Clinical Data warehouse / ETL

Master Public Health Data Science

Vianney Jouhet, MD, PhD

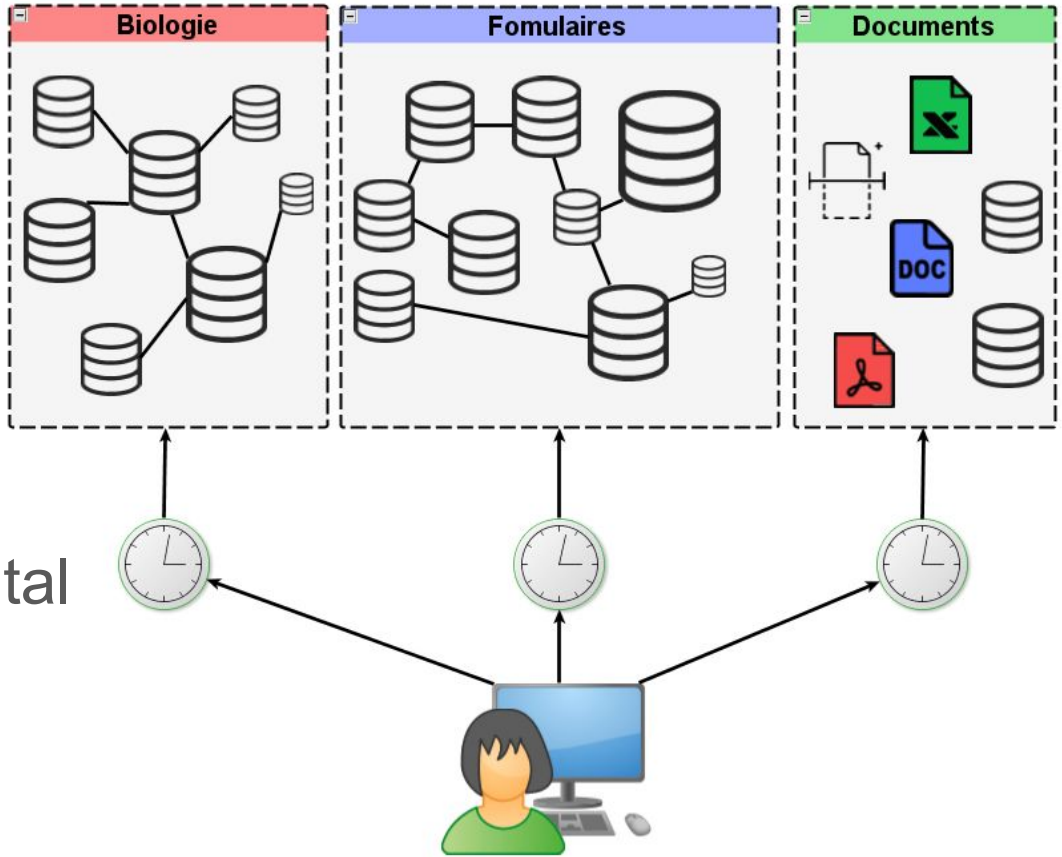
Issues

Separated data (silos)

