Recherche dans un entrepôt de données de santé

Bastien Rance

Université Paris Descartes - Faculté de médecine
AP-HP ; Hôpital Européen Georges Pompidou
INSERM ; Centre de Recherche des Cordeliers ; Equipe Burgun

Avertissement

Cette présentation n’engage que mon point de vue et ne reflète pas la position de l’AP-HP ou d’une autre institution

Hôpital Européen Georges Pompidou (HEGP)

Ouverture en 2000

HIMMS niveau 6

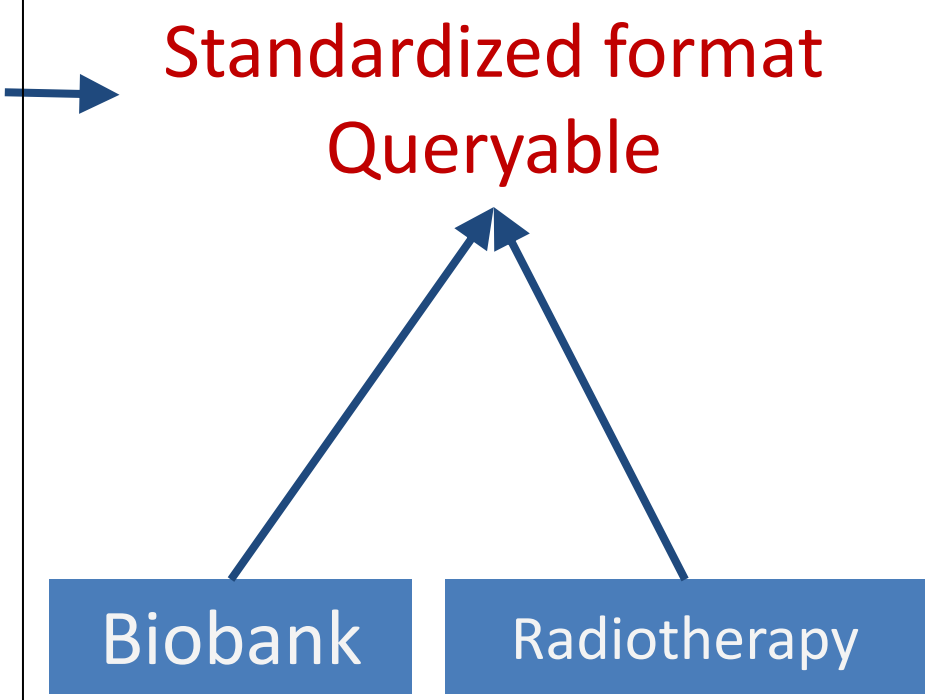
(http://www.himss.eu/node/1116)

Clinical Data Warehouse

Electronic Health Record (EHR)

Diagnosis
Clinical items
Billing codes
Biology (lab)
Nurse transmission
Imaging reports
Pathology reports
Drug prescription
Chemotherapy

Clinical Data Warehouse (CDW)



Clinical Data Warehouse at HEGP

Concept	# observations
EHR concepts	176,524,243
Biology (Laboratory)	**155,173,140**
Nursing transmission	22,001,856
Billing (disease) codes	4,846,602
Rx prescription	91,251,062
Text reports	5,343,224
Imaging reports	1,650,027
Pathology codes	1,434,803

Unstructured data: transformation is needed before reuse

RGPD

RGPD

Pourquoi les entrepôts ?

- Données de “**vraie vie**”
  - Populations différentes des essais cliniques
  - Populations agées ou pédiatriques
  - Des problèmes complexes et combinés
- Des données **exhaustives**
  - Nombreuses modalités
  - ***Contextualisées***
- Granularité fine
- …

Cas d'usage : Phénotypage à haut débit

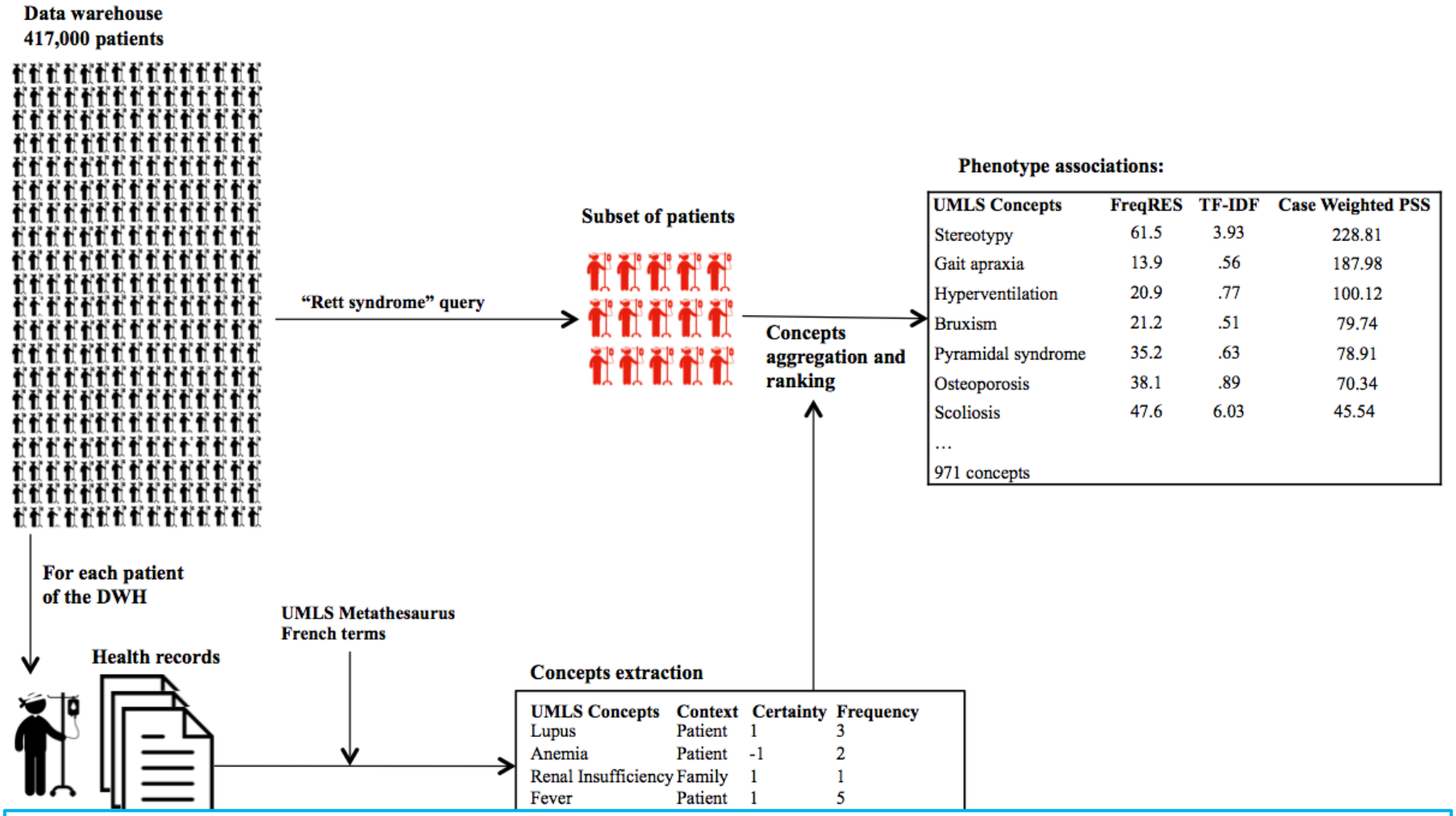
Research | Open Access

Next generation phenotyping using narrative reports in a rare disease clinical data warehouse

Nicolas Garcelon, Antoine Neuraz, Rémi Salomon, Nadia Bahi-Buisson, Jeanne Amiel, Capucine Picard, Nizar Mahlaoui, Vincent Benoit, Anita Burgun and Bastien Rance

Orphanet Journal of Rare Diseases 2018 **13**:85
https://doi.org/10.1186/s13023-018-0830-6
Received: 4 October 2017 | **Accepted:** 23 May 2018 | **Published:** 31 May 2018

Cas d'usage : Phénotypage à haut débit

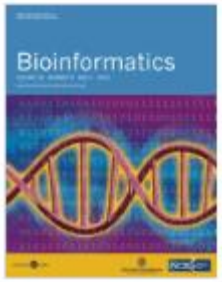


Next generation phenotyping using narrative reports in a rare disease clinical data warehouse. Garcelon et al. *Orphanet Journal of Rare Diseases*

Cas d'usage : Phénotypage à haut débit

Next generation phenotyping using narrative reports in a rare disease clinical data warehouse. Garcelon et al. *Orphanet Journal of Rare Diseases*

