



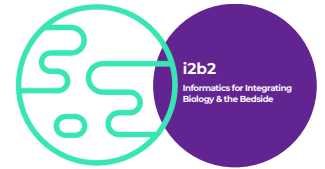
HEALTH
DATA HUB



i2b2

**Informatics for Integrating
Biology & the Bedside**

**Ce document a été réalisé en partenariat avec Veltys
et avec la contribution du CHU de Bordeaux.**

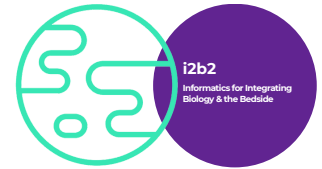


Description et type de standard : i2b2 (Informatics for Integrating Biology & the Bedside) est une **plateforme de recherche clinique** qui contient un **schéma de données**, une **couche d'application** et des **APIs**. Elle a été créée en 2004 par la Harvard Medical School et financée par le **NIH** (National Institutes of Health).

Domaine d'application : i2b2 est utilisée dans les domaines cliniques ainsi qu'en **génomique** dans le cadre de **recherches cliniques et translationnelles**, en particulier pour **l'exploration et l'identification de cohortes de patients**.

Maturité / Utilisation : la version actuelle d'i2b2 est la **v1.7.13** sortie en 2022 et est utilisée par plus de **250 institutions** dans le monde (aux États-Unis dans de nombreuses universités, en France dans plusieurs établissements de santé, par des entreprises privées, etc. voir l'élément « Utilisation sur le marché » en partie 4. Valorisation).

1. Général



PRÉSENTATION

- **Pays d'origine** : États-Unis.
- **Consortium d'origine** : NIH (National Institutes of Health)¹.
- **Type de standard** : Schéma de données.
- **Description** :
 - i2b2 (Informatics for Integrating Biology & the Bedside), créée en **2004** par la Harvard Medical School, est une **plateforme de recherche clinique**² qui permet d'organiser et de transformer des données cliniques orientées patient³. Elle est optimisée pour la recherche génomique clinique. La plateforme i2b2, ou logiciel (« i2b2 Software »), est constituée de plusieurs cellules communicantes via des services web XML, qui forment le **i2b2 Hive**⁴ (voir Figure 1) :
 - Cellules centrales (i2b2 Core Cell) :
 - **Project Management**⁵: gère la configuration et la maintenance du Hive, les projets, les utilisateurs et la sécurité.
 - **Ontology Management**⁶: gère les terminologies typiquement utilisées dans le schéma de données et contient les informations sur les relations entre les concepts dans l'ensemble du Hive.
 - **Data Repository (CRC : Clinical Research Chart)**⁷: contient les données phénotypiques et génotypiques du Hive dans un format structuré et désidentifié.
 - **Workflow Framework**⁸: constitue l'espace de travail des utilisateurs où les informations peuvent être partagées et peut être librement consulté par le chef de projet.
 - Cellules optionnelles (i2b2 Optional Cell) :
 - **File Repository (CRC)**⁹: contient de gros fichiers de données (images radiologiques, séquences génétiques, ...).
 - **Identity Management**¹⁰: gère les données sensibles des patients, en accord avec la réglementation HIPAA (Health Insurance Portability and Accountability Act).

1) [Our History – i2b2 tranSMART Foundation](#)

2) i2b2 est un NBCB (National Center for Biomedical Computing), un programme développé par le NIH (National Institutes of Health) aux États-Unis ayant pour objectif de participer au développement d'une infrastructure informatique universelle pour accélérer le progrès de la recherche biomédicale.

3) Introduction à i2b2 disponible sur le site de la communauté d'i2b2 : [i2b2-workshop at AMIA Nov-2020](#)

4) [i2b2 Community Wiki](#)

5) [Project Management \(PM\) Cell](#)

6) [Ontology Management \(ONT\) Cell](#)

7) [i2b2 Software Architecture - Data Repository \(CRC\) Cell](#)

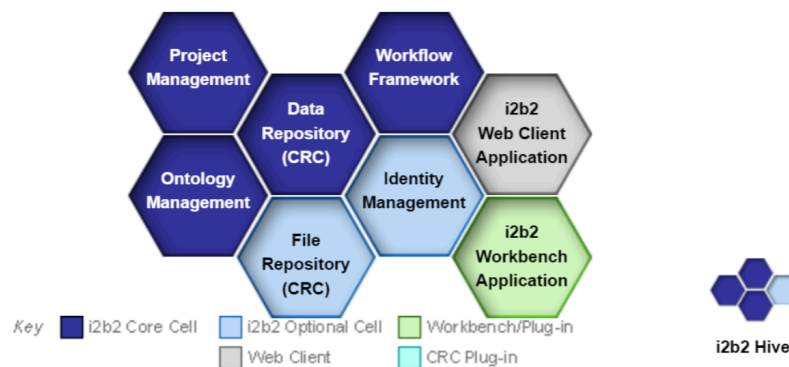
8) [Workplace Framework \(WORK\) Cell](#)

9) [i2b2 Software Architecture - File Repository \(FR\) Cell](#)

10) [i2b2 Software Architecture - Identity Management \(IM\) Cell](#)

- Une **application Web Client**¹¹ qui permet d'interroger, d'analyser et d'afficher les données du i2b2 Hive en communiquant avec les cellules i2b2 (on parle aussi du « Query & Analysis Tool », voir l'élément « Outils compatibles » en partie 5. Valorisation).
- Une **application Workbench** qui réalise les mêmes analyses que le Web Client d'une façon plus approfondie.

Figure 1 : Structure du i2b2 Hive



Source : [i2b2 Community Wiki](#)

APPLICATION

- **Organisme en charge**^{1,12} : **i2b2 tranSMART Foundation.**
- **Domaines d'application en santé** : la plateforme i2b2 est principalement utilisée en **recherche clinique et translationnelle** dans les domaines **cliniques** et en **génomique**¹³.
- **Principaux cas d'usage** :
 - La plateforme facilite en effet la recherche à partir d'un symptôme particulier, d'une histoire médicale, d'un profil génomique, etc., ainsi que l'analyse des résultats⁴.
 - L'outil SHRINE (Shared Health Research Information Network) permet de constituer des **réseaux fédérés** de patients caractérisés et ainsi de compter par exemple le nombre agrégé de patients concernés dans les hôpitaux participants⁴.
 - La suite tranSMART d'exploration et visualisation de données, analyses génomiques et outils d'ETL a été développée par l'industrie pharmaceutique pour des études de **recherche translationnelle**.

11) Démonstration du web client : [i2b2 Web Client](#)

12) En mai 2017, i2b2 et tranSMART Foundations ont fusionné pour intégrer les deux plateformes dominantes utilisées dans la recherche clinique et translationnelle pour faire avancer la médecine de précision dans le monde.

13) <https://www.i2b2.org/about/intro.html>

- **Illustration concrète, exemple d'utilisation sur un cas simple :**

- Hong et al. (2016)¹⁴ détaillent la **création d'un entrepôt de données i2b2** (IDRs : Integrated Data Repository) **pour la recherche sur le cancer ovarien** à la Clinique Mayo aux États-Unis.
 - Le **schéma de données i2b2 a été sélectionné pour structurer les données des patients sur un modèle commun au sein de l'IDR** en raison principalement :
 - De son interface de requête facile à utiliser pour les chercheurs.
 - De ses modules analytiques paramétrables.
 - Les **données** proviennent principalement de **trois sources** :
 - Du registre du cancer ovarien (données de diagnostic, données démographiques, etc.).
 - Des données des tests de laboratoire (valeurs de tests CA-125, CBC, etc.).
 - Du système de gestion des documents (données de traitement par chimiothérapie extraites de documents au format CDM).
 - Les **étapes menant à la centralisation dans l'entrepôt i2b2** sont les suivantes :
 - **Nettoyage et vérification** des données pour s'assurer de leur qualité.
 - **Standardisation** des données (utilisation de l'ontologie d'i2b2 pour harmoniser les terminologies employées, utilisation de codes standards, etc.).
 - **Pseudonymisation** des données (génération d'ID patients aléatoires pour désidentifier les patients).
 - **Transformation** des données dans un schéma i2b2 compatible et chargement dans la base PostgreSQL.
 - Au total, **286 235 observations** ont pu être chargées dans la base i2b2.
- L'outil de requêtage et d'analyse (« Query & Analysis Tool ») intégré à i2b2 est utilisé pour les cas d'usage suivants : **identification de cohortes de patients, analyse des caractéristiques des patients.**

14) Article de Hong Na, Li Zheng, Kiefer C. Richard, Robertson M., Goode E. Whang Chen, Jiang Guoqian (2016) « Building an i2b2-Based Integrated Data Repository for Cancer Research: A Case Study of Ovarian Cancer Registry ». DMAH@VLDB : <https://www.semanticscholar.org/paper/Building-an-i2b2-Based-Integrated-Data-Repository-A-Hong-Li/d71fc7a821d22ff53ef19e3fca39a35a20f0cf3f>

