

OHDSI内では、実名での活動になります。
Zoom参加時も「名前は実氏名で」お願いします。



OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

オデッセイ
ジャパン

OHDSI Japan evening conference #31

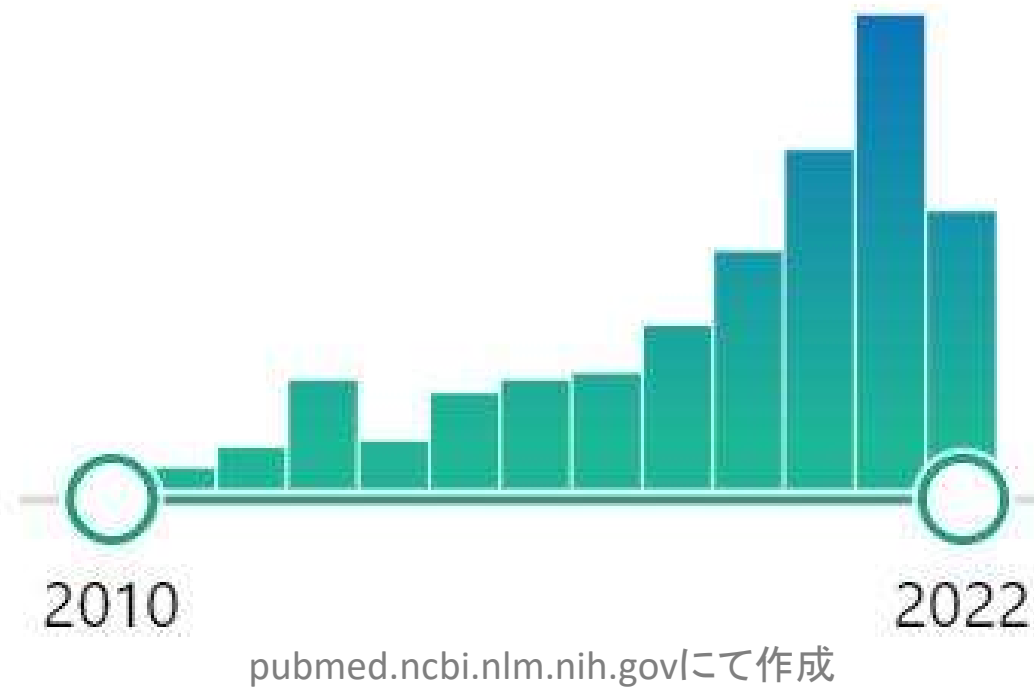
イブニング カンファレンス(第31回)

2022.6.28



OHDSI関連論文

Pubmedで“OHDSI or OMOP”を検索



全期間累計：5月262本→6月269本



細菌検査データをOMOP化しました

➤ [Stud Health Technol Inform.](#) 2022 May 25;294:292-296. doi: 10.3233/SHTI220461.

Usability of OMOP Common Data Model for Detailed Lab Microbiology Results

Emmanuelle Theron¹, Jean-François Gorse¹, Xavier Gansel¹

Affiliations — collapse

Affiliation

¹ bioMérieux, Data Science, Grenoble, France.

PMID: 35612079 DOI: [10.3233/SHTI220461](#)

Abstract

Anti-microbial resistance surveillance systems in Europe are limited by the inability to link laboratory data and patient data. The OMOP Common Data Model (OMOP CDM) is an option to store and use patient data in an international context supporting observational research. Detailed medical microbiology laboratory data are usually not stored in OMOP CDM.

We propose here a solution to deal with the inherent complexity of microbiology data and store those in the OMOP CDM v5.4. We demonstrate the feasibility of our approach by capturing data from a microbiology in vitro diagnostic middleware, modeling in OMOP CDM 5.4 and querying for visualization.

概要

ヨーロッパにおける抗微生物薬耐性サーベイランスシステムは、検査室データと患者データをリンクさせることができないため、制限されています。OMOP CDMは、観察研究をサポートする国際的な文脈で患者データを保存し使用するためのオプションである。医療微生物学の詳細な検査データは通常OMOP CDMに保存されていない。

我々は、微生物検査データ固有の複雑性に対処し、それらを OMOP CDM v5.4 に格納するソリューションを提案する。我々は、微生物学体外診断用ミドルウェアからデータを取得し、OMOP CDM 5.4でモデリングし、可視化のためのクエリを実行することにより、我々のアプローチの実現可能性を実証する。

⇒全欧の薬剤耐性サーベイランス ネットワークの構築を見据える



オランダNICEでは、
FAIRなデータベースを実現するのに
OMOP CDMを使った。

› Stud Health Technol Inform. 2022 May 25;294:367-371. doi: 10.3233/SHTI220476.