Cas d'usage : PheWAS

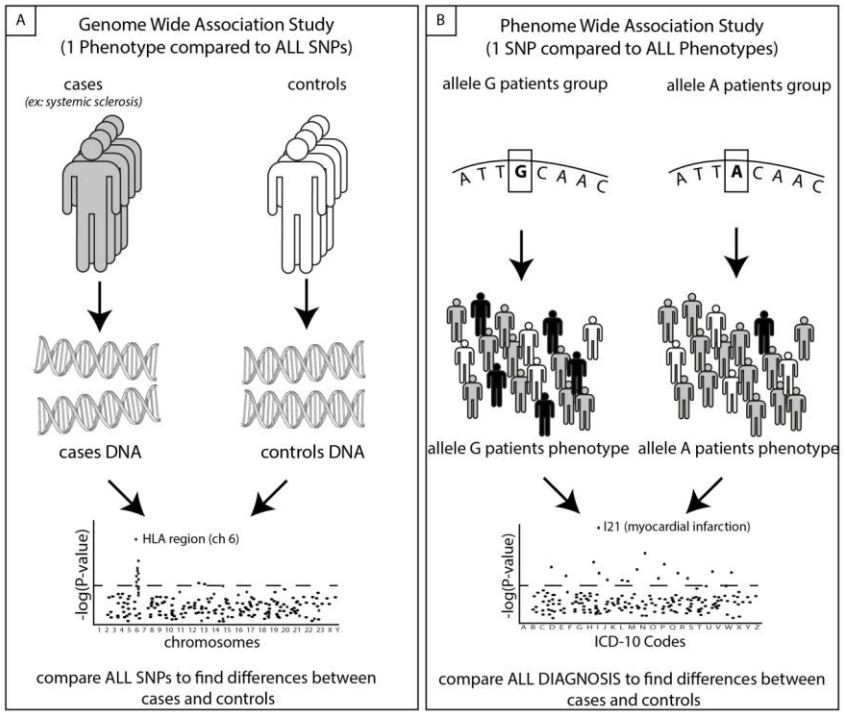


Volume 26, Issue 9
1 May 2010

PheWAS: demonstrating the feasibility of a phenome–wide scan to discover gene–disease associations

Joshua C. Denny, Marylyn D. Ritchie, Melissa A. Basford, Jill M. Pulley, Lisa Bastarache, Kristin Brown-Gentry, Deede Wang, Dan R. Masys, Dan M. Roden, Dana C. Crawford

Bioinformatics, Volume 26, Issue 9, 1 May 2010, Pages 1205–1210,
https://doi.org/10.1093/bioinformatics/btq126
Published: 24 March 2010 **Article history**



McCarthy et al, Nature Reviews Genetics, 2008 / Denny et al, Bioinformatics 2010

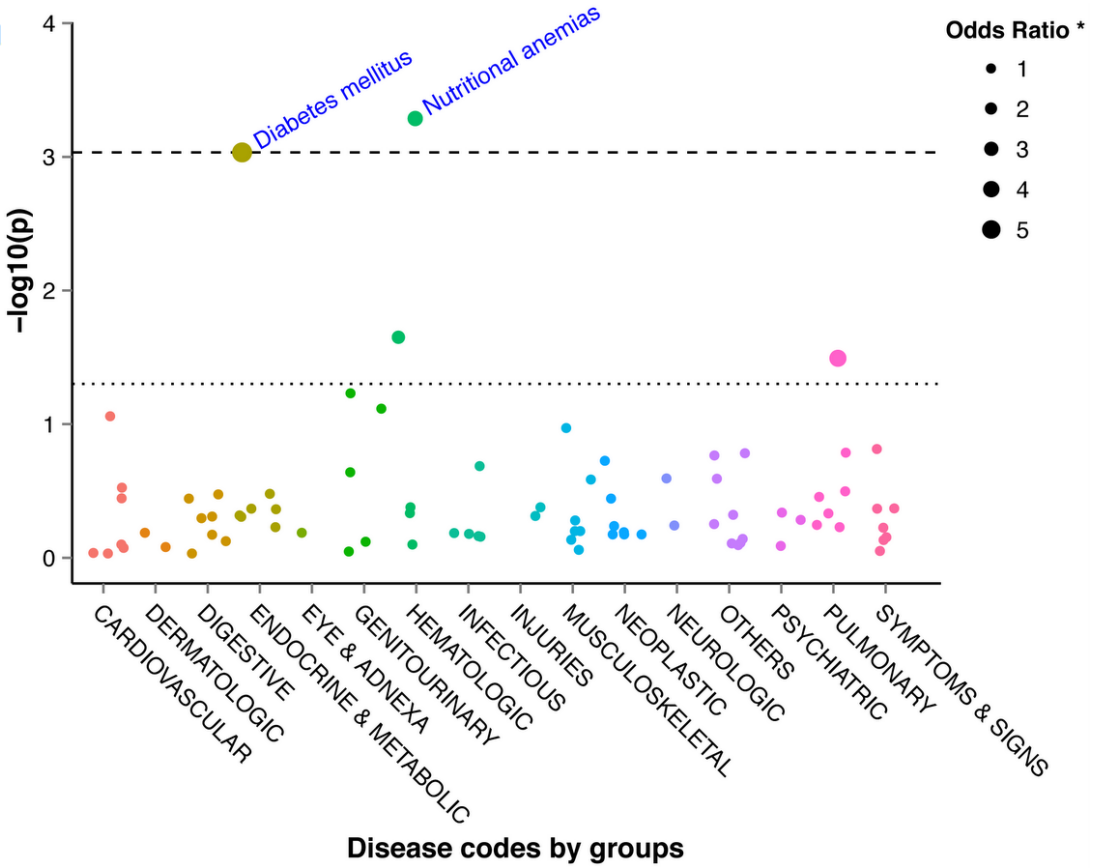
Cas d'usage : PheWAS

Phenome-Wide Association Studies on a Quantitative Trait: Application to TPMT Enzyme Activity and Thiopurine Therapy in Pharmacogenomics

Antoine Neuraz, Laurent Chouchana, Georgia Malamut, Christine Le Beller, Denis Roche, Philippe Beaune, Patrice Degoulet, Anita Burgun, Marie-Anne Loriot, Paul Avillach

Published: December 26, 2013 • https://doi.org/10.1371/journal

PLOS | COMPUTATIONAL BIOLOGY



Cas d'usage : Études cliniques diverses

Journal of Biomedical Informatics 52 (2014) 28–35

Contents lists available at ScienceDirect

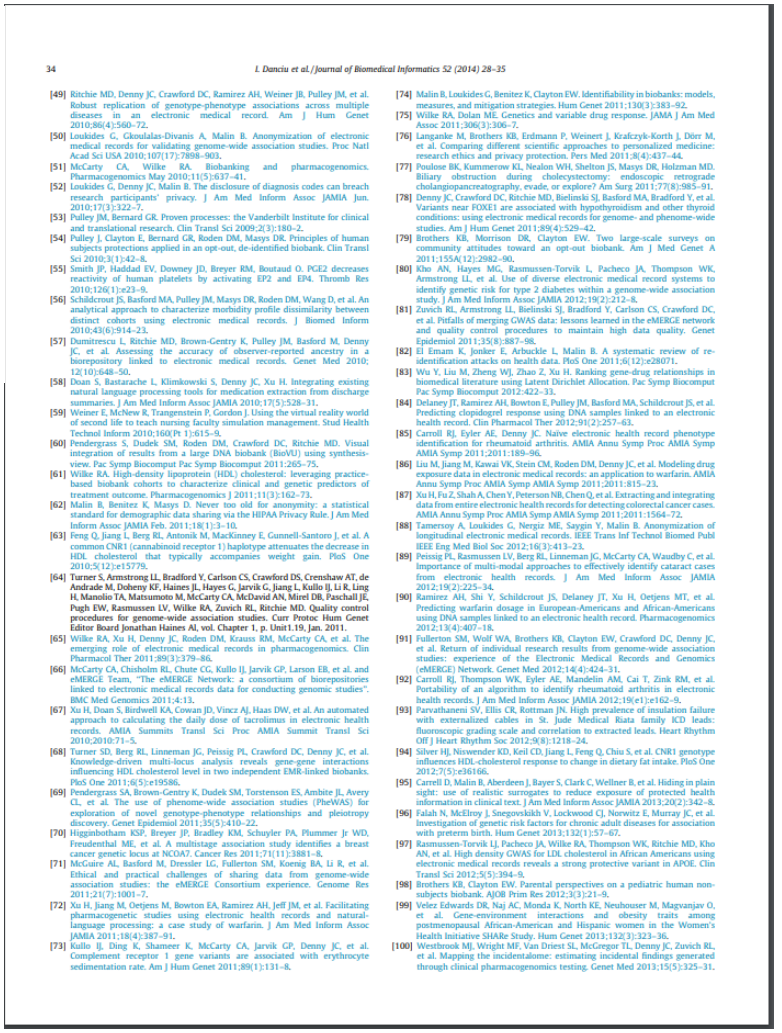
Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Secondary use of clinical data: The Vanderbilt approach

Ioana Danciu [a,b,*], James D. Cowan [a], Melissa Basford [a], Xiaoming Wang [a], Alexander Saip [a], Susan Osgood [a], Jana Shirey-Rice [a], Jacqueline Kirby [a], Paul A. Harris [a,b]

