

The Georges Pompidou University Hospital i2b2 Warehouse: learning and feedback from 8-years' experience



*Anne-Sophie Jannot, Eric Zapletal, Paul Avillach
Marie-France Mamzer, Anita Burgun, Patrice Degoulet*



UNIVERSITÉ
**PARIS
DESCARTES**



ASSISTANCE
PUBLIQUE  HÔPITAUX
DE PARIS

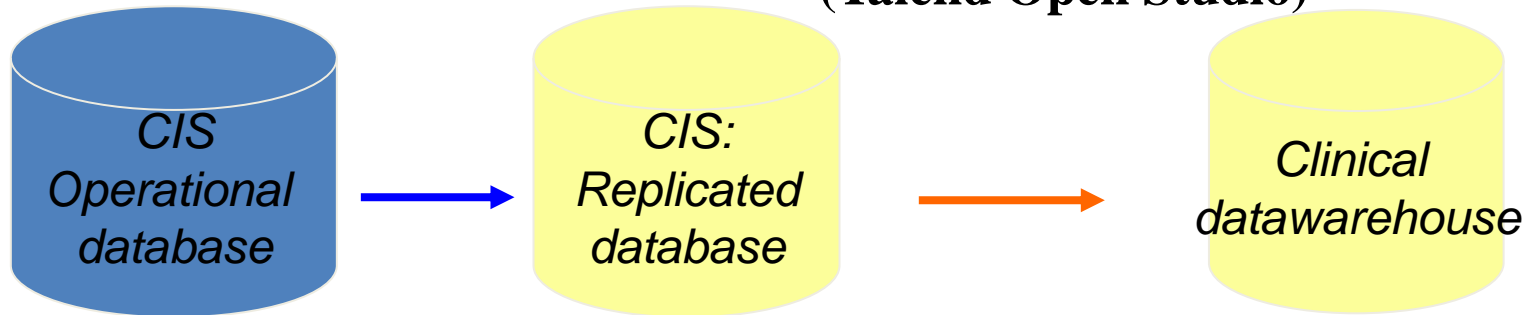


CDW challenges

- Data integration
- Data gouvernance
- Data valorisation / processing

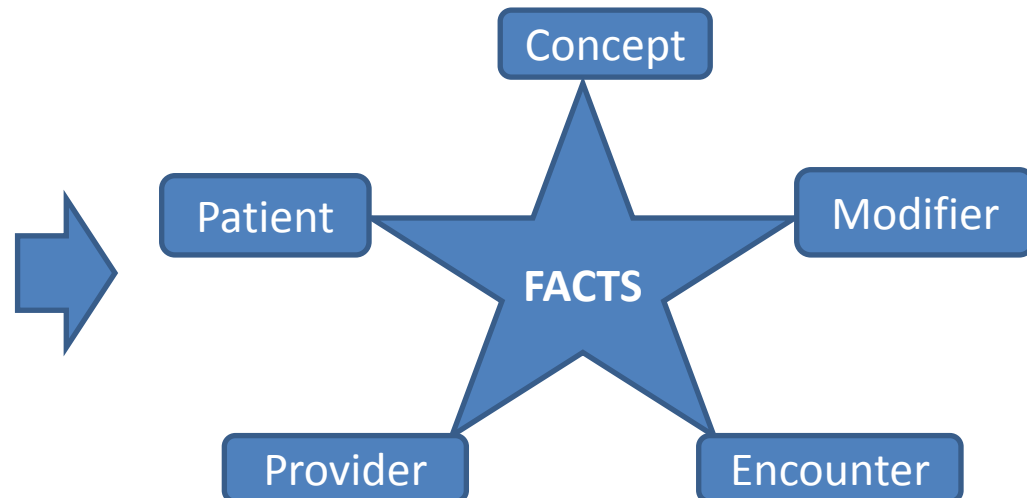
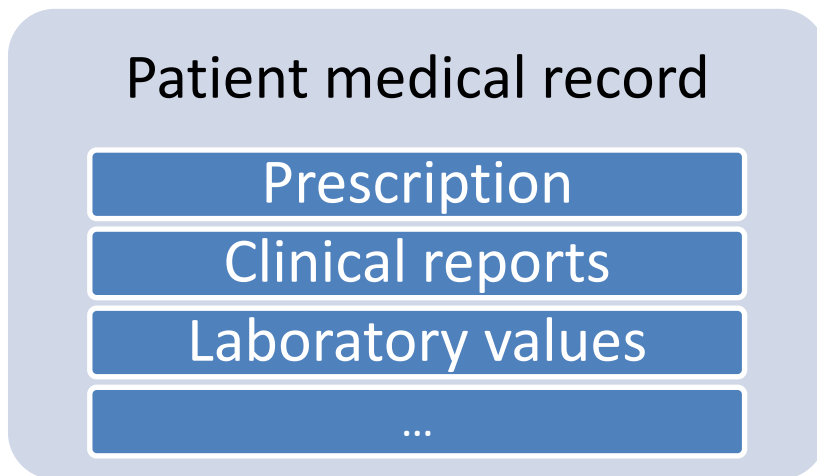
CDW integration in HEGP hospital (2009-)