DONNÉES

- **Typologie de données concernées : données observationnelles** de différentes sources : il peut s'agir de données issues des dossiers de santé électroniques (EHR : Electronic Health Record), de demandes de remboursement, de notes, d'images, de données génomiques, de données d'essais cliniques, etc¹⁵.
- **Type de granularité :**
 - Une observation correspond à un **événement médical** (un diagnostic, une procédure, la prescription ou l'administration d'un médicament, etc.)¹⁶. Chaque ligne concerne un seul patient et une seule observation. Une visite médicale peut cependant donner lieu à plusieurs lignes dans la base pour recenser différentes observations (voir l'élément « Description technique du schéma de données » ci-dessous).
 - Les **modificateurs permettent de préciser les concepts utilisés**. Lorsqu'un modificateur est associé à une observation, il peut y avoir plusieurs lignes pour une même observation (voir la description de la table de dimension « Modificateur » ainsi qu'un exemple dans l'élément « Description technique du schéma de données » ci-dessous).
- **Utilisation dans plusieurs langues** : les concepts sont affichés en anglais. Il n'y a pas d'indications concernant la disponibilité dans une autre langue excepté un cas d'implémentation au Japon qui a conduit à devoir modifier les codes sources pour s'adapter au japonais¹⁷.

DISPONIBILITÉ DE LA DOCUMENTATION D'IMPLÉMENTATION

- **Guide d'implémentation** disponible sur le site officiel de la communauté i2b2¹⁸.
- **Documentation** sur le schéma de données i2b2¹⁹.
- **Description du Hive** et **guide d'installation** de chaque cellule⁵.
- **GitHub** d'i2b2²⁰.

15) <https://community.i2b2.org/wiki/display/BUN/1.+Introduction>

16) <https://community.i2b2.org/wiki/display/BUN/2.+Quick+Start+Guide>

17) Article de Takai-Igarashi T, Akasaka R, Suzuki K, Furukawa T, Yoshida M, Inoue K, Maruyama T, Maejima T, Bando M, Takasaki M, Sakota M, Eguchi M, Konagaya A, Matsuura H, Suzumura T, Tanaka H. « On experiences of i2b2 (Informatics for integrating biology and the bedside) database with Japanese clinical patients' data ». Bioinformatics. 2011 Mar 26 : [On experiences of i2b2 \(Informatics for integrating biology and the bedside\) database with Japanese clinical patients' data](#)

18) Guide d'implémentation : [i2b2 Installation Guide](#)

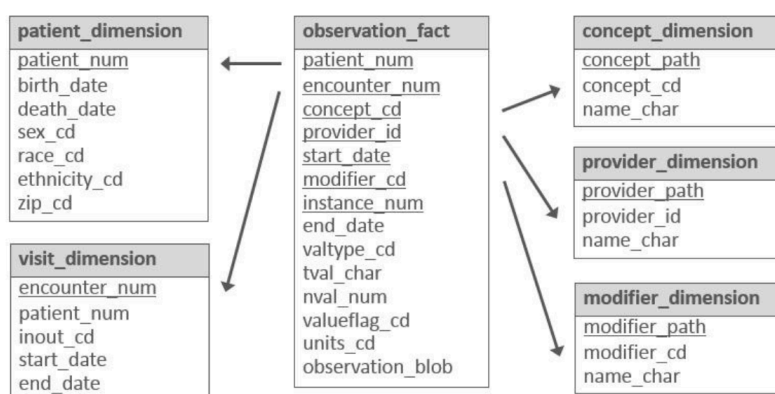
19) [i2b2 Common Data Model Documentation](#)

20) [i2b2 Foundation · GitHub](#)

DESCRIPTION TECHNIQUE DU SCHÉMA DE DONNÉES

- La structure d'i2b2 consiste en un **schéma en étoile** (voir Figure 2 ci-dessous) composé d'une **table de faits** (qui contient les **observations** du modèle ; ex : diagnostics, procédures, résultats de tests, etc.) et de 6 **tables de dimensions** (qui contiennent des informations descriptives sur les faits à travers la définition d'entités ; ex : un patient, un concept, ...).

Figure 2 : Schéma en étoile d'i2b2



Source : [Documentation i2b2](#)

- Liste des tables du schéma de données :
 - Table de faits (observation_fact)**²¹ : elle contient les observations sur un patient collectées lors d'une visite. Une visite peut être divisée en plusieurs lignes pour recenser différentes observations (voir l'exemple dans le Tableau 1 ci-dessous).

Tableau 1 : Exemple d'enregistrement dans la table observation_fact

patient_num	encounter_num	concept_cd	start_date	modifier_cd	valtype_cd	tval_char	nval_num
1000001	123456	ICD9:462	2007-08-09	@			
1000001	123456	LOINC:1751-7	2007-08-10	@	N	E	4.3
1000001	298765	LOINC:6598-7	2007-09-15	@	N	L	0.01
1000002	890123	LOINC:6598-7	2009-03-20	@	T	NEG	
1000002	543210	VITAL:BP	2010-05-01	systolic	N	120	
1000002	543210	VITAL:BP	2010-05-01	diastolic	N	80	
1000002	543210	VITAL:BP	2010-05-01	position	T	standing	

Source : [Documentation i2b2](#)

- Table de dimension Patient (patient_dimension)** : cette table contient les informations démographiques sur les patients (identifiant du patient, sexe, âge, date de naissance, statut marital, code postal, etc.).

21) [OBSERVATION FACT Table](#)

22) [PATIENT_DIMENSION Table](#)

- **Table de dimension Visite (visit_dimension)**²³: cette table contient les informations sur les différentes visites au cours desquelles les observations ont été collectées. Ces sessions peuvent impliquer le patient directement (exemple : visite chez un médecin) ou indirectement (exemple : tests réalisés sur un tube sanguin).
- **Table de dimension Concept (concept_dimension)**²⁴: cette table contient les concepts (diagnostics, procédures, médicaments, tests de laboratoire, etc.). Tous les types de données peuvent être stockés en tant que concepts (données démographiques et génétiques).
- **Table de dimension Prestataire (provider_dimension)**²⁵: cette table contient les informations relatives aux médecins ou prestataires de soins au sein d'une institution (position du prestataire au sein de l'institution, etc.).
- **Table de dimension Modificateur (modifier_dimension)**²⁶: cette table contient tous les modificateurs. Un modificateur permet de préciser un concept (fréquence, voie d'administration, etc.). Par exemple, « systolique », « diastolique » et « position » sont des modificateurs pour le concept de pression artérielle.

NIVEAU DE GÉNÉRALISATION (FACILITÉ DE REMPLISSAGE DES CHAMPS DU STANDARD)

- **Note : 0,8 / 1.** Cette note combine plusieurs sous-critères.
- Elle s'explique par :
 - a. Il n'y a pas de terminologies locales imposées (voir l'élément « Flexibilité dans les choix des terminologies » en partie 3. Technique).
 - b. La flexibilité du standard (voir l'élément, « Flexibilité du standard, personnalisation » en partie 3. Technique).
 - c. L'absence de contraintes d'implémentation (voir l'élément « Contraintes d'implémentation » en partie 3. Technique).
 - d. Une couverture moyenne des cas d'usage (voir l'élément « Principaux cas d'usages » en partie 1. Général).
 - e. Une couverture moyenne des domaines d'application (voir l'élément « Domaines d'application en santé » en partie 1. Général).

La facilité de remplissage des champs du standard dépend :

- De la **disponibilité des données sources** :
 - Les tables du modèle i2b2 font référence à des données que l'on peut retrouver dans les bases de données des hôpitaux, des laboratoires, des cabinets de ville (détails des visites, des médicaments prescrits, des diagnostics, des examens, ...).
 - Toutefois, en pratique plusieurs difficultés peuvent être rencontrées pour remplir ces champs :

23) https://community.i2b2.org/wiki/display/ServerSideDesign/VISIT_DIMENSION+Table

24) [CONCEPT_DIMENSION Table](#)

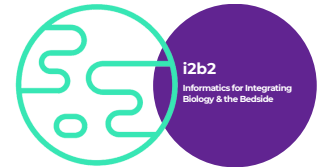
25) [PROVIDER_DIMENSION Table](#)

26) [MODIFIER_DIMENSION Table](#)

- Dans les établissements de santé, ces données sont rarement stockées de manière centralisée.
- Selon les pratiques de l'établissement et des professionnels de santé, le niveau de complétude des champs ainsi que la profondeur de l'historique varient.
- De la facilité à réaliser le **mapping des données sources** :
 - Le **concept EAV** (Entity-Attribute-Value) permet une **approche générique** et une **intégration simple de tous types de données source aux tables i2b2**, sans modification de la structure de la base de données ou du logiciel (voir l'élément « Flexibilité du standard, personnalisation » en partie 3. Technique).
 - Cependant, ce concept EAV peut conduire à des pertes de performance (voir l'élément « Performance » en partie 3. Technique).