Since its inception, the SD has provided infrastructure and support for various research projects that resulted in **87 papers published between 2010 and 2013** [21,23,26,27,32–37,39–42,45–115], with 61 first authors from 18 different departments



34 I. Danciu et al. / Journal of Biomedical Informatics 52 (2014) 28–35

[49] Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. Am J Hum Genet 2010;86(4):560–72.
[50] Loukides G, Gkoulalas-Divanis A, Malin B. Anonymization of electronic medical records for validating genome-wide association studies. Proc Natl Acad Sci USA 2010;107(17):7898–903.
[51] McCarty CA, Wilke RA, Biobanking and pharmacogenomics. Pharmacogenomics May 2010;11(5):637–41.
[52] Loukides G, Denny JC, Malin B. The disclosure of diagnosis codes can breach research participants' privacy. J Am Med Inform Assoc JAMIA Jun. 2010;17(3):322–7.
[53] Pulley JM, Bernard GR. Proven processes: the Vanderbilt Institute for clinical and translational research. Clin Transl Sci 2009;2(3):180–2.
[54] Pulley J, Clayton E, Bernard GR, Roden DM, Masys DR. Principles of human subjects protections applied in an opt-out, de-identified biobank. Clin Transl Sci 2010;3(1):42–8.
[55] Smith JP, Haddad EV, Downey JD, Breyer RM, Boutaud O. PGE2 decreases reactivity of human platelets by activating EP2 and EP4. Thromb Res 2010;126(1):e23–9.
[56] Schildcrout JS, Basford MA, Pulley JM, Masys DR, Roden DM, Wang D, et al. An analytical approach to characterize morbidity profile dissimilarity between distinct cohorts using electronic medical records. J Biomed Inform 2010;43(6):914–23.
[57] Dumitrescu L, Ritchie MD, Brown-Gentry K, Pulley JM, Basford M, Denny JC, et al. Assessing the accuracy of observer-reported ancestry in a biorepository linked to electronic medical records. Genet Med 2010;12(10):648–50.
[58] Doan S, Bastarache L, Klimkowski S, Denny JC, Xu H. Integrating existing natural language processing tools for medication extraction from discharge summaries. J Am Med Inform Assoc JAMIA 2010;17(5):528–31.
[59] Weiner E, McNew R, Trangenstein P, Gordon J. Using the virtual reality world of second life to teach nursing faculty simulation management. Stud Health Technol Inform 2010;160(Pt 1):615–9.
[60] Pendergrass S, Dudek SM, Roden DM, Crawford DC, Ritchie MD. Visual integration of results from a large DNA biobank (BioVU) using synthesis-view. Pac Symp Biocomput Pac Symp Biocomput 2011:265–75.
[61] Wilke RA. High-density lipoprotein (HDL) cholesterol: leveraging practice-based biobank cohorts to characterize clinical and genetic predictors of treatment outcome. Pharmacogenomics J 2011;11(3):162–73.
[62] Malin B, Benitez K, Masys D. Never too old for anonymity: a statistical standard for demographic data sharing via the HIPAA Privacy Rule. J Am Med Inform Assoc JAMIA Feb. 2011;18(1):3–10.
[63] Feng Q, Jiang L, Berg RL, Antonik M, MacKinney E, Gunnell-Santoro J, et al. A common CNR1 (cannabinoid receptor 1) haplotype attenuates the decrease in HDL cholesterol that typically accompanies weight gain. PloS One 2010;5(12):e15779.
[64] Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DS, Crenshaw AT, de Andrade M, Doheny KF, Haines JL, Hayes G, Jarvik G, Jiang L, Kullo IJ, Li R, Ling H, Manolio TA, Matsumoto M, McCarty CA, McDavid AN, Mirel DB, Paschall JE, Pugh EW, Rasmussen LV, Wilke RA, Zuvich RL, Ritchie MD. Quality control procedures for genome-wide association studies. Curr Protoc Hum Genet Editor Board Jonathan Haines Al, vol. Chapter 1, p. Unit1.19, Jan. 2011.
[65] Wilke RA, Xu H, Denny JC, Roden DM, Krauss RM, McCarty CA, et al. The emerging role of electronic medical records in pharmacogenomics. Clin Pharmacol Ther 2011;89(3):379–86.
[66] McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, et al. and eMERGE Team, "The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies". BMC Med Genomics 2011;4:13.
[67] Xu H, Doan S, Birdwell KA, Cowan JD, Vincz AJ, Haas DW, et al. An automated approach to calculating the daily dose of tacrolimus in electronic health records. AMIA Summits Transl Sci Proc AMIA Summit Transl Sci 2010;2010:71–5.
[68] Turner SD, Berg RL, Linneman JG, Peissig PL, Crawford DC, Denny JC, et al. Knowledge-driven multi-locus analysis reveals gene-gene interactions influencing HDL cholesterol level in two independent EMR-linked biobanks. PloS One 2011;6(5):e19586.
[69] Pendergrass SA, Brown-Gentry K, Dudek SM, Torstenson ES, Ambite JL, Avery CL, et al. The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. Genet Epidemiol 2011;35(5):410–22.
[70] Higginbotham KSP, Breyer JP, Bradley KM, Schuyler PA, Plummer Jr WD, Freudenthal ME, et al. A multistage association study identifies a breast cancer genetic locus at NCOA7. Cancer Res 2011;71(11):3881–8.
[71] McGuire AL, Basford M, Dressler LG, Fullerton SM, Koenig BA, Li R, et al. Ethical and practical challenges of sharing data from genome-wide association studies: the eMERGE Consortium experience. Genome Res 2011;21(7):1001–7.
[72] Xu H, Jiang M, Oetjens M, Bowton EA, Ramirez AH, Jeff JM, et al. Facilitating pharmacogenetic studies using electronic health records and natural-language processing: a case study of warfarin. J Am Med Inform Assoc JAMIA 2011;18(4):387–91.
[73] Kullo IJ, Ding K, Shameer K, McCarty CA, Jarvik GP, Denny JC, et al. Complement receptor 1 gene variants are associated with erythrocyte sedimentation rate. Am J Hum Genet 2011;89(1):131–8.
[74] Malin B, Loukides G, Benitez K, Clayton EW. Identifiability in biobanks: models, measures, and mitigation strategies. Hum Genet 2011;130(3):383–92.
[75] Wilke RA, Dolan ME. Genetics and variable drug response. JAMA J Am Med Assoc 2011;306(3):306–7.
[76] Langanke M, Brothers KB, Erdmann P, Weinert J, Krafczyk-Korth J, Dörr M, et al. Comparing different scientific approaches to personalized medicine: research ethics and privacy protection. Pers Med 2011;8(4):437–44.
[77] Poulose BK, Kummerow KL, Nealon WH, Shelton JS, Masys DR, Holzman MD. Biliary obstruction during cholecystectomy: endoscopic retrograde cholangiopancreatography, evade, or explore? Am Surg 2011;77(8):985–91.
[78] Denny JC, Crawford DC, Ritchie MD, Bielinski SJ, Basford MA, Bradford Y, et al. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. Am J Hum Genet 2011;89(4):529–42.
[79] Brothers KB, Morrison DR, Clayton EW. Two large-scale surveys on community attitudes toward an opt-out biobank. Am J Med Genet A 2011;155A(12):2982–90.
[80] Kho AN, Hayes MG, Rasmussen-Torvik L, Pacheco JA, Thompson WK, Armstrong LL, et al. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. J Am Med Inform Assoc JAMIA 2012;19(2):212–8.
[81] Zuvich RL, Armstrong LL, Bielinski SJ, Bradford Y, Carlson CS, Crawford DC, et al. Pitfalls of merging GWAS data: lessons learned in the eMERGE network and quality control procedures to maintain high data quality. Genet Epidemiol 2011;35(8):887–98.
[82] El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. PloS One 2011;6(12):e28071.
[83] Wu Y, Liu M, Zheng WJ, Zhao Z, Xu H. Ranking gene-drug relationships in biomedical literature using Latent Dirichlet Allocation. Pac Symp Biocomput Pac Symp Biocomput 2012:422–33.
[84] Delaney JT, Ramirez AH, Bowton E, Pulley JM, Basford MA, Schildcrout JS, et al. Predicting clopidogrel response using DNA samples linked to an electronic health record. Clin Pharmacol Ther 2012;91(2):257–63.
[85] Carroll RJ, Eyler AE, Denny JC. Naïve electronic health record phenotype identification for rheumatoid arthritis. AMIA Annu Symp Proc AMIA Symp AMIA Symp 2011;2011:189–96.
[86] Liu M, Jiang M, Kawai VK, Stein CM, Roden DM, Denny JC, et al. Modeling drug exposure data in electronic medical records: an application to warfarin. AMIA Annu Symp Proc AMIA Symp AMIA Symp 2011;2011:815–23.
[87] Xu H, Fu Z, Shah A, Chen Y, Peterson NB, Chen Q, et al. Extracting and integrating data from entire electronic health records for detecting colorectal cancer cases. AMIA Annu Symp Proc AMIA Symp AMIA Symp 2011;2011:1564–72.
[88] Tamersoy A, Loukides G, Nergiz ME, Saygin Y, Malin B. Anonymization of longitudinal electronic medical records. IEEE Trans Inf Technol Biomed Publ IEEE Eng Med Biol Soc 2012;16(3):413–23.
[89] Peissig PL, Rasmussen LV, Berg RL, Linneman JG, McCarty CA, Waudby C, et al. Importance of multi-modal approaches to effectively identify cataract cases from electronic health records. J Am Med Inform Assoc JAMIA 2012;19(2):225–34.
[90] Ramirez AH, Shi Y, Schildcrout JS, Delaney JT, Xu H, Oetjens MT, et al. Predicting warfarin dosage in European-Americans and African-Americans using DNA samples linked to an electronic health record. Pharmacogenomics 2012;13(4):407–18.
[91] Fullerton SM, Wolf WA, Brothers KB, Clayton EW, Crawford DC, Denny JC, et al. Return of individual research results from genome-wide association studies: experience of the Electronic Medical Records and Genomics (eMERGE) Network. Genet Med 2012;14(4):424–31.
[92] Carroll RJ, Thompson WK, Eyler AE, Mandelin AM, Cai T, Zink RM, et al. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. J Am Med Inform Assoc JAMIA 2012;19(e1):e162–9.
[93] Parvathaneni SV, Ellis CR, Rottman JN. High prevalence of insulation failure with externalized cables in St. Jude Medical Riata family ICD leads: fluoroscopic grading scale and correlation to extracted leads. Heart Rhythm Off J Heart Rhythm Soc 2012;9(8):1218–24.
[94] Silver HJ, Niswender KD, Keil CD, Jiang L, Feng Q, Chiu S, et al. CNR1 genotype influences HDL-cholesterol response to change in dietary fat intake. PloS One 2012;7(5):e36166.
[95] Carrell D, Malin B, Aberdeen J, Bayer S, Clark C, Wellner B, et al. Hiding in plain sight: use of realistic surrogates to reduce exposure of protected health information in clinical text. J Am Med Inform Assoc JAMIA 2013;20(2):342–8.
[96] Falah N, McElroy J, Snegovskikh V, Lockwood CJ, Norwitz E, Murray JC, et al. Investigation of genetic risk factors for chronic adult diseases for association with preterm birth. Hum Genet 2013;132(1):57–67.
[97] Rasmussen-Torvik LJ, Pacheco JA, Wilke RA, Thompson WK, Ritchie MD, Kho AN, et al. High density GWAS for LDL cholesterol in African Americans using electronic medical records reveals a strong protective variant in APOE. Clin Transl Sci 2012;5(5):394–9.
[98] Brothers KB, Clayton EW. Parental perspectives on a pediatric human non-subjects biobank. AJOB Prim Res 2012;3(3):21–9.
[99] Velez Edwards DR, Naj AC, Monda K, North KE, Neuhouser M, Magvanjav O, et al. Gene-environment interactions and obesity traits among postmenopausal African-American and Hispanic women in the Women's Health Initiative SHARe Study. Hum Genet 2013;132(3):323–36.
[100] Westbrook MJ, Wright MF, Van Driest SL, McGregor TL, Denny JC, Zuvich RL, et al. Mapping the incidentalome: estimating incidental findings generated through clinical pharmacogenomics testing. Genet Med 2013;15(5):325–31.