Extract / Transform / Load
(Talend Open Studio)



i2b2

Informatics for Integrating Biology & the Bedside



HEGP CDW sources and volumetry (current figures)

HEGP CDW data sources	Number of observations (millions)
Demographic patient data (gender, age, family status and vital status)	1.6
Biology and pathology results	136.7
Drug prescriptions and pharmacists validations	11.4
Final diagnosis	8.4
Procedures	5.4
Diagnoses and procedures coded using legacy terminologies (DRG)	1.2
Patient pathways	7.1
Electronic Health Record (EHR) structured forms items	130.6
Clinical and imagery text reports	4.1
Clinical parameters (heart frequency, temperature, ...)	14.2
Chemotherapy prescriptions	1.4
Nurses notes	18.9
Radiation therapy data	< 0.1

Data governance (2011-): patients' information

■ ■ ■ ■ ■ Grâce à vos données informatisées, vous êtes co-acteur de la recherche en santé

Les Hôpitaux Universitaires Paris Ouest sont rattachés à la faculté de médecine Paris Descartes. Ils assurent des missions de soins, d'enseignement et de recherche. Vos données médicales informatisées, comme celles des autres patients, peuvent être réutilisées dans un but de recherche en santé. Ce sont des recherches dites rétrospectives car elles utilisent a posteriori les données recueillies pour vos soins. Elles ne nécessitent aucune participation de votre part. Pour garantir l'absence de mésusage de la réutilisation des données et organiser ces recherches, en particulier la constitution d'un Comité d'éthique qui peut donner un avis sur ces recherches.

Institutional review board

Les recherches sur données sont menées en parallèle aux recherches biomédicales classiques. Elles peuvent être réalisées par tout professionnel de santé hospitalier, y compris ceux qui ne vous suivent pas.

Patients involvement

Vous pouvez vous-même proposer des idées de recherche en santé en vous adressant au 01 56 09 20 30 ou par mél à

Right to refuse

Vous avez le droit de vous opposer à la réutilisation pour la recherche de vos données de soins, à tout moment et sans vous en justifier, en vous adressant à la Direction de l'Informatique Hospitalière au 01 56 09 20 30 ou mesdonnees.info@egp.aphp.fr.

Cette opposition de votre part n'affectera en rien votre prise en charge à l'hôpital, la qualité des soins qui vous seront fournis, ni la relation avec votre médecin référent.

Pour en savoir plus : www.cnil.fr et www.aphp.fr (<http://www.aphp.fr/patient/je-connais-mes-droits/>)

Data governance (2011-): definition of several levels of data access

- First level : aggregated data
 - All physicians and pharmacist from HEGP having participated to a training session on i2b2
 - Training sessions opened to all physicians and pharmacist including residents

Data governance (2011-): aggregated data access

The screenshot shows the i2b2 Query & Analysis Tool web interface. The browser address bar displays <http://cdw.egp.aphp.fr/default.htm>. The page title is "i2b2 Query & Analysis Tool". The user is identified as "User: Anne-Sophie Jannot".