La couche applicative reste ancienne dans son architecture et les technologies implémentées. Cependant, les travaux dans le contexte de la fondation i2b2/transmart tendent progressivement à converger vers des technologies plus à l'état de l'art.

2. Gouvernance



LIBRE ACCÈS AUX SCHÉMAS DE DONNÉES OUI²⁷.

MODALITÉS D'ACCÈS ET DISTRIBUTION DES SOLUTIONS BASÉES SUR CE STANDARD

- Les composantes du logiciel i2b2 (« Software ») sont **open source** publiées sous la licence libre **Mozilla Public License, version 2.0**^{27,28} : il s'agit des cellules centrales du Hive i2b2, des applications Web Client et Workbench. Des données de démonstration ainsi que de la documentation sont également disponibles en téléchargement.
- Les différents composants du logiciel i2b2 sont disponibles en téléchargement pour Windows ou pour Mac²⁹.

PROCESSUS DE PRISE DE DÉCISION SUR LE STANDARD

- Il existe **3 groupes de travail, accessibles à tous** :
 - « **ETL Working Group** » : il présente les meilleures pratiques pour le chargement de données dans i2b2³⁰.
 - « **Ontology Working Group** » : il présente de la documentation, des outils et des tutoriels pour aider les utilisateurs à comprendre, modifier et construire une ontologie³¹.
 - « **User Interface Working Group** » : ce groupe de travail réunit à la fois les utilisateurs d'i2b2, de tranSMART et des plateformes i2b2-tranSMART. Il a pour objectif de collecter les problèmes et suggestions des utilisateurs, et de les transmettre aux concepteurs des interfaces i2b2 et tranSMART³².
- Les problèmes rencontrés par les utilisateurs et les suggestions peuvent être remontés via i2b2 Bug Tracker (une inscription est nécessaire)³³.

27) [i2b2 Software License](#)

28) [i2b2 Software](#)

29) [2.6 i2b2 Requirements](#)

30) Ce groupe de travail se réunit tous les deuxièmes mardis du mois de 10h EST à 11h EST : [ETL Working Group](#)

31) Ce groupe de travail se réunit tous les troisièmes jeudis de chaque mois de 11h EST à 12h EST : [Welcome to the Ontology Working Group space!](#)

32) Ce groupe de travail se réunit tous les troisièmes mercredis de chaque mois de 12h EST à 13h EST. Voir : [the i2b2 tranSMART User Interface Working Group space!](#)

33) [The i2b2 data repository](#) et [Log in - i2b2 JIRA](#)

MATURITÉ DU STANDARD

- **Fréquence de mise à jour :**
 - La version actuelle est la **v1.7.13** sortie en juin 2022³⁴, **compatible avec SHRINE**³⁵ (voir l'élément « Outils compatibles » en partie 5. Utilisation).
 - Depuis la version v1.3 sortie en 2008³⁶, 4 versions principales (v1.4, v1.5, v1.6) et 31 mises à jour mineures³⁷ ont été publiées.
 - Entre 2008 et 2013, une version principale est sortie **presque tous les ans en moyenne**, à une fréquence irrégulière : la v1.3 est sortie en 2008, la v1.4 et la v1.5 sont sorties en 2010 (janvier et juillet), la v1.6 est sortie en 2011, la v1.7 est sortie en 2013.
 - Depuis 2013, il n'y a pas eu de nouvelle version principale mais uniquement des mises à jour.

MATURITÉ

- **Note : 0,5 / 1.**
- Cette note combine plusieurs sous-critères. Elle s'explique par :

- a) La publication d'une version stable.
- b) La fréquence moyenne faible de mise à jour des versions principales.
- c) L'âge élevé du standard.
- d) Son utilisation dans le monde restreinte à l'échelle d'institutions (voir l'élément « Adoption du standard » en partie 4. Valorisation).
- e) L'absence d'adoption officielle par un ou plusieurs pays ou par une organisation de référence (voir l'élément « Adoption du standard » en partie 4. Valorisation).

EXISTENCE DE FINANCEMENTS POUR STANDARDISATION

En mai 2023, nous n'identifions pas de financements pour la standardisation de données vers i2b2.

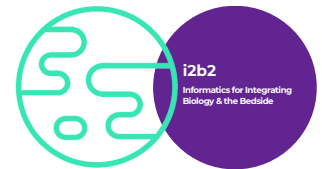
34) [All Releases](#)

35) Il existe une table de compatibilité entre les versions d'i2b2 (à partir de la v1.7.05) et les versions de SHRINE (à partir de la version v1.19.2). Voir la matrice de compatibilité d'i2b2 et de SHRINE : <https://open.catalyst.harvard.edu/wiki/display/SHRINE/SHRINE-i2b2+Compatibility+Matrix>

36) [1.3 Versions](#)

37) Liste des mises à jour mineures : 2 mises à jour pour la v1.3, 2 mises à jour pour la v1.5, 8 mises à jour pour la v1.6 et 19 mises à jour pour la v1.7. Consulter l'historique des versions d'i2b2 (v1.3 à v1.6) : <https://community.i2b2.org/wiki/display/RM/1.6+and+Older+Release+Notes>, [Earlier 1.7.x Release Notes](#) (v1.7.01 à v1.7.08b) et <https://community.i2b2.org/wiki/display/RM/Latest+Release+Notes> (v1.7.09c à v1.7.13)

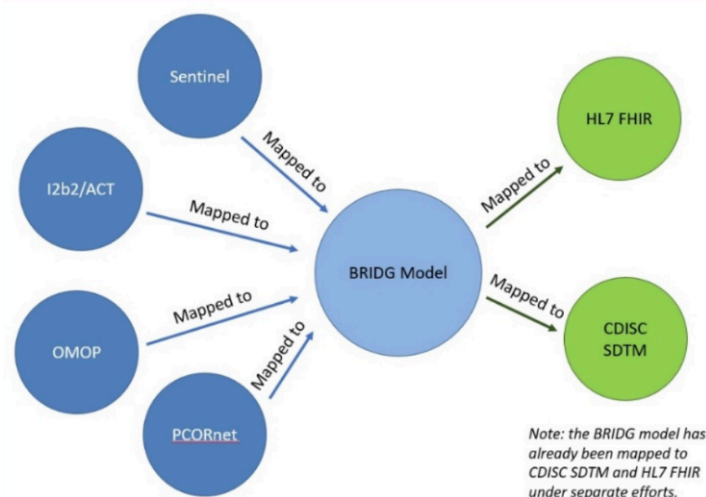
3. Technique



CAPACITÉ DE TRADUCTION VERS UN AUTRE STANDARD (INTRA TYPES DE STANDARDS)

- Le projet Common Data Model Harmonization (CDMH)³⁸ vise à harmoniser les 4 schémas de données suivants : **PCORnet, OMOP-CDM, i2b2 et Sentinel**.
 - L'objectif est d'avoir un outil d'accès unifié à ces données permettant aux chercheurs d'accéder à un réseau plus large de patients et à des données variées (EHR, demandes de remboursement, données issues des essais cliniques, ...).
 - Le projet consiste à mapper chacun de ces schémas de données vers le modèle intermédiaire BRIDG v3.2³⁹ (voir Figure 3 ci-dessous). Ce modèle a été choisi comme modèle intermédiaire car il a été mappé dans une première étape à FHIR et CDISC SDTM.

Figure 3 : Procédure de mapping



Source : [Common Data Model Harmonization \(CDMH\) and Open Standards for Evidence Generation](#)

Il existe un processus de transformation depuis i2b2 vers OMOP-CDM par les scripts SQL développés dans le cadre du projet ARCH-OMOP^{40,41}.