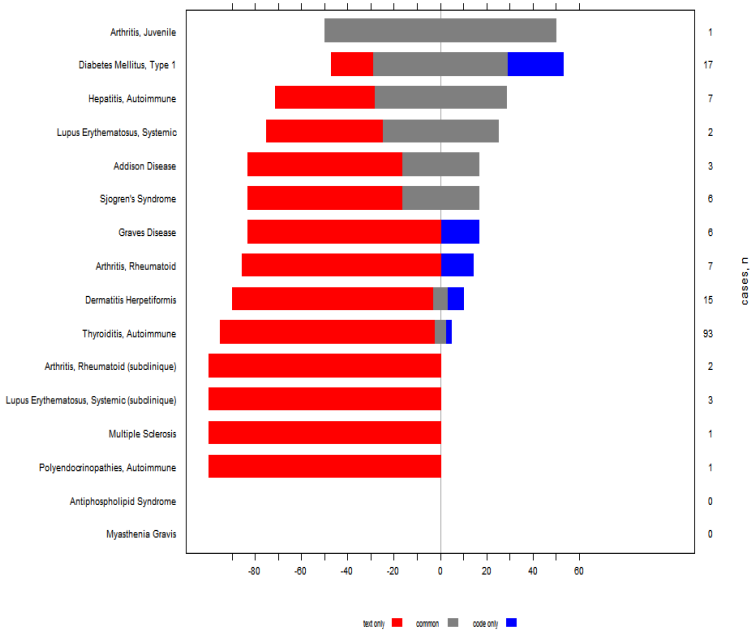
Quelles données dans les entrepôts ?

Des données **très hétérogènes**

- Structurées
  - E.g.:
    - Code diagnostics
    - Code de procédures médicales
- Non-structurées
  - Textes narratifs cliniques (comptes rendus)
  - Imagerie

Les trésors textuels

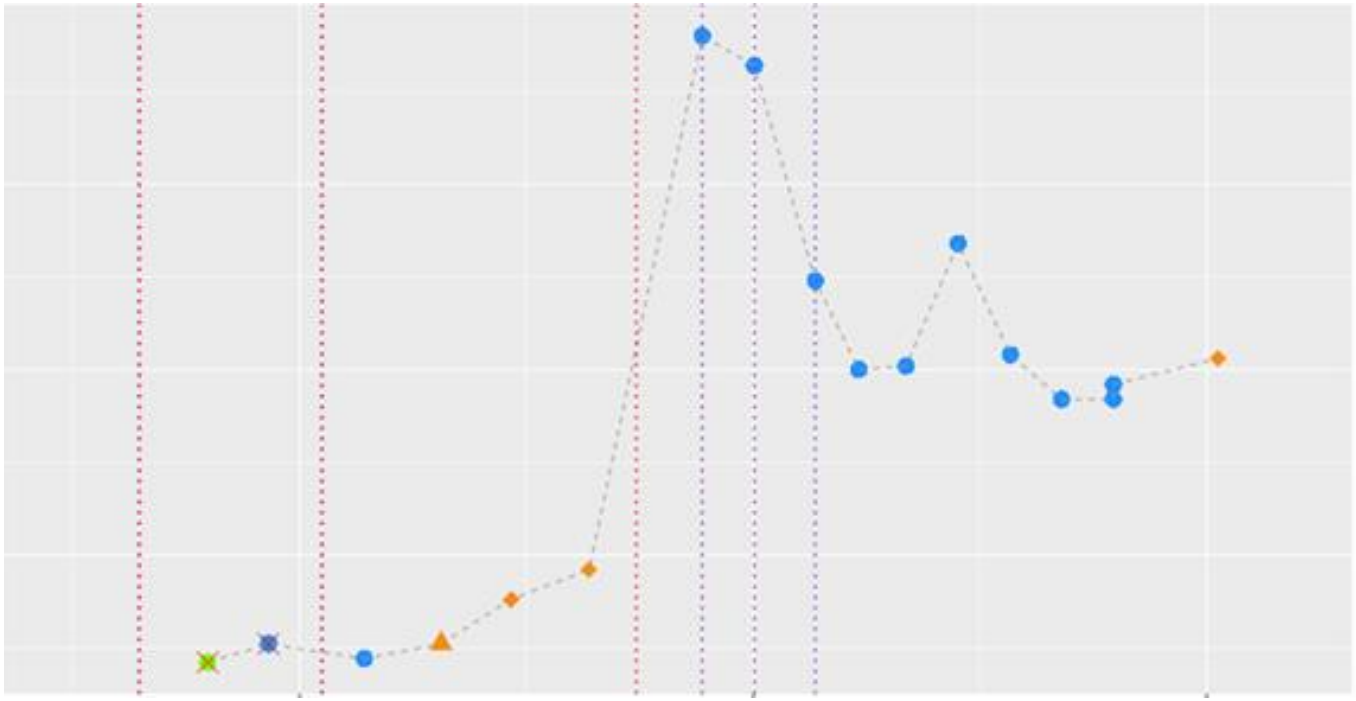
80 à 90%



de l'information médicale est contenue uniquement dans les textes libres

Séries temporelles

Horodatage de toutes les données



Sémantique du temps : Saisonnalité

“Sens” de la métadonnées (un dosage biologique à 3h du matin ≠ dosage à 8h)

