	1	IG_POSRE_OUTPATIENTV	OutpatientV	4			
10	SS_N2222	SE_RESOURCEUSE	Resource use	1	IG_POSRE_OUTPATIENTV	OutpatientV	5			
11	SS_N2222	SE_RESOURCEUSE	Resource use	1	IG_POSRE_OUTPATIENTV	OutpatientV	6			
12	SS_N2222	SE_RESOURCEUSE	Resource use	1	IG_POSRE_OUTPATIENTT	OutpatientT	1			
13	SS_N2222	SE_RESOURCEUSE	Resource use	1	IG_POSRE_IMAGINGPROCEDURES	ImagingProcedures	1			
14	SS_N2222	SE_RESOURCEUSE	Resource use	1	IG_POSRE_IMAGINGPROCEDURES	ImagingProcedures	2			
15	SS_N2222	SE_RESOURCEUSE	Resource use	1	IG_POSRE_PATHOLOGY	Pathology	1			
16	SS_N2222	SE_BIOMARKERS	Biomarkers		IG_POSBI_BIOMARKERS	Biomarkers	1			
17	SS_N2222	SE_BIOMARKERS	Biomarkers		IG_POSBI_BIOMARKERS	Biomarkers	2			
18	SS_N2222	SE_BIOMARKERS	Biomarkers		IG_POSBI_BIOMARKERS	Biomarkers	3			
19	SS_N2222	SE_BIOMARKERS	Biomarkers		IG_POSBI_BIOMARKERS	Biomarkers	4			
20	SS_N2222	SE_BIOMARKERS	Biomarkers		IG_POSBI_BIOMARKERS	Biomarkers	5			
21	SS_N2222	SE_BIOMARKERS	Biomarkers		IG_POSBI_BIOMARKERS	Biomarkers	6			
22	SS_N081							1	1	1
23	SS_N081	SE_HOSPITALS	Hospitals		IG_POSHO_HOSPITALS	Hospitals	1			
24	SS_N081	SE_HOSPITALS	Hospitals		IG_POSHO_HOSPITALS	Hospitals	2			
25	SS_N081	SE_TREATMENTS	Treatments	1						
26	SS_N081	SE_TREATMENTS	Treatments	1	IG_POSTR_DRUG	drug	1			
27	SS_N081	SE_TREATMENTS	Treatments	2						