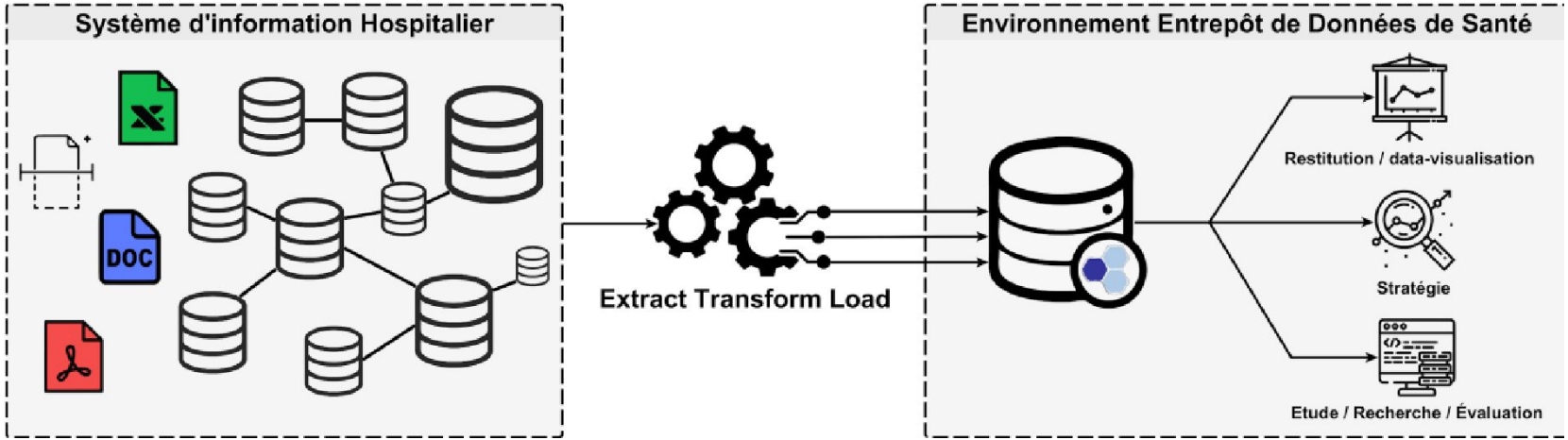
Data Heterogeneity

Bordeaux University Hospital

- 100+ Applications
- 10 000+ Tables



Clinical data warehouse (CDW)



ETL - Extract

- Extract data from Hospital Information System
  - Need knowledge of data model
  - This usually needs to be reverse engineered
- Another solution
  - Use exchange standard
  - Leverage standardised messages
  - Necessitate to listen data flows in real time
  - FHIR may offer easier solution for data extraction

ETL - Transform

Transform data from HIS

- Technical transformation
  - Relational model to CDW information model
- Semantic transformation
  - Harmonize terminologies
  - Annotate information (meta modeling)
  - Leverage Ontologies and terminologies
  - May benefit from multi terminology server