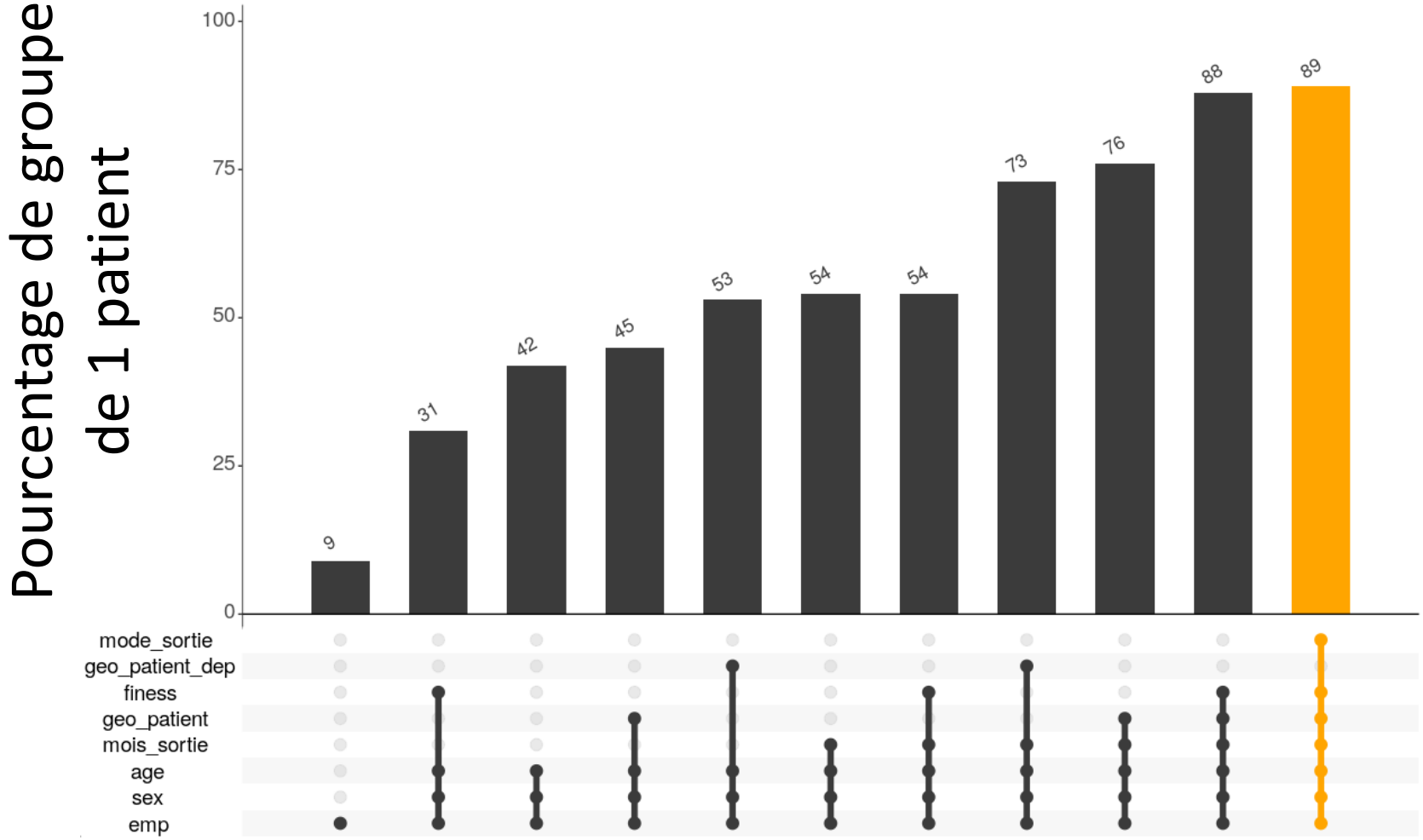
Génomiques et environnement

Partiellement ou pas encore dans les entrepôts, mais probablement indispensable

Exposome - Trajectoires de vie

Partiellement ou pas encore dans les entrepôts,
mais probablement indispensable

L’anonymat dans les bases médicales

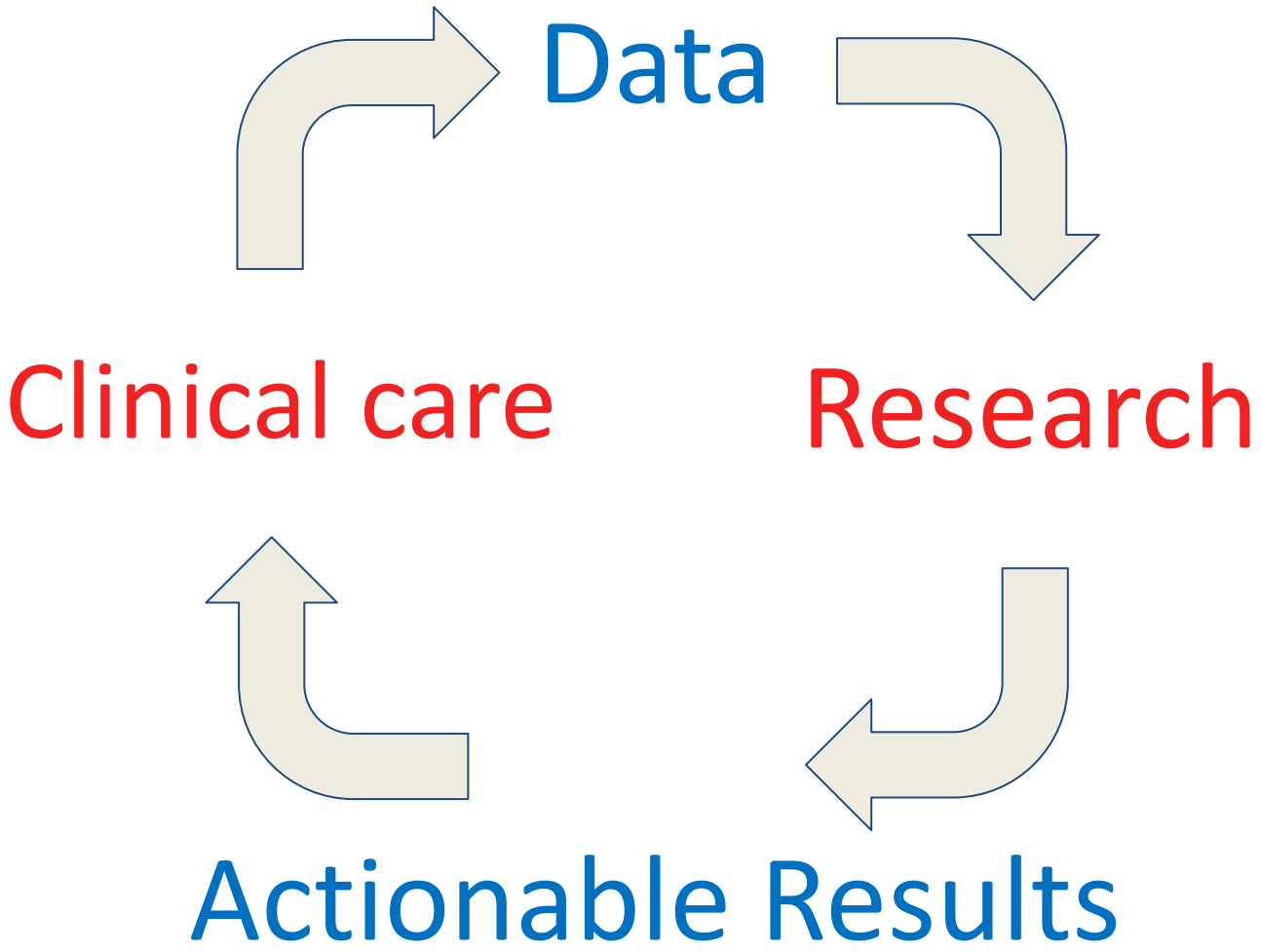


Anonymat du patient dans le PMSI : quel leurre est-il ? Blum D. *EMOIS 2011*

Les limites des entrepôts; besoin du retour aux sources

- Données **incomplètes** et volatiles
  - Besoin de contrôle
- **Hétérogène** dans la qualité
- Contexte pour la validation et **compléter** l'information
- Biais induits par les algorithmes d'extractions

Vers une recherche translationnelle



Vers une recherche translationnelle

- **Porosité** croissante entre soin et recherche
- Découvertes fortuites
- Evolution des connaissances
  - Données figées, mais amélioration des savoirs
- Maladies rares et essais thérapeutiques
  - **Médecine de précision**
  - Stratification des cohortes

Perspective FREE PREVIEW

The NEW ENGLAND JOURNAL of MEDICINE

Evidence-Based Medicine in the EMR Era

Jennifer Frankovich, M.D., Christopher A. Longhurst, M.D., and Scott M. Sutherland, M.D.