The interface is divided into several sections:

- Navigate Terms / Find Terms:** A list of medical terms is displayed, including "10. Démographie", "20. Biologie", "30. Médicaments", "40. Diagnostics (CIM10)", "50. Actes (CCAM)", "60. Groupe HEGP-CCL-VGR", "65. Fournisseurs HEGP-CCL-VGR", "80. Questions (DxCare)", "90. Type de dossier", "92. Comptes rendus cliniques", "94. Comptes rendus d'exams complémentaires", "96. Projets", "98. Bientôt disponible", and "99. Divers". The term "40. Diagnostics (CIM10)" is highlighted.
- Workplace:** A section for managing queries, currently showing a single query with ID "3252998".
- Query Tool:** The main area for constructing queries. It includes a "Query Name" field, a "Temporal Constraint" dropdown set to "Treat all groups independently", and three columns for defining query groups (Group 1, Group 2, and Group 3). Each column has headers for "Dates", "Occurs > 0x", and "Exclude". Below these headers are large empty boxes for defining the query logic. A yellow box in the first group area contains the text "drop a term on here".

At the bottom of the interface, there are buttons for "Run Query", "Clear", and "Print Query", along with a status indicator showing "0 Groups" and a "New Group" button.

Data governance (2011-): aggregated data access

The screenshot displays the i2b2 Web Client interface. The browser address bar shows <http://cdw.egp.aphp.fr/default.htm>. The interface includes a menu bar with 'Fichier', 'Edition', 'Affichage', 'Favoris', and 'Outils'. Below the menu bar, there are tabs for 'Navigate Terms' and 'Find Terms'. The 'Navigate Terms' pane on the left shows a hierarchical list of medical categories, with '40. Diagnostics (CIM10)' selected. The 'Find Terms' pane on the right shows a search results table with columns for 'Group 2', 'Group 3', 'Occurs > 0x', 'Exclude', and 'Dates'. The table is currently empty, and a yellow box with the text 'drop a term on here' is visible in the center of the workspace area. The bottom of the interface features buttons for 'Run Query', 'Clear', and 'Print Query', along with a status bar indicating '0 Groups' and a 'New Group' button.

http://cdw.egp.aphp.fr/default.htm

i2b2 Web Client

Fichier Edition Affichage Favoris Outils ?

Sites suggérés

izb2 Query

Navigate Terms Find Terms

40. Diagnostics (CIM10)

- (A00-B99) Certaines maladies infectieuses et parasitaires
- (C00-D48) Tumeurs
- (D50-D89) Maladies du sang et des organes hématopoïétiques
- (E00-E90) Maladies endocriniennes, nutritionnelles et métaboliques
- (F00-F99) Troubles mentaux et du comportement
- (G00-G99) Maladies du système nerveux
- (H00-H59) Maladies de l'oeil et de ses annexes
- (H60-H95) Maladies de l'oreille et de l'apophyse mastoïde
- (I00-I99) Maladies de l'appareil circulatoire
 - (I00-I02) Rhumatisme articulaire aigu
 - (I05-I09) Cardiopathies rhumatismales chroniques
 - (I10-I15) Maladies hypertensives
 - (I20-I25) Cardiopathies ischémiques

Workplace

3252998

drop a term on here

Run Query Clear Print Query

0 Groups

New Group

Sophie Jannot Find Patients | Analysis Tools | Help | Logout

Treat all groups independently

Group 2	Group 3
Occurs > 0x	Occurs > 0x
Exclude	Exclude
independently	Treat Independently

Data governance (2011-): aggregated data access

The screenshot displays the i2b2 Web Client interface, which is used for managing and querying aggregated data. The browser address bar shows the URL <http://cdw.egp.aphp.fr/default.htm>. The interface includes a menu bar with options like 'Fichier', 'Edition', 'Affichage', 'Favoris', and 'Outils'. Below this, there are tabs for 'Navigate Terms' and 'Find Terms'. The main content area shows a hierarchical tree of medical terms, with the following structure:

- 40. Diagnostics (CIM10)
 - (A00-B99) Certaines maladies infectieuses et parasitaires
 - (C00-D48) Tumeurs
 - (D50-D89) Maladies
 - (E00-E90) Maladies
 - (F00-F99) Troubles
 - (G00-G99) Maladies
 - (H00-H59) Maladies
 - (H60-H95) Maladies
 - (I00-I99) Maladies de l'appareil circulatoire
 - (I00-I02) Rhumatisme articulaire aigu
 - (I05-I09) Cardiopathies rhumatismales chroniques
 - (I10-I15) Maladies hypertensives
 - (I20-I25) Cardiopathies ischémiques
 - (I26-I28) Affections cardio-pulmonaires et maladies de la circulation
 - (I30-I52) Autres formes de cardiopathies
 - (I60-I69) Maladies cérébrovasculaires
 - (I60) Hémorragie sous-arachnoïdienne
 - (I61) Hémorragie intracérébrale
 - (I62) Autres hémorragies intracrâniennes non traumatiques
 - (I63) Infarctus cérébral
 - (I64) Accident vasculaire cérébral, non précisé comme étant hémorragique ou ischémique
 - (I65) Occlusion et sténose des artères précérébrales, non

The interface also includes a 'Workplace' section on the left with a file named '3252998'. At the bottom, there are buttons for 'Run Query', 'Clear', and 'Print Query', along with a status bar indicating '0 Groups' and a 'New Group' button.

Data governance (2011-): definition of several levels of data access

- First level : aggregated data
 - All physicians and pharmacist from HEGP having participated to a training session on i2b2
 - Training sessions opened to all physicians and pharmacist including residents
- Level 2 and 3 : individual data (IRB)
 - Level 2 : encoded identifiers
 - Level 3: hospital identifier or complete name

Data governance (2011-): access to individual data for a research project

Research protocol

IRB assessment

Research datamart

Analysis: vizualisation tool, statistical support

Are all units having produced relevant data for the project involved in the research?

FastVisu

Select a serie: JBE_MC5 - 0.catalogue autoimmun v3.1 (847) | Select a patient: 6) 2005648058 (1) | Select a regex catalog: 0.catalogue autoimmun v3.1

Select a poll: JBE_MC5 Aiv3 | Select allowed vote values: OUI/NON | Select a voter: JBE (72)

Show categories without match

<< R >>

Elle ne décrit aucun symptôme en particulier pas de trouble digestif à type de diarrhée, pas de perte de poids, pas de douleur articulaire, pas d'aphtose et a quelques troubles du sommeil. Elle se plaint essentiellement de problèmes ORL récidivants et a bénéficié d'une amygdalotomie à l'âge de 30 ans. Elle a toujours des sinusites à répétition. Elle a une fille de 8 ans qu'elle n'a pas dépisté et je lui conseille de le faire. Elle n'a pas d'antécédents familiaux très spécifiques dans le cadre des maladies auto-immunes.

Diabetes Mellitus, Type 1 ☒ Oui ☐ Non

thyroïdiennes, diabète par exemple s thyroidiennes, diabète par exemple s

Thyroiditis, Autoimmune ☐ Oui ☐ Non

et d'une TSH pour rechercher une éventuelle thyroïdite auto-immune ass et d'une TSH pour rechercher une éventuelle thyroïdite auto-immune ass

rechercher une éventuelle thyroïdite auto-immune ass

ylroglibilines et d'une TS ylroglibilines et d'une TS ylroglibilines et d'une TS

existe également un risque statistique de développer d'autres maladies en particulier des maladies thyroïdiennes, diabète par exemple si le régime n'est pas suivi.

doute un peu délicat à adapter du fait de sa profession d'organisatrice d'évènements.

Je lui conseille de réaliser un nouveau bilan après un an du régime sans gluten afin de savoir si les écarts qu'elle pourrait réaliser ont empêché la repousse villositaire. Si cela est le cas il faudra intensifier le régime sans gluten et le rendre le plus strict possible.