38) Page HL7 décrivant le projet : [HL7.FHIR.US.CDMH\IG Home Page](#) et le rapport final « Common Data Model Harmonization (CDMH) and Open Standards for Evidence Generation », U.S Food & Drug Administration, NIH, The Office of the National Coordinator for Health Information Technology, 2020 ([Common Data Model Harmonization \(CDMH\) and Open Standards for Evidence Generation](#))

39) Page d'implémentation : [Reference Implementations | Biomedical Research Integrated Domain Group](#)

40) Article de Klann JG, Joss MAH, Embree K, Murphy SN (2019) « Data model harmonization for the All Of Us Research Program: Transforming i2b2 data into the OMOP common data model ». PLoS ONE (2019) : [Data model harmonization for the All Of Us Research Program: Transforming i2b2 data into the OMOP common data model | PLoS ONE](#)

41) GitHub : [GitHub - i2b2-omop/i2o-transform: PCORnet Ontology to OMOP - beta!](#)

COMMUNICATION AVEC D'AUTRES STANDARDS (INTER TYPOLOGIES DE STANDARDS)

- Klann et al. (2016)⁴² ont développé, au sein d'un réseau de 12 établissements aux États-Unis, un processus permettant de générer une base de données **PCORnet CDM** directement à partir de systèmes i2b2 déjà existants.
 - Cette méthodologie a pu être validée par 8 des 12 sites concernés, ce qui a permis de constituer un réseau de 10 millions de patients.
 - La méthodologie est **efficace** : elle construit un réseau qui peut s'appuyer sur les fonctionnalités analytiques d'i2b2 et de PCORnet, **sans avoir à développer un processus ETL supplémentaire**. Les 8 établissements peuvent en effet :
 - Exécuter des requêtes fédérées dans SHRINE (voir « Outils compatibles » en partie 4. Valorisation).
 - Participer à l'analyse distribuée réalisée au niveau national sur données retrospectives à travers PCORnet.
- Dans le contexte de la réflexion sur la standardisation, il est important de préciser qu'i2b2 spécifie avant tout un modèle de données compact et abstrait pour le stockage des données qui concernent les individus.
 - Ce modèle présente plusieurs avantages :
 - Grâce à sa formalisation en étoile, il est centré sur les observations : il concentre toutes les informations concernant un patient ayant eu un contact avec une structure à une date donnée, fournie par une entité identifiable.
 - Egalement, il présente très peu de contraintes, ce qui facilite l'intégration des données et lui donne un avantage concurrentiel.
 - Cependant, on note un certain nombre de limites :
 - La gestion du texte libre est sous-optimale (index full text des SGBD) et nécessite souvent de mettre en place des solutions spécifiques pour ces données (solr, elastic).
 - La gestion du lien entre plusieurs observations est complexe : la modification de celles-ci permet seulement l'ajout d'un niveau de relation, ce qui peut poser problème pour certains usages.
 - Enfin, dans des structures ayant une certaine masse critique, le volume des observations implique une gestion très fine des indexes, du partitionnement et des paramètres de la base de données afin de conserver des performances raisonnables lors de l'interrogation - cela est possible avec ce standard, néanmoins des compétences poussées en administration / optimisation de bases de données sont requises pour un passage à l'échelle.
 - Cette difficulté est également retrouvée avec le modèle OMOP.

42) Voir l'article de Klann JG, Abend A, Raghavan VA, Mandl KD, Murphy SN. « Data interchange using i2b2. » J Am Med Inform Assoc. 2016 Sep : [Data interchange using i2b2](#)

FLEXIBILITÉ DANS LES CHOIX DES TERMINOLOGIES

- La plateforme i2b2 utilise plusieurs terminologies standards pour représenter ses concepts (voir Tableau 2 ci-dessous)^{43,44} :
 - CIM-9 ou CIM-10 pour les maladies.
 - LOINC pour les tests de laboratoire.
 - NDC (National Drug Codes : nomenclature des médicaments aux États-Unis) pour les médicaments.

Tableau 2 : Quelques terminologies standard utilisées dans i2b2

Data Domains	Typical Standards
Demographics	HL7 Administrative
Diagnoses	ICD
Procedures	ICD, CPT, HCPCS ⁴⁵
Medications	RxNorm, VA Classes hierarchy, NDC
Labs	LOINC
Vital Signs	LOINC

Source : [Documentation i2b2](#)

- Cependant, il est possible d'utiliser d'autres terminologies grâce au concept EAV⁴⁶ (Entity-Attribute-Value), voir l'élément « Flexibilité du standard, personnalisation » ci-après). Il est par exemple possible de combiner l'utilisation de codes SNOMED CT et LOINC⁴⁷. En effet, il est relativement simple de passer d'un modèle EAV à un autre (par exemple, d'i2b2 vers OMOP) ; les enjeux concernent principalement le mapping vers les terminologies standards.

43) [Ontologies 101](#)

44) [Ontologies | i2b2](#)

45) [HCPCS \(Healthcare Common Procedure Coding System\) est une collection de codes standardisés représentant des procédures médicales, des fournitures. Ces codes sont produits par les CMS \(Centers for Medicare and Medicaid Services\) et sont utilisés pour faciliter le traitement des demandes d'assurance maladie par Medicare entre autres.](#)

46) [Explication du concept EAV \(avantages et inconvénients\) : 711. Background](#)

47) [Employing SNOMED CT and LOINC to make EHR data sensible and interoperable for clinical research](#)

- Points d'attention :
 - Les terminologies dans i2b2, qui permettent de capturer la signification d'une observation, se trouvent dans un schéma dédié et nécessitent d'alimenter des tables de métadonnées spécifiques.
 - i2b2 ne fournit pas de terminologie standard, hormis pour le jeu de données d'exemple, ou le contexte de réseaux spécifiques : c'est à l'intégrateur de données de déterminer les terminologies (locales et standards) qu'il utilisera et de les implémenter dans l'application.
 - Cette implémentation est relativement simple mais nécessite une courbe d'apprentissage.
 - Par ailleurs, dans le contexte d'un EDS hospitalier, les terminologies locales peuvent être de grandes dimensions, ce qui complique l'implémentation et rend la tâche de mapping vers les standards complexe et chronophage.

FLEXIBILITÉ DU STANDARD, PERSONNALISATION

- Le schéma de la plateforme i2b2 est fondé sur le **concept EAV** (Entity-Attribute-Value) permettant d'**ajouter un attribut** qui n'existe pas dans l'ontologie par défaut d'i2b2. Pour cela, il suffit d'ajouter les valeurs de cet attribut dans la table de dimension concernée et les lignes de données dans la table de faits.
 - Par exemple, pour utiliser SNOMED CT, qui ne fait initialement pas partie des terminologies standards d'i2b2, il faut ajouter les codes SNOMED CT dans la table `concept_dimension` et ajouter les lignes d'observation dans la table `observation_fact`. Dans le cas de SNOMED CT, et plus généralement dans le cas des terminologies, le lien entre les deux tables se fait par le champ « `concept_cd` ».
- Ainsi, de **nouveaux types d'observations** (concernant par exemple des symptômes, des instruments utilisés, etc.) peuvent être ajoutés à i2b2 par les institutions, en étendant simplement l'ontologie¹⁶, sans modifier la structure de la base de données ou du logiciel (voir exemple ci-dessus). Cela permet la création d'outils d'interrogation des données, de visualisation et d'analyse qui peuvent être généralisés à différents types de données (ex : codes locaux utilisés dans les établissements de santé).
- i2b2 est un modèle performant pour l'intégration de données issues de plusieurs sources, en revanche celui-ci ne peut pas être perçu comme un standard de partage de données à l'échelle de plusieurs structures (car il ne contraint pas le vocabulaire et les terminologies nativement).
 - i2b2 peut devenir un standard dès lors que l'on définit en sus du modèle de données une ou plusieurs terminologie(s) cible(s), que l'on implémente dans i2b2/shrine sur le modèle de PCORNET et 4CE par exemple.

PERFORMANCE

- **Note : 1 / 1** en raison du faible nombre de jointures nécessaires pour réaliser des requêtes (voir ci-dessous).
- La plateforme i2b2 utilise le **concept EAV** (Entity-Attribute-Value) pour optimiser le schéma de données ⁴⁷.
 - Grâce à ce concept EAV :
 - Les **schémas des données** n'ont **pas besoin d'être modifiés** pour ajouter des nouveaux types d'observations (voir l'élément « Flexibilité du standard, personnalisation » ci-dessus). Cela permet de restreindre le nombre de jointures nécessaires pour effectuer l'analyse des données. Pour la plupart des requêtes, trois tables principales sont utilisées (observation_fact, patient_dimension, concept_dimension).
 - Dans certains cas, le nombre de jointures utilisées dans les requêtes pour analyser les données peut aller jusqu'à 5 (en ajoutant des jointures avec visit_dimension, observer_dimension et modifier_dimension).
 - Cependant ce concept EAV peut conduire à ⁴⁷ :
 - Une possible **explosion de la table de faits** (observation_fact). En effet, dans le cas d'un schéma EAV, tous les faits sont stockés dans une base unique, tandis que dans d'autres modèles, des tables individuelles peuvent être ajoutées.

COMPLEXITÉ DU MODÈLE

- **Note : 0,2 / 1.** Cette note combine plusieurs sous-critères. Elle s'explique par :
 - a. Le faible nombre de tables avec des liens entrants et sortants.
 - b. Le faible niveau de normalisation.
 - c. L'absence de tables larges.
- Le schéma de données i2b2 n'est **pas normalisé**, c'est-à-dire qu'il ne contient pas de table de dimension pour toutes les entités. Par exemple, il ne contient pas de table de dimension pour le statut marital (voir Figure 4 ci-dessous). Par conséquent, le **niveau de redondance d'information est élevé**.