Results of Electronic Search of Patient Medical Records (for a Cohort of 98 Pediatric Patients with Lupus) Focused on Risk Factors for Thrombosis Relevant to Our 13-Year-Old Patient with Systemic Lupus Erythematosus.*

Outcome or Risk Factor	Keywords Used to Conduct Expedited Electronic Search	Prevalence of Thrombosis	Relative Risk (95% CI)
		no./total no (%)	
Outcome — thrombosis	"Thrombus," "Thrombosis," "Blood clot"	10/98 (10)	Not applicable
Thrombosis risk factor			
Heavy proteinuria (>2.5 g per deciliter)			
Present at any time	"Nephrosis," "Nephrotic," "Proteinuria"	8/36 (22)	7.8 (1.7–50)
Present >60 days	"Urine protein"	7/23 (30)	14.7 (3.3–96)
Pancreatitis	"Pancreatitis," "Lipase"	5/8 (63)	11.8 (3.8–27)
Antiphospholipid antibodies	"Aspirin"	6/51 (12)	1.0 (0.3–3.7)

« We made the decision on the basis of the best data available »
« in the light of experience as guided by intelligence. »

Découvertes fortuites et incidentalomes

Natural language processing of radiology reports for the detection of thromboembolic diseases and clinically relevant incidental findings

Anne-Dominique Pham †, Aurélie Névéol †, Thomas Lavergne, Daisuke Yasunaga, Olivier Clément, Guy Meyer, Rémy Morello and Anita Burgun

BMC Bioinformatics

https://doi.org/10.1186/1471-2105-15-266

Journal of Biomedical Informatics 73 (2017) 51–61

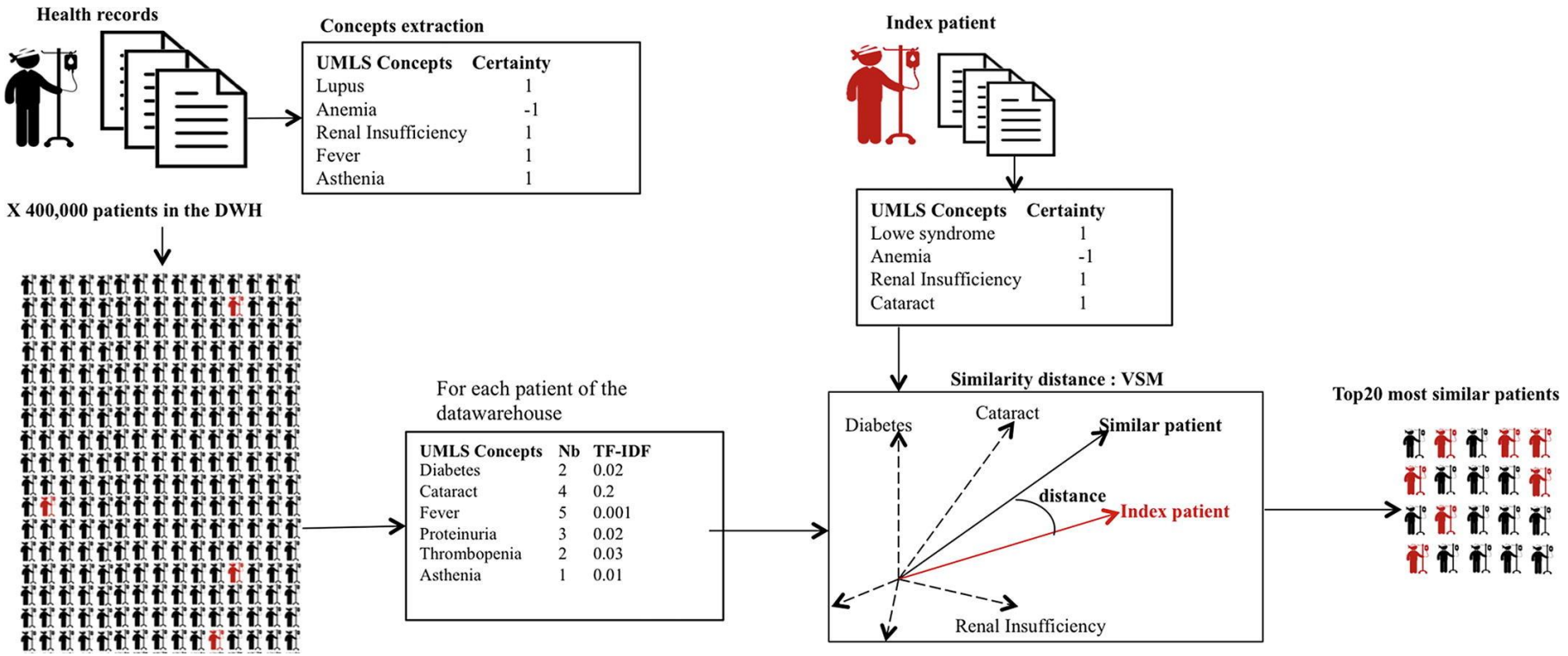
Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Finding patients using similarity measures in a rare diseases-oriented clinical data warehouse: Dr. Warehouse and the needle in the needle stack

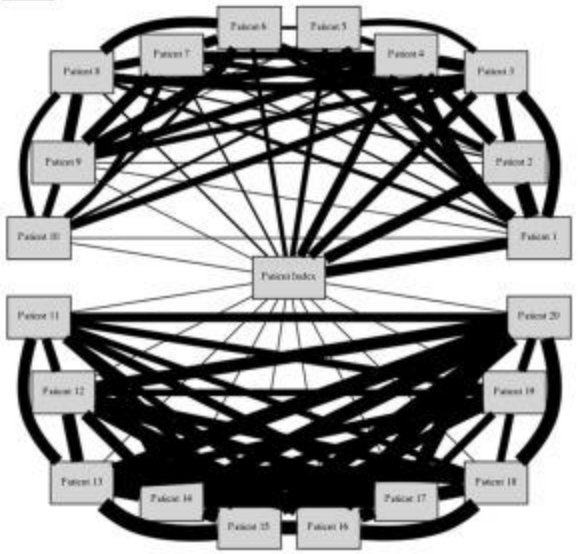
Nicolas Garcelon [a,b,c,*], Antoine Neuraz [c,d], Vincent Benoit [a,b], Rémi Salomon [a,b,e], Sven Kracker [a,b,f], Felipe Suarez [a,b,g], Nadia Bahi-Buisson [a,b,h], Smail Hadj-Rabia [a,b,i], Alain Fischer [a,b,j,k,l], Arnold Munnich [a,b,m,n], Anita Burgun [c,d,o]



Patient F,
47 ans

Documents | Biologie | TimeLine | Parcours | Cohorte | Concepts | Similarité

Limiter aux patients avec ces termes dans les comptes rendus: Pré-calculer Nb patients :
Exclure les patients avec ces termes dans les comptes rendus : cd40 ligand
Exclure certains codes UMLS pour le calcul de similarité (Pour évaluation) :
Nb minimum de concepts par patient : 10
Limite nombre de patients similaires : 20
Anonymiser :
Exclure cette cohorte des résultats :
calculer



Search:

Patient	Similarité	Concepts communs	Concepts patient
Patient 1 (statut cohorte APDS : 1) (statut cohorte APDS2 : 1) (statut cohorte DIH inclus CEREDIH : 1) Afficher le comparateur	81	Conjonctivite, Toux, DYSPHAGIE, dyspnee, Fievre, Hyperplasie, deficit immunitaire, Lymphopenie, mycose, Otite, Douleur, Sinusite, syndrome d'apnee du sommeil, Surinfection, Sinusite chronique, larmoiement, Rhinorrhee purulente, syndrome hyper ige, Adenopathie, Intervention chirurgicale, Gene respiratoire, rhinosinusite, Cicatrice, Recidive	Vegetations, **stenose pharyngee**, Gaz, Epanchement, Retrait, hypertrophie des amygdales, suppuration, pi3k, Thrombus, deficit immunitaire humoral, Ptosis, Croute, septicemie, dacryocystite, hypertrophie, Observation, obstruction laryngee, Bouton, Respiration, Aphte, tor, Complication, Thrombose, Hernie inguinale bilaterale, obstruction chronique des voies aeriennes, nutrition, Stenose laryngee, Oedeme, Hypopnee, vesicule, obstruction nasale bilaterale, Deficit en IgG, Apnee, Obstruction, Lesions muqueuses, Nodule, SAOS, Circulation collaterale, Douleurs abdominales, adherence, hypertrophie des vegetations adenoides, apds, appareil respiratoire
Patient 2 (statut cohorte APDS : 1) (statut cohorte APDS2 : 1) (statut cohorte DIH inclus CEREDIH : 1) Afficher le comparateur	80	Anorexie, Asthenie, dilatation des bronches, rhinite, Infection, Conjonctivite, Toux, Fievre, gastroenterite, cephalees, Hyperplasie, deficit immunitaire, Lymphopenie, Otite, Douleur, syndrome d'apnee du sommeil, Surinfection, baisse, Sinusite chronique, Adenopathie cervicale, Tolerance, Ecoulement, Rhinorrhee purulente, syndrome hyper ige, Adenopathie, Parasinusite, Rhinorrhee, Cicatrice, Recidive	Enuresie, bruits, Aphte, Hypopnee, 3 6, bronchite chronique, Ptosis, Toux nocturne, Hypogammaglobulinemie, somnolence diurne, deviation de la cloison nasale, Gastro-enterite aigue, ANEMIE MICROCYTAIRE, Diphterie, Apnee, 3 4, Anemie, Injection intraveineuse, baisse de l'acuite visuelle, **Conjonctivite chronique**, Ronchi, crepitants, deficit de l'immunite humorale, Conjonctivite purulente, pal, lymphoproliferation, erytheme localise, Infections respiratoires, PIK3R1, cauchemars, ronflement, Obstruction, Enuresie nocturne, Gynecomastie, hypertrophie, Urgences, endormissement, Epistaxis, Rhinite chronique, **pi3k**, **apds**, Irritable, bronchorrhee, Agitation, Prise de poids, Toux chronique, Taux d'anticorps, cavite, vomissement, Sueurs nocturnes, SAOS, Frequence cardiaque maximale, sommeil agite, Respiration buccale, Toux persistante
Patient 3 (statut cohorte DIH inclus CEREDIH : 1) (statut cohorte APDS : 1)		douleur abdominale, Asthme, ATELECTASIE, dilatation des bronches, rhinite, Infection, Toux, Diarrhee,	vesicule, Decouverte fortuite, Toux chronique, Douleurs abdominales, Rhinite chronique, regression, bruits, PIK3CD, HPA3, cavite, PIK3, Vegetations, distension, Diarrhee chronique, hypertrophie, Retrecissement, CD31, Purpura, **apds**, ANEMIE MICROCYTAIRE, consideration, eruption, fenetre aorto-pulmonaire, adenopathies hilaires, Sequestration splenique, positive, Hematome, douleur dentaire, Hypersplenisme, Cytolyse hepatique, Infection virale, Toux nocturne, Moelle riche, EMPHYSEME, Hypogammaglobulinemie, selles glaireuses, malformation pulmonaire, Respiration, Sueurs nocturnes, 70%, Pancytopenie, pi3k, Detresse