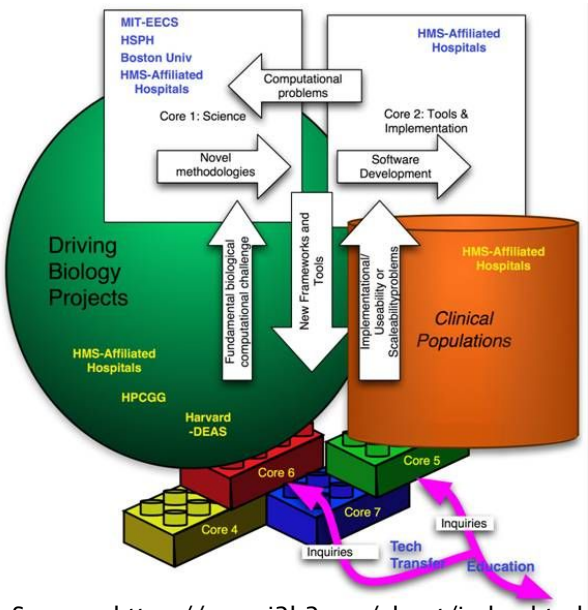
ETL - Load

Load data into the EDS information model

- Trade of
    - Update data (add, remove, update)
    - Drop and replace

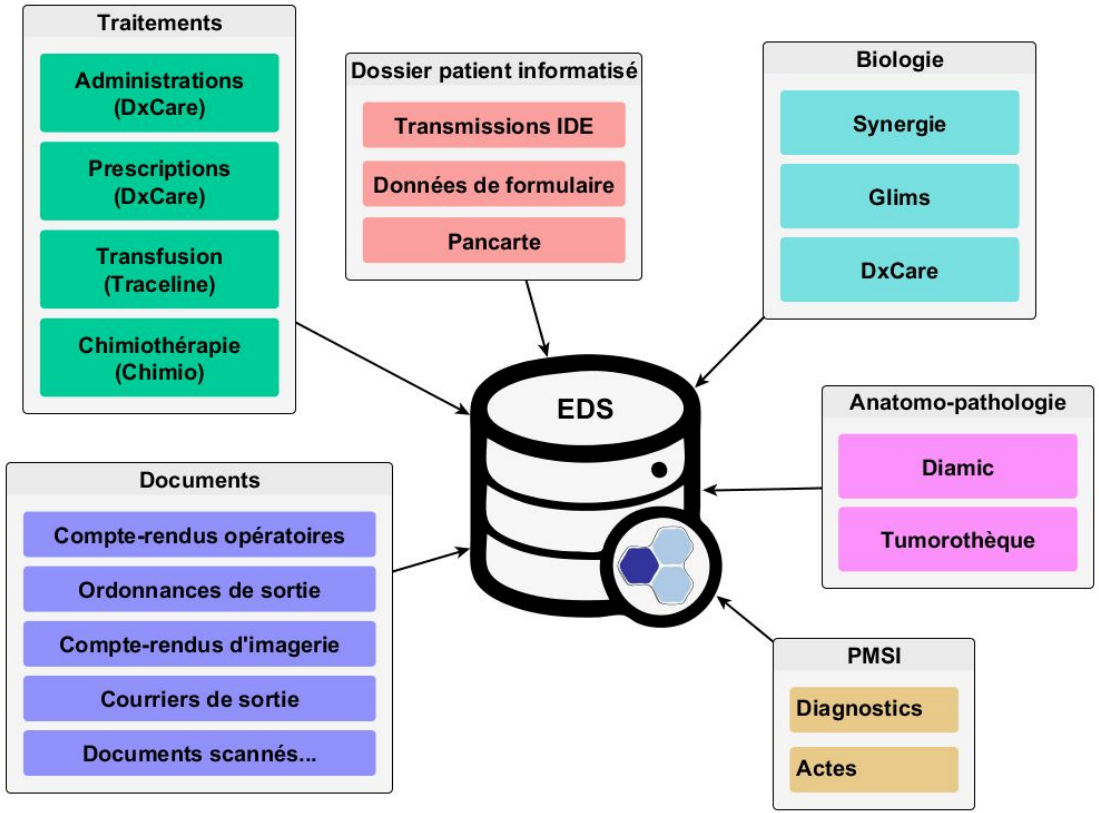
Example I2b2

- NIH (national institute of health) funded
- NCBC (national Center for Biomedical Computing)
  - Director Isaac Kohane
  - Developed since 2004
- Open source
- Aims at enabling translational research