Dans l'immédiat, il serait souhaitable de disposer au moins d'anticorps anti-thyroglibilines et d'une TSH pour rechercher une éventuelle thyroïdite auto-immune associée.

Je suis prêt à la revoir si tu le juges nécessaire

Bien amicalement.

Professeur Christophe CELLIER

20, rue Leblanc
75908 PARIS Cedex 15

Tél. : 01 56 09 20 00
Fax : 01 56 09 21 18

En attente de egp-svlwarp2.egp.aphp.fr...

Escudié J-B, Jannot A-S, Zapletal E, Cohen S, Malamut G, Burgun A, et al. Reviewing 741 patients records in two hours with FASTVISU. AMIA Annu Symp. Proc AMIA Symp AMIA Symp. 2015; 2015: 553–9.

Data valorisation (2011-): projects approved by local IRB

Year	Number of projects	Number of bioclinical departments	Clinical epidemiology	Health service research	Clinical Research
2011	13	5	8	5	0
2012	4	4	1	3	0
2013	13	10	8	4	1
2014	22	11	14	5	3
2015	22	10	9	13	0
2016	21	10	6	15	0
Total (%)	95	19 (79%)	46 (49%)	45 (47%)	4 (4%)

Data valorisation (2011-): an epidemiological study on a rare condition

Methods

Matched case control
study

Large inclusion
period

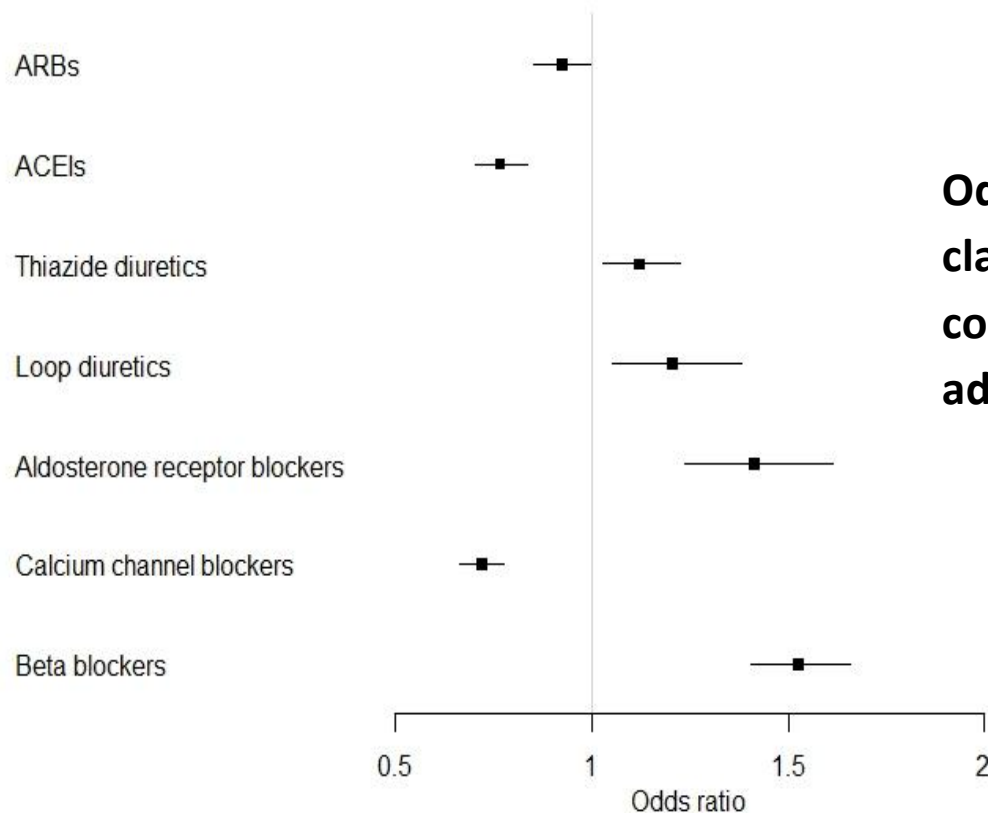
We conducted a retrospective monocentric study of RTR diagnosed with PCa after transplantation between 2000 and 2013. Demographics, PCa characteristics and treatment were assessed. For each RTR in radical prostatectomy (RP) subset, we included 4 non-organ transplant patients who underwent RP by the same surgeons, and compared pre-operative and post-operative oncological outcomes and biochemical recurrence (BCR) rate.