Anonymiser les données ?

Besoin de pouvoir **retourner** :

- Au patient
- A son dossier

Données **profondément identifiantes** (souvent indirectement)

Equité devant les méthodes d'anonymisation ?

- Pseudonymisation (quelles données d'entrainement)

Anonymiser les données ?

Besoin de pouvoir **retourner** :

- Au patient
- A son dossier

Données **profondément identifiantes** (souvent indirectement)

Equité devant les méthodes d'anonymisation ?

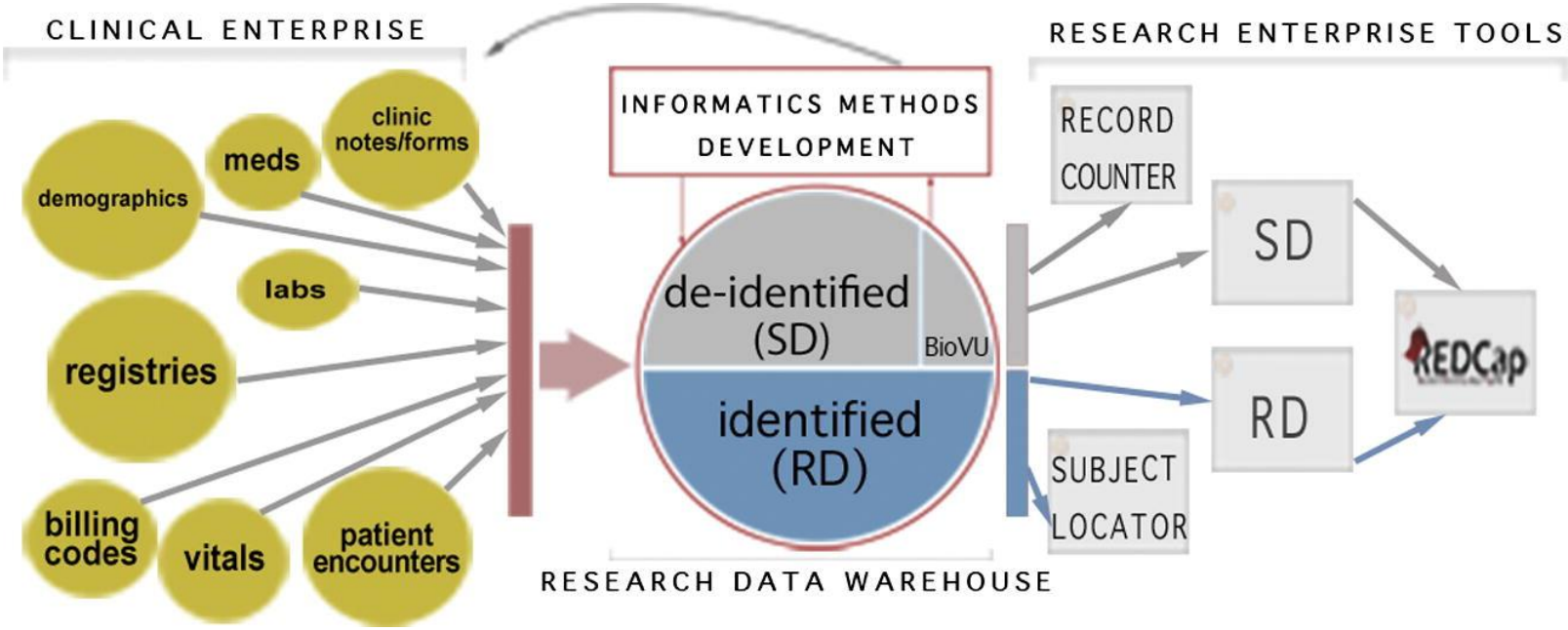
- Pseudonymisation (quelles données d'entrainement)

Difficile de garantir un résultat absolu

Comment protéger les données ?

- Contrôles techniques, organisationnels, et réglementaires
- Les données **restent** sur le lieux de génération
- Les actions sur les données sont journalisées

Moyens techniques



Secondary use of clinical data: The Vanderbilt approach. Journal of Biomedical Informatifs. Danciu et al. 2014