Source : https://www.i2b2.org/about/index.html

EDS : Data integrated



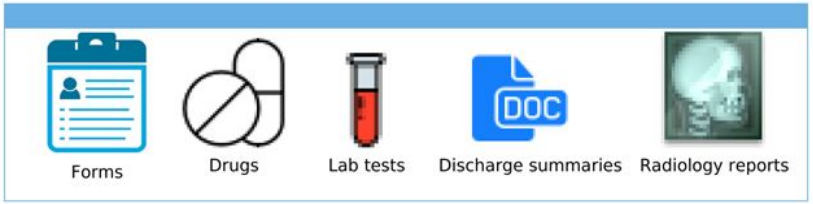
EDS : Data size

1 650 454

Patients

12 098 270

Venues

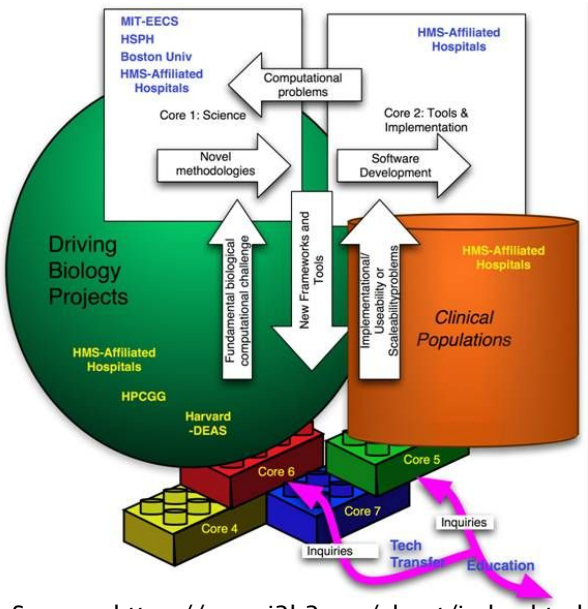


1 237 180 900

Observations

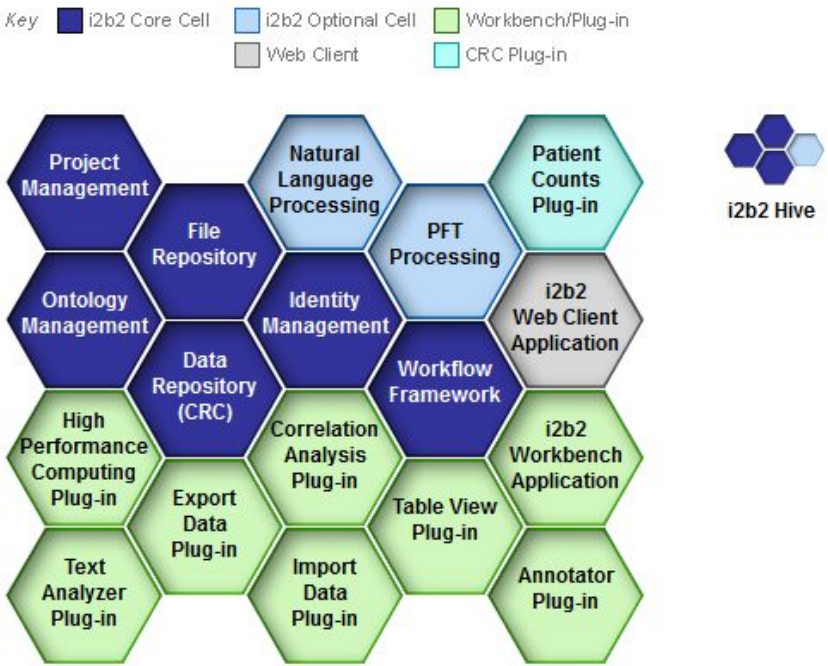
Exemple I2b2

- Sponsored by the NIH
- NCBC (national Center for Biomedical Computing)
  - Directeur Isaac Kohane
  - Developed since 2004 (Harvard)
- Translational research



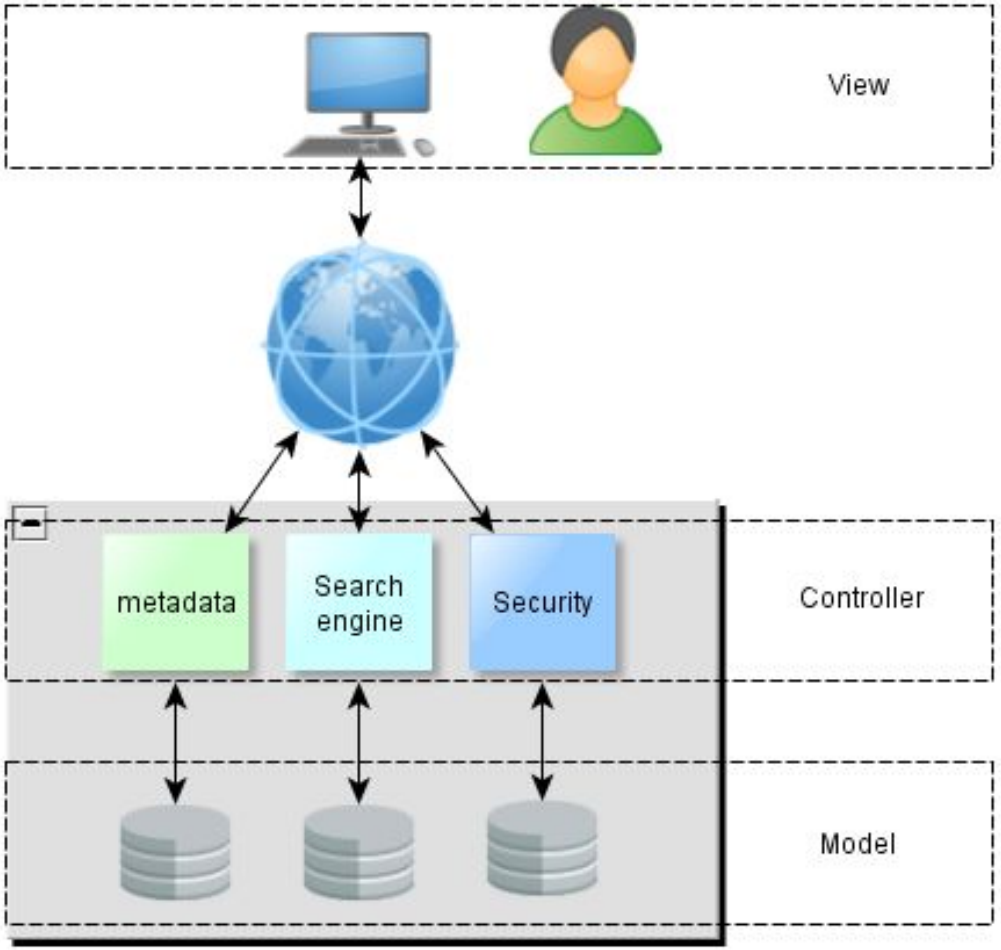
Source : https://www.i2b2.org/about/index.html