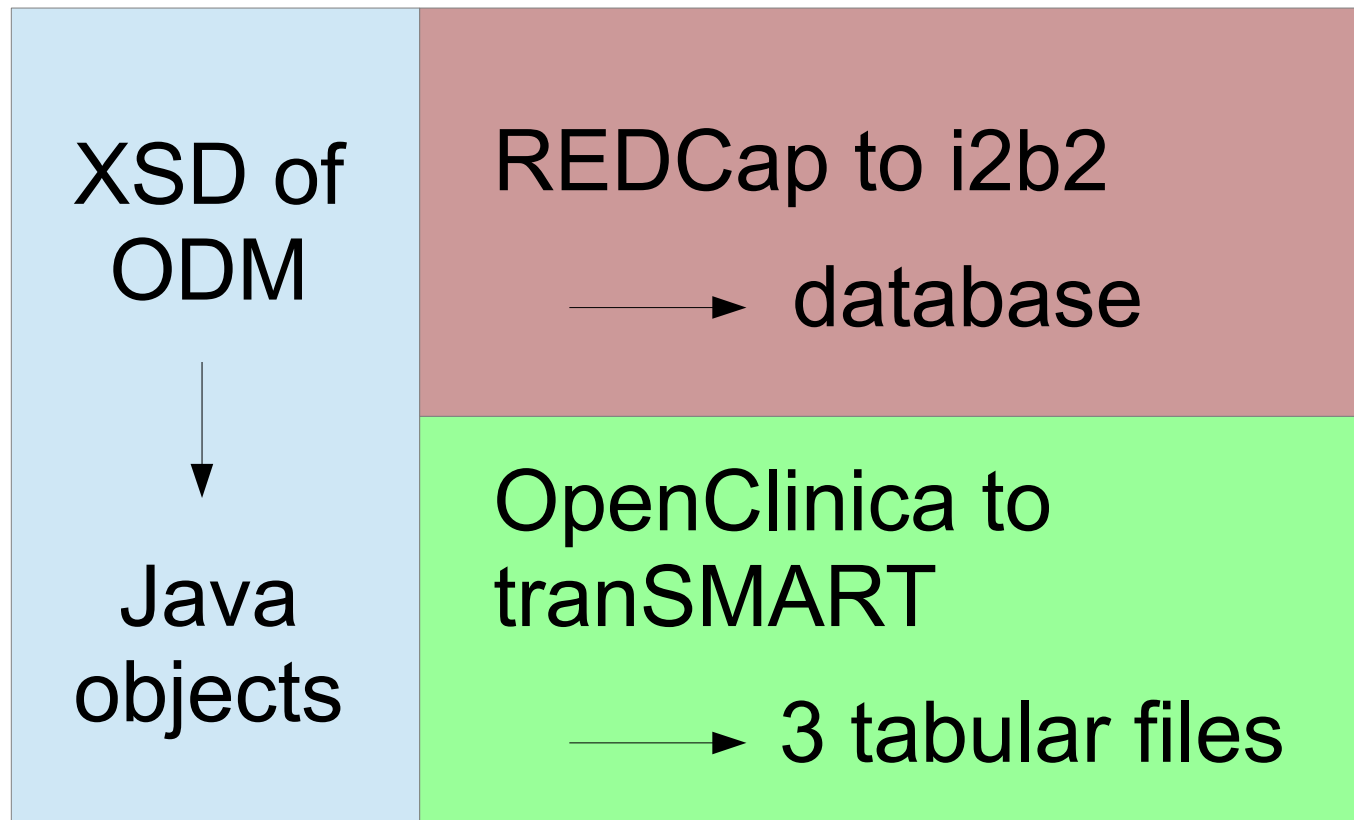
3 tabular files

- Columns file: puts concepts in the hierarchy

	A	B	C	D	E	F	G
1	Filename	Category Code	Column Number	Data Label	Data Label Source	Control Vocab Cd	
2	Poseidon_clinical_data.txt		1	SUBJ_ID			
3	Poseidon_clinical_data.txt	ID	2	encounter_id			
4	Poseidon_clinical_data.txt	ID	3	encounter_name			
5	Poseidon_clinical_data.txt	ID	4	encounter_repeat_key			
6	Poseidon_clinical_data.txt	ID	5	item_group_id			
7	Poseidon_clinical_data.txt	ID	6	item_group_name			
8	Poseidon_clinical_data.txt	ID	7	item_group_repeat_key			
9	Poseidon_clinical_data.txt	Selection Criteria+Pos: Selection criteria - v0.2+IG_POSSE_	8	The patient was diagnosed with non-small cell lung cancer (NSC			
10	Poseidon_clinical_data.txt	Selection Criteria+Pos: Selection criteria - v0.2+IG_POSSE_	9	This first diagnosis (preferably using pathology) was made after			
11	Poseidon_clinical_data.txt	Selection Criteria+Pos: Selection criteria - v0.2+IG_POSSE_	10	The patient did not have (diagnosed) NSCLC before 01/06/201			
12	Poseidon_clinical_data.txt	Hospitals+Pos: Hospitals - v0.6+Hospitals	11	Hospital - Hospital			
13	Poseidon_clinical_data.txt	Hospitals+Pos: Hospitals - v0.6+Hospitals	12	Contact information, if relevant (name department /clinician) -			
14	Poseidon_clinical_data.txt	Hospitals+Pos: Hospitals - v0.6+Hospitals	13	Data availability - Data availability			
15	Poseidon_clinical_data.txt	Hospitals+Pos: Hospitals - v0.6+Hospitals	14	Data entry for this hospital is - Data entry for this hospital is			
16	Poseidon_clinical_data.txt	Hospitals+Pos: Hospitals - v0.6+Hospitals	15	If not finished, complete up to: (dd-mmm-yyyy) - If not finished,			
17	Poseidon_clinical_data.txt	Diagnosis+Pos: Diagnosis - v0.8+IG_POSDI_UNGROUPED	16	Date of Birth - dd-mmm-yyyy			
18	Poseidon_clinical_data.txt	Diagnosis+Pos: Diagnosis - v0.8+IG_POSDI_UNGROUPED	17	Gender			
19	Poseidon_clinical_data.txt	Diagnosis+Pos: Diagnosis - v0.8+IG_POSDI_UNGROUPED	18	Date of first NSCLC-related hospital visit - dd-mmm-yyyy			
20	Poseidon_clinical_data.txt	Diagnosis+Pos: Diagnosis - v0.8+IG_POSDI_UNGROUPED	19	Hospital of first histology/cytology diagnosis			
21	Poseidon_clinical_data.txt	Diagnosis+Pos: Diagnosis - v0.8+IG_POSDI_UNGROUPED	20	Mode of diagnosis			
22	Poseidon_clinical_data.txt	Diagnosis+Pos: Diagnosis - v0.8+IG_POSDI_UNGROUPED	21	Date of first histology/cytology diagnosis - dd-mmm-yyyy			
23	Poseidon_clinical_data.txt	Diagnosis+Pos: Diagnosis - v0.8+IG_POSDI_UNGROUPED	22	Tumor type			
24	Poseidon_clinical_data.txt	Diagnosis+Pos: Diagnosis - v0.8+IG_POSDI_UNGROUPED	23	WHO Performance status at diagnosis			
25	Poseidon_clinical_data.txt	Diagnosis+Pos: Diagnosis - v0.8+IG_POSDI_UNGROUPED	24	Smoking			
26	Poseidon_clinical_data.txt	Diagnosis+Pos: Diagnosis - v0.8+IG_POSDI_UNGROUPED	25	Comorbidity (relevant at diagnosis and/or later)			
27	Poseidon_clinical_data.txt	Diagnosis+Pos: Diagnosis - v0.8+IG_POSDI_UNGROUPED	26	Cardiac			

REDCap to i2b2

- Fork of an open source project



Thanks for your attention!