Figure 4 : Exemple de redondance d'information

RACE_CD	MARITAL_STATUS_CD	RELIGION_CD	ZIP_CD	STATECITYZIP_PATH
black	married	roman catholic	02140	Zip codes\Massachusetts\Cambridge\02140\
white	married	roman catholic	02142	Zip codes\Massachusetts\Cambridge\02142\
asian	widow	agnostic	01821	Zip codes\Massachusetts\Billerica\01821\
black	widow	atheist	02303	Zip codes\Massachusetts\Brockton\02303\
hispanic	single	roman catholic	01889	Zip codes\Massachusetts\North Reading\01889\
black	divorced	atheist	01889	Zip codes\Massachusetts\North Reading\01889\
asian	single	muslim	01061	Zip codes\Massachusetts\Northampton\01061\
hispanic	divorced	roman catholic	02303	Zip codes\Massachusetts\Brockton\02303\
hispanic	single	christian	02124	Zip codes\Massachusetts\Boston\02124\
hispanic	married	atheist	02140	Zip codes\Massachusetts\Cambridge\02140\
white	widow	roman catholic	02152	Zip codes\Massachusetts\Winthrop\02152\

Source : [Traitement Veltys](#)

- Le schéma de données i2b2 ne contient que **10 tables (132 champs)**. Par ailleurs, la plupart des requêtes utilise seulement **deux jointures principales** entre les tables observations_fact, patient_dimension et concept_dimension. En effet, les tables patient_dimension et concept_dimension contiennent la plupart des critères d'analyse (l'âge des patients, le sexe, les maladies, etc.). Enfin, le schéma des données ne contient **pas de relations physiques** entre les tables, c'est-à-dire qu'il ne définit pas de clés étrangères. Cela réduit la complexité du modèle, mais ne permet pas d'utiliser de mécanisme interne de base de données pour vérifier l'intégrité référentielle.

CONTRAINTES D'IMPLEMENTATION

- La plateforme i2b2 n'impose pas de contrainte d'implémentation. L'un des trois SGBDs suivants peut être utilisé⁴⁸ :
 - Oracle (v12 ou 18).
 - PostgreSQL (9.x).
 - MS SQL Server (de 2012 à 2019).
- Le i2b2 Hive comprend plusieurs cellules qui représentent chacune une unité fonctionnelle. Ces cellules sont installées sur un ou plusieurs « serveurs i2b2 ». L'environnement d'exécution Java 8 est nécessaire pour déployer les cellules centrales (« Core Cells ») sur le ou les serveurs i2b2⁴⁹.
- L'application Web Client peut être hébergée par tous les serveurs web qui supportent HTML, JavaScript, CSS et les formats d'image GIF / JPG / PNG. Le composant suivant doit cependant être installé sur le serveur web⁵⁰: PHP Hypertext Preprocessor (PHP).

TYPE DE TECHNOLOGIE DE REQUÊTAGE

Plusieurs types de technologies de requêtage peuvent être utilisés :

- **i2b2 « Query & Analysis Tool »**¹²: outil graphique de requêtage intégré à i2b2.
- Requêtes **SQL**.
- **Interfaces de programmation d'application (API)**⁵¹.

NEUTRALITÉ TECHNOLOGIQUE

Pour l'installation native, Java 8⁵² est obligatoire pour pouvoir installer i2b2.

- Il existe la version containerisée (créée par la communauté) pour laquelle Docker Compose est obligatoire⁵³.

48) [2.2 Database Requirements](#)

49) [2.4.1 Java JDK](#)

50) [2.5 Web Server Requirements](#)

51) <https://community.i2b2.org/wiki/display/ServerellsMessagingHome/>

52) [Quick Installation for 1.8.0](#)

53) [docker images for core i2b2 container](#)

INTENSITÉ DE LA PERTE DE DONNÉES AU MAPPING

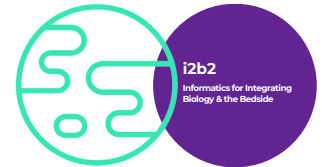
- Mate et al. (2011)⁵⁴ ont développé un processus ETL (Extract Transform Load) pour charger les données stockées dans les EMR (Electronic Medical Records) des établissements de santé dans la base de données i2b2.
 - Leur approche n'utilise pas de programme d'import/export propriétaires, ni de code SQL, mais un langage générique flexible (Web Ontology Language).
 - Dans le cadre de la construction de cette méthodologie, ils ont développé plusieurs briques qui peuvent être utilisées pour faciliter la mise en place d'un processus ETL (OntoEdit, OntoGen, OntoExport).
 - Ils ont implémenté cette méthodologie dans deux établissements de santé en Allemagne : 75 % des données des EMR ont pu être mappées à i2b2.

COMPÉTENCES TECHNIQUES ET MÉTIER NÉCESSAIRES POUR UTILISER LE STANDARD

- Un profil Data Engineer est nécessaire pour transformer les données en respectant le concept EAV.
- Les profils optionnels suivants sont également intéressants :
 - Database administrator pour l'optimisation des SGBDs.
 - DevOps pour l'installation/la gestion d'i2b2 containerisé.

54) Voir l'article de Mate S., Bürkle T., Köpcke F., Breil B., Wullich B., Dugas M., Prokosch H., Ganslandt T. « Populating the i2b2 Database with Heterogeneous EMR Data : a Semantic Network Approach ». Stud Health Technol Inform, 2011 : <https://ebooks.iospress.nl/publication/14218>

4. Valorisation



ACCESSIBILITÉ À DES RESSOURCES DE FORMATION

- **Vidéos** sur YouTube⁵⁵ et en particulier les vidéos de la chaîne de i2b2 tranSMART Foundation⁵⁶.
- **Vidéos** Coursera⁵⁷.
- Le CCTS (Center for Clinical and Translational Science) propose des **vidéos de formation** sur i2b2⁵⁸.
- **GitHub** d'i2b2²¹.

DISPONIBILITÉ DE LA DOCUMENTATION SCIENTIFIQUE DÉMONTRANT L'INTÉRÊT

- En mai 2023, **environ 439 articles**⁵⁹ traitant d'i2b2 **entre 2010 et 2023 sont disponibles sur PubMed** (voir la requête et un extrait de la liste des publications en « Annexe n°1 »).
- Selon Murphy et Wilcox (2014)⁶⁰, i2b2 est devenu un **outil central pour les chercheurs** dans le domaine de la santé. Cette plateforme open source repose sur une communauté de développeurs qui améliorent sans cesse les capacités analytiques et inventent de nouvelles fonctionnalités.
- Majeed et al. (2021)⁶¹ soulignent l'intérêt d'i2b2 et notamment de sa **plateforme utilisateur simple et intuitive**, permettant des analyses de données :
 - Si le modèle OMOP-CDM est souvent choisi pour la structuration des données médicales, il ne dispose pas d'un outil d'interrogation des données permettant aux utilisateurs non-formés de faire des requêtes simples.
 - Les auteurs développent un algorithme permettant de traduire des requêtes créées dans le Web Client i2b2 en des requêtes SQL pouvant être exécutées sur une base OMOP-CDM.

55) https://www.youtube.com/results?search_query=i2b2

56) [i2b2 tranSMART Foundation - YouTube](#)

57) <https://www.coursera.org/lecture/clinical-data-models-and-data-quality-assessments/a-quick-tour-of-a-common-data-model-i2b2-Fjbppp>

58) [i2b2 - Center for Clinical and Translational Science | UAB](#)

59) L'estimation du nombre d'articles traitant d'i2b2 sur PubMed a été faite en saisissant le terme « i2b2 » dans la barre de recherche du site PubMed (<https://pubmed.ncbi.nlm.nih.gov/>). Voir en « Annexe n°1 » pour la méthodologie utilisée et un extrait de la liste des publications.