I2b2 hive



Sources : https://www.i2b2.org/software/index.html

I2B2 Architecture

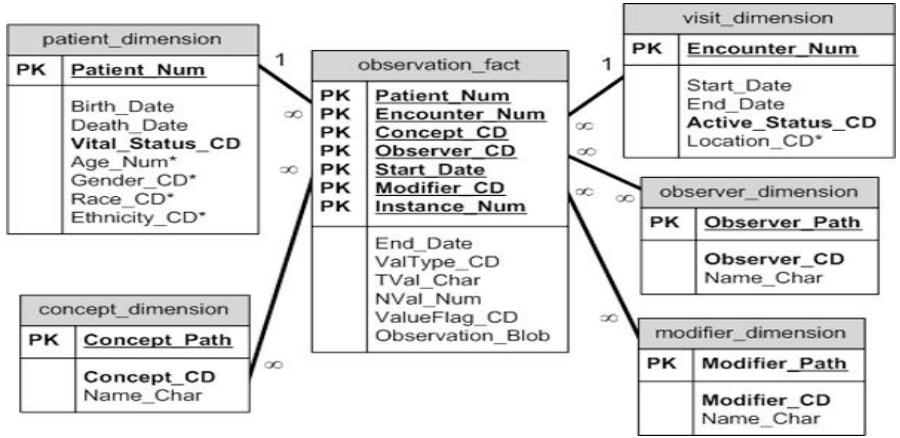


I2b2 data integration

- Can integrate
  - Clinical data
  - Genomic data
  - Biological data
- Handles data heterogeneity
  - Syntaxique
  - Sémantic
  - Unit of measure

I2b2 – Modèle de données

- Denormalized model
- 5 Tables (patient, venue, provider, observation, concept, modifier)
- All data format
- Data and semantics are seperated



Patient_dimension

PATIENT_NUM	Encoded i2b2 patient number	100001
VITAL_STATUS_CD		
BIRTH_DATE		
SEX_CD		
...	...	...

Visit_dimension

ENCOUNTER_NUM	Encoded i2b2 patient visit number	10001
PATIENT_NUM	Encoded i2b2 patient number	100001
START_DATE		
END_DATE		
...	...	...

Observation Fact

ENCOUNTER_NUM	Encoded i2b2 patient visit number	10001
PATIENT_NUM	Encoded i2b2 patient number	100001
CONCEPT_CD	Code for the observation of interest (i.e. diagnoses, procedures, medications, lab tests)	ICD10:C50.1
PROVIDER_ID	Practitioner or provider id	1254
START_DATE	Starting date-time of the observation (mm/dd/yyyy)	25/01/2019
VALTYPE_CD	Format of the concept	N = Numeric T = Text (enums / short messages) B = Raw Text (notes / reports) NLP = NLP result text

Integration example

Patient ID	DDN	Sexe	diag	type_diag	acte	id_sejour	DateEntrée
1	20/05/1923	F	Bladder k	Principal	Surgery	112006	08/07/2009
1	20/05/1923	F	Bladder k	Complemen	Sample	125485	10/11/2009
2	11/08/1943	F	Breast k	…	RMI	113005	…

Patient dimension

Patient ID	DDN	Sexe	diag	type_diag	acte	id_sejour	DateEntrée
1	20/05/1923	F	Bladder k	Principal	Surgery	112006	08/07/2009
1	20/05/1923	F	Bladder k	Complemen	Sample	125485	10/11/2009
2	11/08/1943	F	Breast k	…	RMI	113005	…

PATIENT_NUM	BRITH_DATE	SEXE_CD
1	20/05/1923	F
2	11/08/1943	F

Visit dimension

Patient ID	DDN	Sexe	diag	type_diag	acte	id_sejour	DateEntrée
1	20/05/1923	F	Bladder k	Principal	Surgery	112006	08/07/2009
1	20/05/1923	F	Bladder k	Complemen	Sample	125485	10/11/2009
2	11/08/1943	F	Breast k	…	RMI	113005	…

ENCOUNTER_NUM	PATIENT_NUM	START_DATE
1	1	08/07/2009
2	1	10/11/2009
1	2	...

Observation_fact

Patient ID	DDN	Sexe	diag	type_diag	acte	id_sejour	DateEntrée
1	20/05/1923	F	Bladder k	Principal	Surgery	112006	08/07/2009
1	20/05/1923	F	Bladder k	Complemen	Sample	125485	10/11/2009
2	11/08/1943	F	Breast k	…	RMI	113005	…

ENCOUNTER_NUM	PATIENT_NUM	START_DATE	CONCEPT_CD	...
1	1	08/07/2009	Bladder_principal	
1	1	08/07/2009	surgery	
2	1	10/11/2009	Bladder complementary	
1	2	...	Breast	