Rare disease

Twenty-four RTR were included (PCa incidence 1.5%). Mean follow-up was 47 months. PCa was mostly localized (n=21, 87.5%) with treatments including RP (n=16, 76.2%), brachytherapy (n=3, 14.3%), radiation therapy (n=1, 4.7%) and active surveillance (n=1, 4.7%). No graft loss due to PCa treatment was reported. Nineteen RTR with localized PCa (90.5%) were free from BCR. Considering RP subset, no difference in PCa characteristics at diagnosis and BCR rate was found between RTR (n=16) and control patients (n=64).

Pettenati, C., Jannot, A. S., Hurel, S., Verkarre, V., Kreis, H., Housset, M., ... & Timsit, M. O. (2016). Prostate Cancer Characteristics and Outcome in Renal Transplant Recipients: Results from a Contemporary Monocentric Study. Clinical transplantation.

Data valorisation (2011-): an health service research on prescription practice evaluation



Odds of women being treated by a given class of anti-hypertensive drug at the first consultation compared to men after adjustment over the 2000-2015 period

Reuse of data from hypertension consultation

Deborde, T., Amar, L., Bobrie, G., Postel-Vinay N, Battaglia C, Tache A, Chedid A, Dhib MM, Chatellier, G., Plouin PF, Burgun, A., Azizi M, Jannot, AS. Sex differences in antihypertensive treatment in France among 17,856 patients in a tertiary hypertension unit. Journal of hypertension, 2017.

Data valorisation (2011-): a clinical research project on patient recruitment

Methods

We selected three institutional studies covering various medical domains. The studies included a total of 67 inclusion and exclusion criteria and ran in two University Hospitals. We described the steps required to use the EHR4CR platform for a feasibility study. We also defined metrics to assess each of the steps (including criteria complexity, normalization quality, and data completeness of EHRs).

Results

We identified 114 distinct medical concepts from a total of 67 eligibility criteria. Among the 114 concepts: 23 (20.2%) corresponded to non-structured data (i.e. for which transformation is needed before analysis), 92 (81%) could be mapped to terminologies used in EHR4CR, and 86 (75%) could be mapped to local terminologies. We identified 51 computable criteria following the normalization process. The normalization was considered by experts to be satisfactory or higher for 64.2% (43/67) of the computable criteria. All of the computable criteria could be expressed using the EHR4CR platform.

Girardeau, Y., Doods, J., Zapletal, E., Chatellier, G., Daniel, C., Burgun, A., ... & Rance, B. (2017). Leveraging the EHR4CR platform to support patient inclusion in academic studies: challenges and lessons learned. BMC medical research methodology, 17(1), 36.

Learning and feedback

- Data governance is central
- Clinical researchers often ignore clinical data complexity (thousands of concepts)
- Clinicians not prepared to formulate complex queries and navigate through the different nomenclatures that populate a CDW.
- Strong collaboration between clinicians, biomedical informatics, biostatistics and epidemiology specialists is needed to complete successfully research project using a CDW

Perspectives

- Link with other datawarehouse (AP-HP, SNIIRAM)
- Integration of other types of data (omics, patient reported outcomes)

The Georges Pompidou University Hospital i2b2 Warehouse: learning and feedback from 8-years' experience



*Anne-Sophie Jannot, Eric Zapletal, Paul Avillach
Marie-France Mamzer, Anita Burgun, Patrice Degoulet*



UNIVERSITÉ
**PARIS
DESCARTES**



ASSISTANCE
PUBLIQUE  HÔPITAUX
DE PARIS