60) Article de Murphy S, Wilcox A. « Mission and Sustainability of Informatics for Integrating Biology and the Bedside (i2b2) ». EGEMS (Wash DC). 2014 Sep : [Mission and Sustainability of Informatics for Integrating Biology and the Bedside \(i2b2\) - PMC](#)

61) Article de Majeed RW, Fischer P, Günther A. « Accessing OMOP Common Data Model Repositories with the i2b2 Webclient - Algorithm for Automatic Query Translation ». Stud Health Technol Inform. 2021 : [IOS Press Ebooks - Accessing OMOP Common Data Model Repositories with the i2b2 Webclient - Algorithm for Automatic Query Translation](#)

ADOPTION DU STANDARD

- Adoptions officielles : pas d'adoption officielle identifiée en date de mai 2023.
- Utilisation sur le marché : plus de **250 institutions** dans le monde utilisent i2b2⁶² :
 - Des **universités et des instituts de recherche** :
 - À l'international :
 - Aux États-Unis : Mayo Clinic⁶³, universités (Johns Hopkins, Columbia, Harvard, Stanford, ...).
 - Universités en Autriche, Norvège, Suède, Finlande, Islande, Italie, Espagne, Japon, Allemagne, Arabie Saoudite, Corée, etc.
 - En France :
 - Le Cancéropôle Île-de-France (INCa)⁶⁴ met à disposition de ses établissements une interface i2b2 SHRINE, déjà adoptée par l'Institut Curie, Hôpital Européen Georges Pompidou et AP-HP (i2b2 va être remplacé par une solution développée en interne⁶⁵).
 - Inserm (Rennes).
 - Des **organismes d'assurance maladie** (Group Health Cooperative, Kaiser Permanente).
 - Des **entreprises** : BIOMERIS, Johnson and Johnson, TriNetX, Partner HealthCare⁶⁶, etc.

FOURNISSEURS DE SERVICE AYANT L'EXPERTISE EN FRANCE

Nous n'identifions pas de fournisseur de service ayant l'expertise en France en mai 2023.

62) [i2b2 Installations](#) et [i2b2 Community Wiki](#)

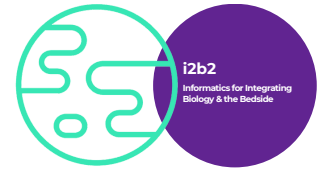
63) [i2b2: Informatics for Integrating Biology and the Bedside - Center for Clinical and Translational Science](#)

64) Le cancéropôle d'Île-de-France est un réseau d'institutions franciliennes impliquées dans la recherche sur le cancer, composé des 7 institutions suivantes : Assistance Publique – Hôpitaux de Paris (AP-HP), Fondation Jean Dausset – CEPH, Gustave Roussy, Institut Curie, Institut Universitaire d'hématologie – Université Paris-Diderot, Institut Pasteur, Sorbonne Université. Le cancéropôle met à disposition une interface i2b2 (SHRINE) pour déterminer des tailles de cohortes. La solution est déjà installée à l'hôpital européen Georges Pompidou, à l'AP-HP et à l'Institut Curie. [Application i2b2/SHRINE - Cancéropôle Île-de-France](#)

65) [i2b2 & Cohort360 | Entrepôt de Données de Santé](#)

66) Article de Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. « Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2) ». J Am Med Inform Assoc. 2010 Mar-Apr : [Serving the enterprise and beyond with informatics for integrating biology and the bedside \(i2b2\) - PMC](#)

5. Utilisation



SIMPLICITÉ D'USAGE

- **Note : 0,7 / 1**

- Cette note combine plusieurs sous-critères. Elle s'explique par :

- a)** L'accès à des ressources officielles de formation (voir l'élément « Accessibilité à des ressources de formation » en partie 4. Valorisation).
- b)** La lisibilité du schéma par un humain (voir l'élément « Lisible par un humain » en partie 5. Utilisation).
- c)** L'absence d'outils de gestion de la qualité des données (voir l'élément « Qualité des données » en partie 4. Valorisation).
- d)** Le nombre relativement faible de compétences requises pour l'implémentation et l'usage (voir l'élément « Compétences techniques et métier nécessaires pour utiliser le standard » en partie 3. Technique).

- Ainsi, l'usage de la plateforme i2b2 est relativement simple, son installation peut se révéler cependant plus complexe :
 - D'après Majeed et al. (2021)⁶⁵, le **i2b2 « Query & Analysis Tool »** et ses **plug-ins** associés sont simples d'utilisation et permettent d'interroger facilement les données.
 - Waghlikar et al. (2018)⁶⁷ expliquent que l'installation d'i2b2 (du Hive) peut être difficile en raison de l'infrastructure complexe d'i2b2, et ce malgré l'existence d'une documentation riche et d'une communauté active. Pour cette raison, ils ont développé un package d'installation automatisé, appelé i2b2-quickstart, qui télécharge automatique la dernière version d'i2b2 et ses dépendances et configure les cellules i2b2 de manière à créer une installation fonctionnelle⁶⁸.

EXISTENCE D'UNE COMMUNAUTÉ EN LIGNE ET DEGRÉ D'ACTIVITÉ

- Il existe plusieurs **groupes de travail** (voir l'élément « Processus de prise de décision sur le standard » dans la partie 2. Gouvernance). Pour rejoindre ou pour créer un nouveau groupe de travail, le remplissage d'un formulaire est nécessaire.⁶⁹
- **Réunions bimensuelles** de la **communauté i2b2 tranSMART**⁷⁰ (présentation des nouvelles fonctionnalités, informations sur les groupes de travail, ...).

67) Article de Waghlikar KB, Mendis M, Dessai P, Sanz J, Law S, Gilson M, Sanders S, Vangala M, Bell DS, Murphy SN. « Automating Installation of the Integrating Biology and the Bedside (i2b2) Platform ». Biomed Inform Insights. 2018 Jun 4 : [Automating Installation of the Integrating Biology and the Bedside \(i2b2\) Platform - PMC](#)

68) Page GitHub : [GitHub - i2b2/i2b2-quickstart: Quick automated installation of i2b2 in VMs, Amazon-Web-Service and Docker](#)

69) <https://i2b2transmart.org/working-groups-2/get-involved-contact-us/>

70) Agenda des meetings à venir : [i2b2 tranSMART Community Meeting](#). Il est possible de visionner les enregistrements des sessions passées.

OUTILS DE MAPPING

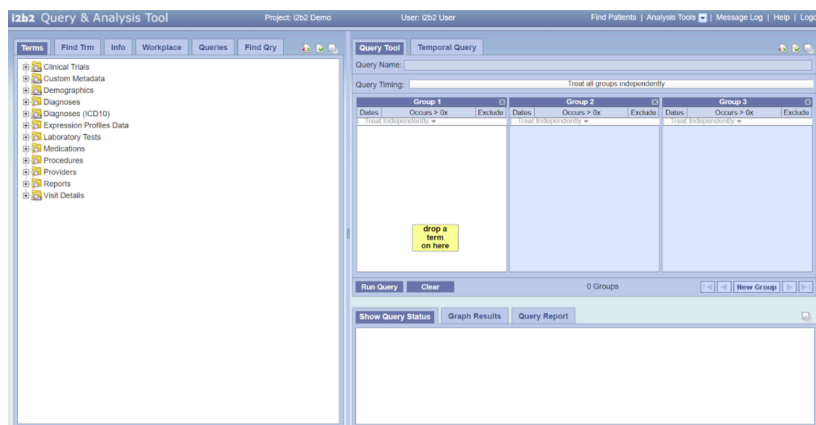
- **Mapping sémantique** : **i2b2 Workbench**⁷¹ possède un Mapping Tool pour les ontologies. Il permet de vérifier, modifier, ajouter des mappings entre deux ontologies (par exemple en ICD-9 et ICD-10).

OUTILS COMPATIBLES

Les outils compatibles avec i2b2 sont les suivants¹² :

- **« Query & Analysis Tool » (on parle aussi de l'application Web Client)** : outil de requêtage pour identifier des cohortes de patients (voir Figure 5 ci-dessous).

Figure 5 : Page d'accueil du « Query & Analysis Tool »



Source : démonstration du Web Client ([i2b2 Web Client](#))

- Plug-ins associés au « Query & Analysis Tool » :
 - **« ExportXLS »** : permet d'exporter sous forme de fichier .csv ou .xls les concepts sélectionnés et observés sur un ensemble de patients, sous forme de tableau (voir Figure 6 ci-dessous).