Moyens organisationnels

Dr Warehouse

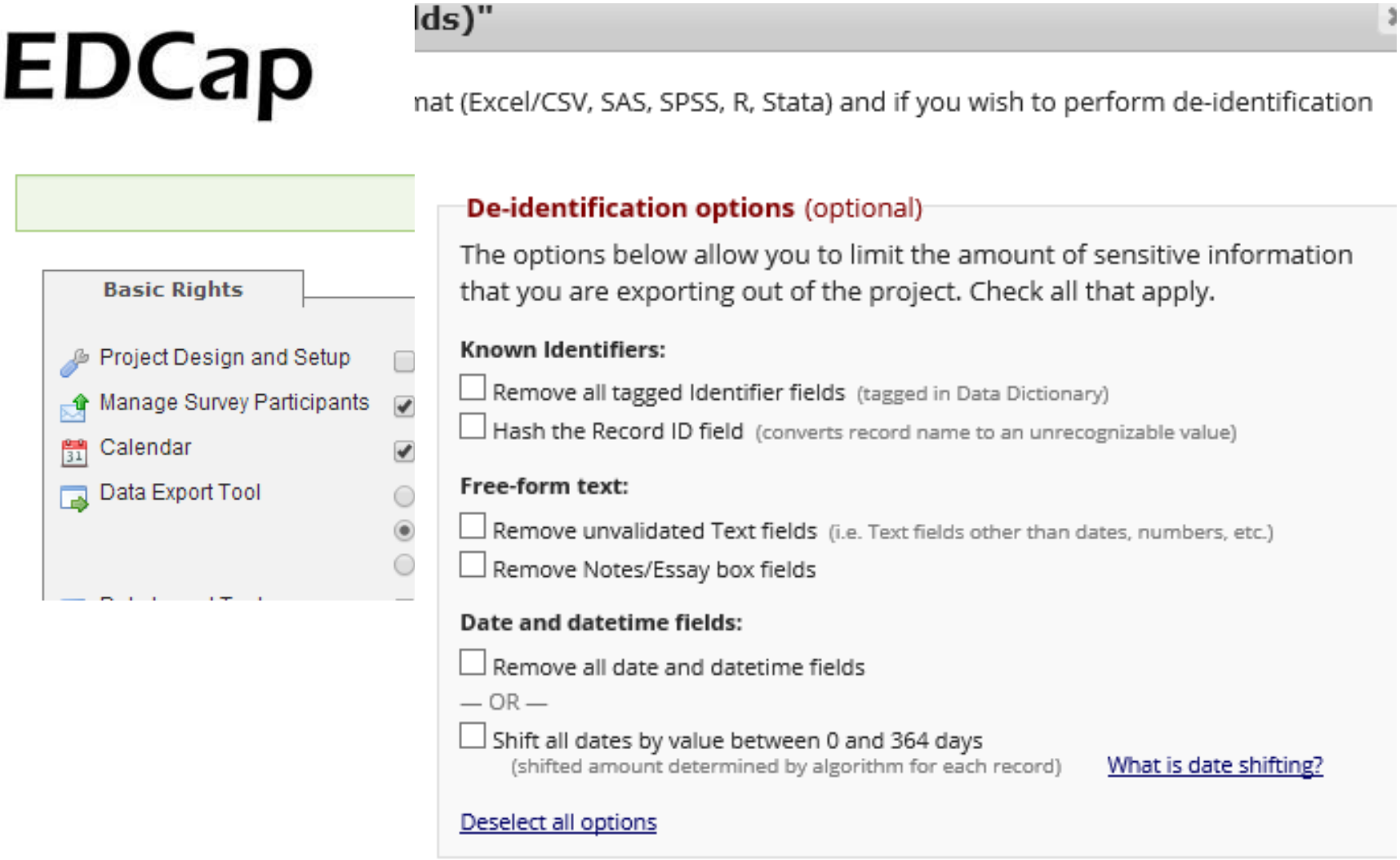
i2b2

HEGP; AP-HP

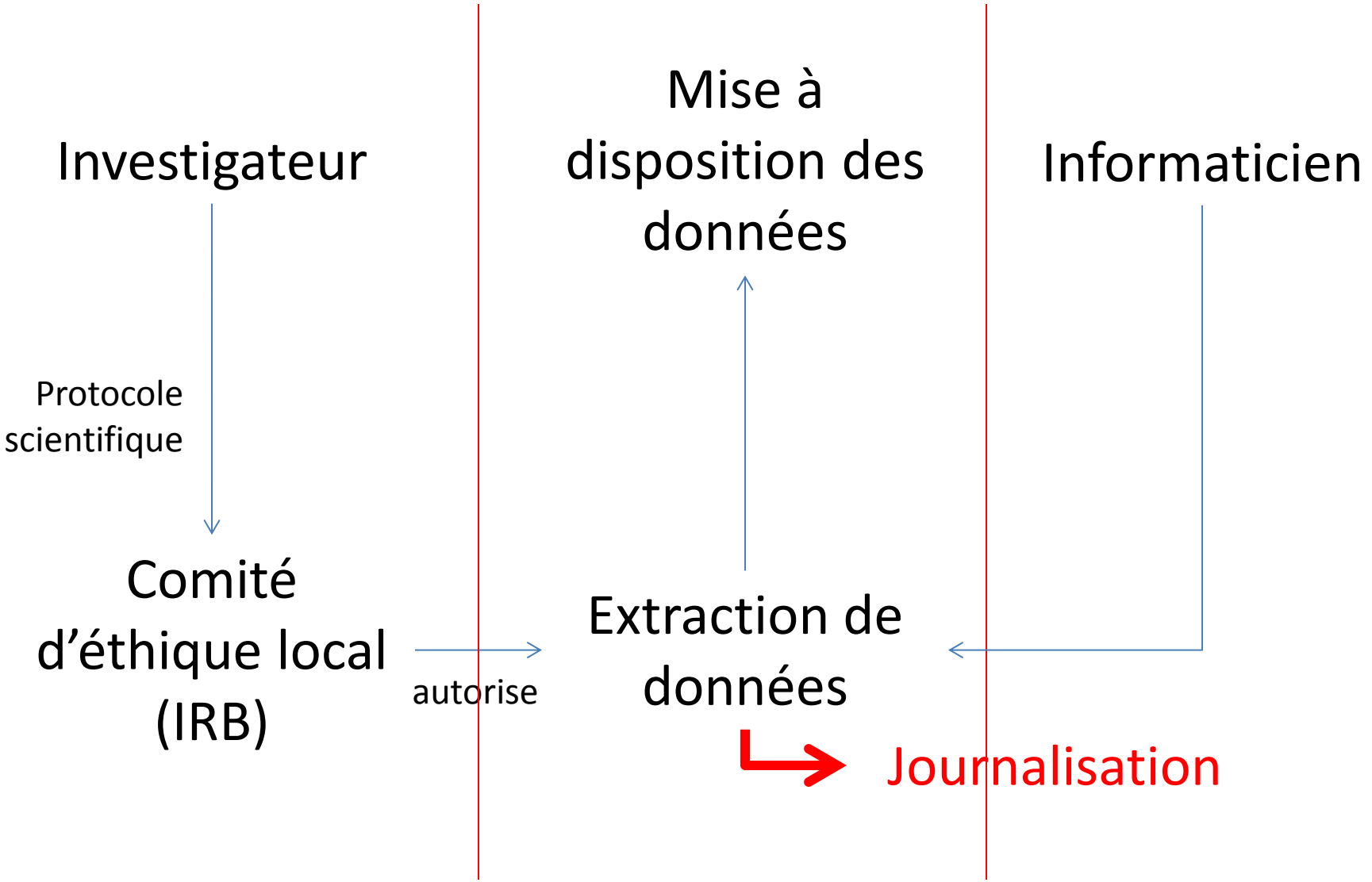
- Refus de participations, suppression des données
- Textes dé-identifiés (recherche par les noms et prénoms d'un patient impossible)
- Profils selon les utilisateurs
  - Médecins du service de prise en charge
  - Autres utilisateurs

Moyens organisationnels

REDCap



Moyens réglementaires : circuits d'accès aux données



Renforcer les aspects de formation et responsabilité des acteurs

Exemple : accès à la base MIMIC III

Required Modules		
	Date Completed	Score
Belmont Report and Its Principles (ID: 1127)	Incomplete	0/0 (0%)
History and Ethics of Human Subjects Research (ID: 498)	Incomplete	0/0 (0%)
Basic Institutional Review Board (IRB) Regulations and Review Process (ID: 2)	Incomplete	0/0 (0%)
Records-Based Research (ID: 5)	Incomplete	0/0 (0%)
Genetic Research in Human Populations (ID: 6)	Incomplete	0/0 (0%)
Populations in Research Requiring Additional Considerations and/or Protections (ID: 16680)	Incomplete	0/0 (0%)
Research and HIPAA Privacy Protections (ID: 14)	Incomplete	0/0 (0%)
Conflicts of Interest in Human Subjects Research (ID: 17464)	Incomplete	0/0 (0%)
Massachusetts Institute of Technology (ID: 1290)	Incomplete	0/0 (0%)

Challenges

Patients **partenaires**

- Mieux impliquer
- Fournisseurs de données (M.J. Fox Foundation, associations de parents…)

Expérience du **CARPEM** (Site intégré de rech.sur le cancer de l'Univ. Paris Descartes)

- Comité des patients
  - Population générale et population en situation de maladie

Partage des données, SNDS, Health Data Hub…

JAMA Internal Medicine

Public Preferences About Secondary Uses of Electronic Health Information FREE

David Grande, MD, MPA[1,2]; Nandita Mitra, PhD[3]; Anand Shah, MD, MSHP[4]; Fei Wan, MS[3]; David A. Asch, MD, MBA[1,2,5]

Willingness to share data (1: low, 10: high)

Respondent Willingness to Share Personal Health Information by Conjoint Scenario (unadjusted means)

Conjoint Scenario			Willingness to Share Personal Health Information (1=low, 10=high)				
Sensitivity	Use	User	Overall	Non-Hispanic African-American	Hispanic	Non-Hispanic White	P-Values*
Low	Research	University Hospital	6.82	6.58	6.38	6.98	0.04
High	Research	University Hospital	6.72	6.60	6.30	6.73	0.84
Low	Research	Drug Company	5.90	5.71	5.92	5.94	0.84
High	Research	Drug Company	5.86	5.79	5.79	5.90	0.86
Low	Research	Public Health Dept.	6.10	6.23	6.14	6.06	0.84
High	Research	Public Health Dept.	6.18	6.07	6.12	6.22	0.84

High sensitivity
University Hospital
Research

Willingness **6.72**

JAMA Internal Medicine

Public Preferences About Secondary Uses of Electronic Health Information

FREE

David Grande, MD, MPA[1,2]; Nandita Mitra, PhD[3]; Anand Shah, MD, MSHP[4]; Fei Wan, MS[3]; David A. Asch, MD, MBA[1,2,5]

JMIR Publications
The Leading eHealth Publisher

Patient Perspectives on Sharing Anonymized Personal Health Data Using a Digital System for Dynamic Consent and Research Feedback: A Qualitative Study

Karen Spencer[1*], PhD ; Caroline Sanders[2*], PhD ; Edgar A Whitley[3], PhD ; David Lund[4], PhD ; Jane Kaye[5], PhD ; William Gregory Dixon[1,6,7], MRCP(UK), PhD

JAMA Internal Medicine

Secondary Use of Health Information
Are We Asking the Right Question?

Isaac S. Kohane, MD, PhD[1]

Remerciements

HEGP - Necker

Informatique médicale

- A. Neuraz
- A. Burgun

Institut Imagine

- N. Garcelon

HEGP

Informatique médicale

- AS. Jannot
- M. Wack
- E. Zapletal
- H. Countouris
- W. Digan
- D. Baudoin