Figure 6 : Exemple d'utilisation du plug-in ExportXLS

Click on one of the buttons on the right to download the following table in the appropriate format. [CSV Export](#) [HTML/XLS Export](#)

Patient ID	ARTERIAL BLOOD GAS PCO2 (LOINC:2019-8)	Operations on cardiovascular system
1	1000000025	ICD9:36.01 (1) ICD9:36.07 (2)
2	1000000045	ICD9:36.01 (1) ICD9:36.06 (2)
3	1000000055	LOINC:2019-8 = 79.00000 mmHg.
4	1000000056	ICD9:36.99 (2)
5	1000000056	ICD9:36.01 (1) ICD9:36.06 (2) ICD9:37.22 (1)

Patient Information (for patients 1 - 5)
of Patient Set 'Femot@17:44:12 [3-3-2023] [demo] [PATIENTSET_58258]'

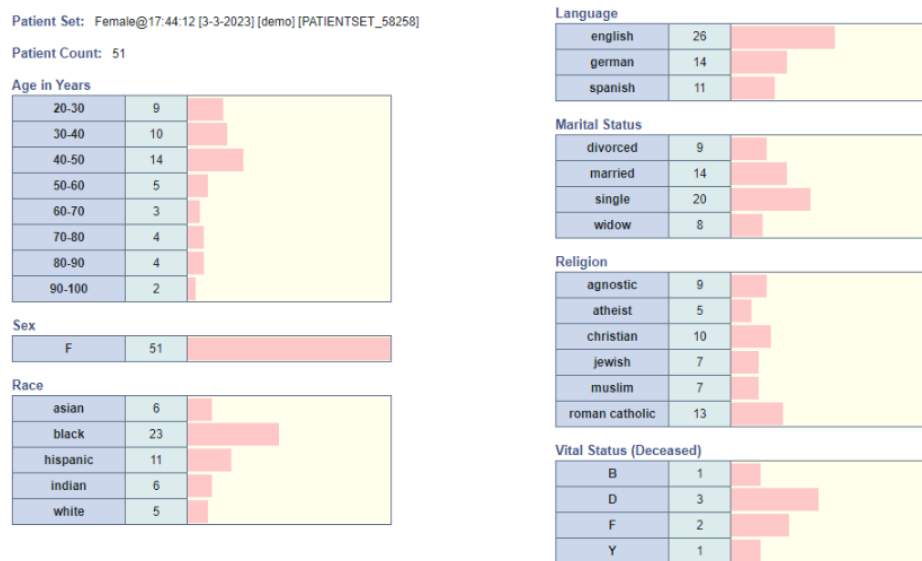
NOTE: only patients with any of the selected concepts are included above.

Source : démonstration du Web Client ([i2b2 Web Client](#))

71) [Mapping tools version 1.1](#)

- « **Timeline** » : permet de créer une représentation visuelle du moment où les concepts sélectionnés sont observés dans un ensemble de patients.
- « **Demographics (1 Patient Set) – Simple Counts** » : permet d'afficher les informations démographiques pour un ensemble de patients (voir Figure 7 ci-dessous).

Figure 7 : Exemple d'utilisation du plug-in Demographics (1 Patient Set) (pour la population Femmes)

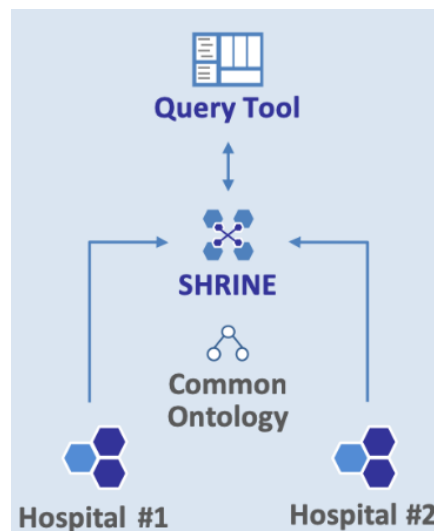


Source : démonstration du Web Client ([i2b2 Web Client](#))

- « **Demographics (2 Patient Sets) – Simple Counts** » : permet de comparer les informations démographiques de 2 ensembles de patients.
- « **WISE (Working Items Sharing Enhancement) – Searcher** » : permet de faciliter la recherche d'éléments ou d'objets dans le panneau Workplace du Web Client i2b2.
- « **CARE (Cohort Analysis & Refinement Expeditor) – Concept Demographics Histograms** » : permet de générer des histogrammes comparatifs des répartitions démographiques pour un sous-ensemble de patients par rapport aux patients de ce sous-ensemble associés à des concepts spécifiques.
- « **CARE – Concept Observation Tally Demographics Histograms** » : permet de générer des histogrammes comparatifs des répartitions démographiques pour un sous-ensemble de patients par rapport aux patients de ce sous-ensemble associés à plusieurs observations d'un concept spécifié.
- « **Communicator Tool** » : permet d'interagir directement avec les objets standard de Cell Communicator depuis la plateforme.

- **SHRINE**⁴ : outil qui permet de relier les systèmes i2b2 provenant de différents sites ou établissements avant l'utilisation du « Query & Analysis Tool » (voir Figure 8 ci-dessous). Cet outil permet ainsi de compter le nombre de patients qui vérifient certains critères dans les établissements participants. Les chercheurs l'utilisent pour inclure un nombre suffisant de patients dans leurs études⁷².

Figure 8 : La fédération de systèmes i2b2 par SHRINE



Source : [Software – i2b2 tranSMART Foundation](#)

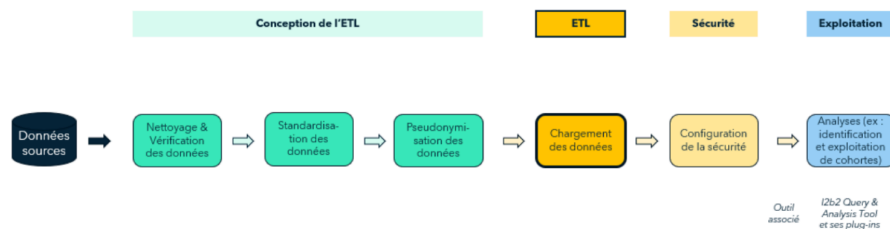
- **tranSMART**⁴ : suite d'outils d'exploration et de visualisation de données, d'analyses génomiques et d'ETL développés par des industries pharmaceutiques pour des avancées en recherche translationnelle.
- Waghlikar et al. (2017)⁷³ ont développé une interface permettant d'implémenter **SMART on FHIR à partir d'une plateforme i2b2** :
 - La plateforme i2b2 est utilisée dans une approche « sidecar » : le logiciel stocke une copie des données patient issues de l'EHR et exécute des requêtes sur ces données pour un usage secondaire de recherche (en parallèle de l'EHR qui continue d'être utilisé pour un usage clinique).
 - SMART on FHIR permet de transformer un EHR (ou bien son « sidecar » dans le cas d'i2b2) en une plateforme de type « App store pour la santé » où les utilisateurs ont accès à de nombreuses applications.
 - L'implémentation de SMART on FHIR à partir d'i2b2 permet ainsi (1) de faciliter le déploiement d'applications SMART, (2) d'avoir un mécanisme additionnel d'accès et de requêtage de la plateforme i2b2 et (3) de migrer les données depuis l'EHR, en passant par la réplication dans i2b2 jusqu'à une base FHIR-compatible.

73) Article de Waghlikar KB, Mandel JC, Klann JG, Wattanasin N, Mendis M, Chute CG, Mandl KD, Murphy SN. « SMART-on-FHIR implemented over i2b2 ». J Am Med Inform Assoc. 2017 Mar : [SMART-on-FHIR implemented over i2b2 - PMC](#)

SIMPLICITÉ D'USAGE

- Pour standardiser une base de données au format i2b2, il faut créer un processus ETL (Extract Transform Load).
- Processus de standardisation (ETL) :
 1. **Nettoyage et vérification des données** : cette étape vise à s'assurer de la qualité des données notamment en éliminant les redondances, les données manquantes ou erronées.
 2. **Standardisation des données (utilisation de l'ontologie d'i2b2 pour harmoniser la terminologie employée, utilisation de codes standards, etc.)** : il est possible de définir d'autres ontologies en cas de besoin.
 3. **Pseudonymisation des données** : pendant cette étape, des identifiants aléatoires sont attribués à chaque patient et une table de mapping associée est générée.
 4. **Chargement des données en base (Oracle, SQL Server ou PostgreSQL)** : dans cette étape, les données sont transformées en un schéma en étoile compatible avec i2b2.
 5. **Configuration de la sécurité et contrôle d'accès** :
 - a. Les utilisateurs ont besoin de s'authentifier pour accéder aux données.
 - b. En fonction du rôle de l'utilisateur, l'utilisateur ne peut accéder qu'aux catégories auxquelles il a été autorisé à accéder.
 - c. Il est possible de chiffrer certains champs pour gérer la confidentialité des données (Patient Notes, Observation_Blob).

Figure 9 : Processus ETL utilisé pour charger les données dans i2b2



Source : Schéma réalisé par Veltys

EXISTENCE D'EXTENSIONS CERTIFIÉES :

Plusieurs extensions existent :

- **i2b2-quickstart**⁶⁹ : facilite l'installation de la plateforme i2b2.
- **i2b2-etl** : facilite le chargement des données vers la base de données de i2b2.

BIBLIOTHÈQUE DE REQUÊTES TYPES

La page d'aide du **i2b2 « Query & Analysis Tool »** contient des exemples de requêtes qui sont accessibles dans la version de démonstration de l'outil¹²(voir Figure 10 ci-dessous).

Figure 10 : Extrait de la page d'aide de l'outil « Query & Analysis Tool »

i2b2 Web Client Help

- 1. Quick Start
- 2. Webclient Layout
- 3. Query Tool
 - 3.1 Layout
 - 3.2 Creating a Query
 - 3.3 Running a Query
- 4. Navigate Terms
 - Navigate Terms Tab
 - Find Tab
- 5. Query Status
 - Show Query Status
 - Graph Results
 - Query Report
- 6. Previous Queries
- 7. Workplace
- 8. Other Operations

[Example]

Find all patients who have at least one diagnosis of *Acute respiratory infections* .

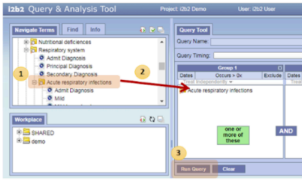
1. Find *Acute respiratory infections*
 1. Go to 'Navigate Terms' panel
 2. Select *Diagnosis* → *Respiratory system* → *Acute respiratory infections*
2. Drag and drop
 1. Drag the *Acute respiratory infections* term into 'Query Tool > Group 1'
3. Run Query
 1. Click 'Query Tool > Run Query' → 'OK'
4. Get Results
 1. See the 'Graph Results' tab for the query result
 2. The query is now stored in the 'Previous Queries' panel

1.2 Adding an AND term

Terms in different groups have a logical **AND** relationship. When terms are combined with the **AND** operator, retrieved records must contain all the terms.

[How to]

1. Find a Term
 1. Go to 'Navigate Terms' tab
 2. Click term folders to narrow down the concept
 3. Select a term
2. Drag and drop to 'Group 1'
 1. Drag the selected term into 'Query Tool > Group 1'

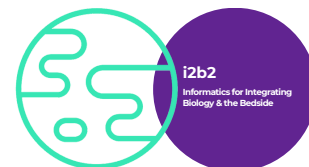


Source : démonstration du Web Client ([i2b2 Web Client](#))

LISIBLE PAR UN HUMAIN : OUI

- Les données dans i2b2 sont présentées sous format tabulaire, le nom des tables et des variables est compréhensible et renseigne directement sur leur contenu.

6. Glossaire des acronymes



1. **AP-HP** : Assistance Publique – Hôpitaux de Paris
2. **API REST** : Application Programming Interface REpresentational State Transfert
3. **ARCH-OMOP** : Accessible Research Commons for Health – Observational Medical Outcomes Partnership
4. **BRIDG** : Biomedical Research Integrated Domain Group
5. **CA-125** : Cancer Antigen 125
6. **CARE** : Cohort Analysis & Refinement Expeditor
7. **CBC** : Complete Blood Count (Numération Sanguine Complète)
8. **CCTS** : Center for Clinical and Translational Science
9. **CDISC** : Clinical Data Interchange Standards Consortium
10. **CDM** : Common Data Model
11. **CDMH** : Common Data Model Harmonization
12. **CIM** : Classification Internationale des Maladies (ICD : International Classification of Diseases)
13. **CRC** : Client Research Chart
14. **CSV** : Comma-Separated Values
15. **DBA** : DataBase Administrator
16. **DDL** : Data Definition Language
17. **EAV** : Entity-Attribute-Value
18. **EHR** : Electronic Health Record
19. **EMR** : Electronic Medical Records
20. **ETL** : Extract Transform Load
21. **FHIR** : Fast Health Interoperability Resources
22. **HIPAA** : Health Insurance Portability and Accountability Act
23. **HL7** : Health Level 7
24. **HCPCS** : Healthcare Common Procedure Coding System
25. **HTML** : HyperText Markup Language
26. **HTTPD** : HyperText Transfer Protocol Daemon
27. **GIF** : Graphics interchange Format
28. **I2B2** : Informatics for Integrating Biology & the Bedside
29. **ID** : IDentifiant
30. **IDR** : Integrated Data Repository
31. **IIS** : Internet Information Services
32. **INCa** : Institut National du Cancer
33. **Inserm** : Institut national de la santé et de la recherche médicale
34. **JPG / JPEG** : Joint Photographic Experts Group
35. **LOINC** : Logical Observation Identifiers Names and Codes
36. **MS SQL Server** : MicroSoft Structured Query Language Server
37. **NCBC** : National Center for Biomedical Computing
38. **NDC** : National Drug Code
39. **NIH** : National Institutes of Health
40. **OAuth** : Open Authorization
41. **OMOP-CDM** : Observational Medical Outcomes Partnership – Common Data Model
42. **PCORnet CDM** : Patient-Centered Outcomes Research Network Common Data Model
43. **PHP** : Hypertext Preprocessor
44. **PNG** : Portable Network Graphics
45. **SGBD** : Système de Gestion de Base de Données
46. **SHRINE** : Shared Health Research Informatics Network
47. **SMART** : Substitutable Medical Applications and Reusable Technologies
48. **SNOMED CT** : Systemized NOMenclature of MEDicine Clinical Terms
49. **SQL** : Structured Query Language
50. **WISE** : Workplace Items Sharing Enhancement
51. **XLS** : eXcel Spreadsheets
52. **XML** : Extensible Markup Language

Annexe n°1 : Analyse quantitative de la littérature sur i2b2 et extrait de la liste des publications

Sur PubMed ([PubMed](https://pubmed.ncbi.nlm.nih.gov/)), on réalise une requête générale sur FHIR. La recherche du mot-clé « i2b2 » renvoie 439 résultats. Voici ci-dessous un extrait de la liste des publications obtenues en résultat :

1: Bucalo M, Gabetta M, Chiudinelli L, Larizza C, Bellasi A, Zambelli A, Barbarini N. i2b2 to Optimize Patients Enrollment. *Stud Health Technol Inform.* 2021 May 27;281:506-507. doi: 10.3233/SHTI210217. PMID: 34042623.

2: Waghlikar KB, Mendis M, Dessai P, Sanz J, Law S, Gilson M, Sanders S, Vangala M, Bell DS, Murphy SN. Automating Installation of the Integrating Biology and the Bedside (i2b2) Platform. *Biomed Inform Insights.* 2018 Jun 4;10:1178222618777749. doi: 10.1177/1178222618777749. PMID: 29887730; PMCID:PMC5989048.

3: Pedrera-Jimenez M, Garcia-Barrio N, Hernandez-Ibarburu G, Baselga B, Blanco A, Calvo-Boyero F, Gutierrez-Sacristan A, Quiros V, Cruz-Bermudez JL, Bernal JL, Meloni L, Perez-Rey D, Palchuk M, Kohane I, Serrano P. Building an i2b2-Based Population Repository for COVID-19 Research. *Stud Health Technol Inform.* 2022 May 25;294:287-291. doi: 10.3233/SHTI220460. PMID: 35612078.

4: Paris N, Mendis M, Daniel C, Murphy S, Tannier X, Zweigenbaum P. i2b2 implemented over SMART-on-FHIR. *AMIA Jt Summits Transl Sci Proc.* 2018 May 18;2017:369-378. PMID: 29888095; PMCID: PMC5961782.

5: Waghlikar KB, Ainsworth L, Zelle D, Chaney K, Mendis M, Klann J, Blood AJ, Miller A, Chulyadyo R, Oates M, Gordon WJ, Aronson SJ, Scirica BM, Murphy SN. I2b2-etl: Python application for importing electronic health data into the informatics for integrating biology and the bedside platform. *Bioinformatics.* 2022 Oct 14;38(20):4833-4836. doi: 10.1093/bioinformatics/btac595. PMID:36053173; PMCID: PMC9563689.

6: Waghlikar KB, Mandel JC, Klann JG, Wattanasin N, Mendis M, Chute CG, Mandl KD, Murphy SN. SMART-on-FHIR implemented over i2b2. *J Am Med Inform Assoc.* 2017 Mar 1;24(2):398-402. doi: 10.1093/jamia/ocw079. PMID:27274012; PMCID: PMC5391721.

7: Klann JG, Abend A, Raghavan VA, Mandl KD, Murphy SN. Data interchange using i2b2. *J Am Med Inform Assoc.* 2016 Sep;23(5):909-15. doi: 10.1093/jamia/ocv188. Epub 2016 Feb 5. PMID: 26911824; PMCID: PMC4997035.

8: Sun W, Rumshisky A, Uzuner O. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. *J Am Med Inform Assoc.* 2013 Sep-Oct;20(5):806-13. doi: 10.1136/amiajnl-2013-001628. Epub 2013 Apr 5. PMID: 23564629; PMCID: PMC3756273.

9: González L, Pérez-Rey D, Alonso E, Hernández G, Serrano P, Pedrera M, Gómez A, De Schepper K, Crepain T, Claerhout B. Building an I2B2-Based Population Repository for Clinical Research. *Stud Health Technol Inform.* 2020 Jun 16;270:78-82. doi: 10.3233/SHTI200126. PMID: 32570350.

10: Stöhr MR, Majeed RW, Günther A. Metadata Import from RDF to i2b2. *Stud Health Technol Inform.* 2018;253:40-44. PMID: 30147037.