FAIRifying a Quality Registry Using OMOP CDM: Challenges and Solutions

Daniel Puttmann^{1 2}, Nicolette De Keizer^{1 2}, Ronald Cornet², Eric Van Der Zwan^{1 2},
Ferishta Bakhshi-Raiez^{1 2}

Affiliations — collapse

Affiliations

- ¹ The National Intensive Care Evaluation (NICE) registry, The Netherlands.
- ² Department of Medical Informatics, Amsterdam Public Health Research Institute, Amsterdam UMC, University of Amsterdam, The Netherlands.

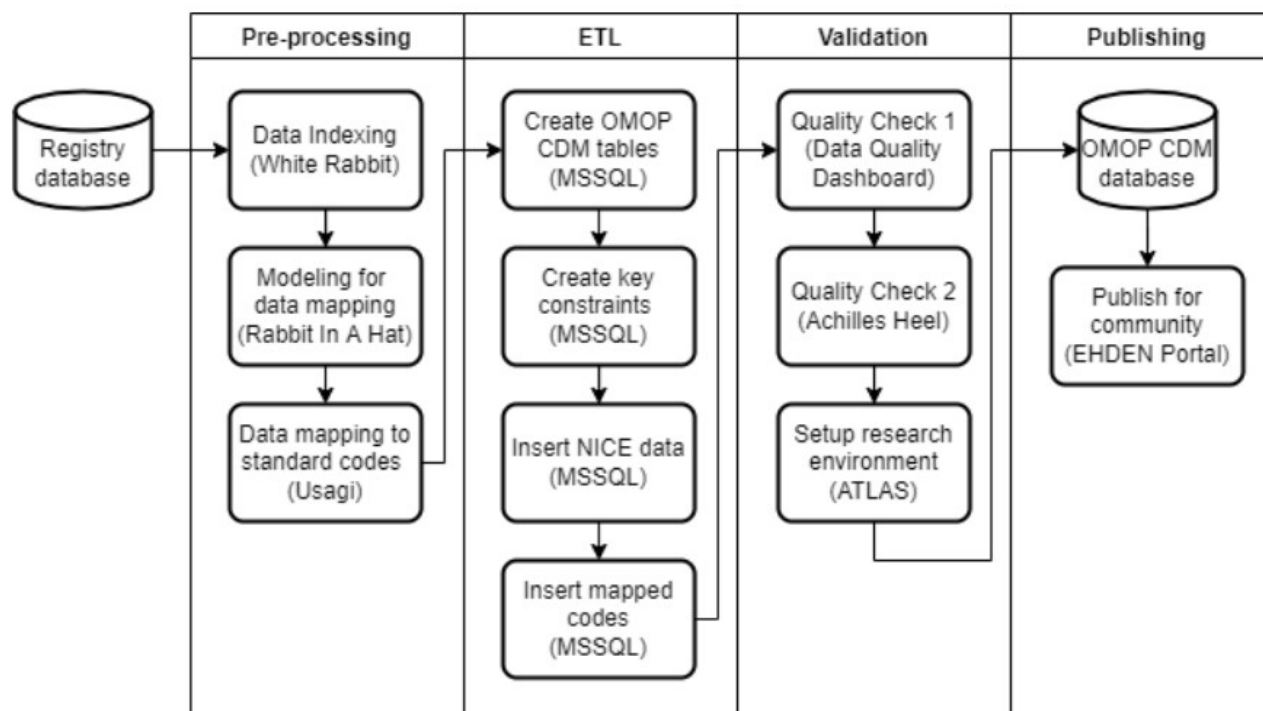
PMID: 35612098 DOI: [10.3233/SHTI220476](https://doi.org/10.3233/SHTI220476)

Abstract

The need for health data to be internationally Findable, Accessible, Interoperable and Reusable (FAIR) and thereby support integrative analysis with other datasets has become crystal clear in the ongoing pandemic. The Dutch National Intensive Care Evaluation (NICE) quality registry adopted the Observational Medical Outcomes Partnership Common Database Model (OMOP CDM) to achieve a FAIR database. In the process of adopting the OMOP CDM, many modeling, technical, and communication challenges needed to be solved. Through communication with the OMOP CDM implementation community, previously done research and trial-and-error we found solutions that we believe can help other healthcare institutions, especially ICU quality registries, FAIRify their databases.

要旨

健康データは国際的にFindable, Accessible, Interoperable and Reusable (FAIR) であるべきであり、それによって他のデータセットとの統合的解析をサポートする必要性が、進行中のパンデミックにおいて明らかにされている。オランダのNational Intensive Care Evaluation (NICE) quality registryは、FAIRデータベースを実現するためにOMOP CDMを採用した。OMOP CDMを採用する過程では、多くのモデリング、技術、コミュニケーションの課題を解決する必要があった。OMOP CDMの導入コミュニティとのコミュニケーション、先行研究、試行錯誤を経て、他の医療機関、特にICU品質レジストリのデータベースのFAIR化に役立つと思われる解決策を見出すことができた。





OMOP標準へのマッピングがない
ドイツOPSコードが2割もあるよ。

➤ [Stud Health Technol Inform. 2022 May 25;294:405-406. doi: 10.3233/SHTI220484.](#)

Evaluation and Challenges of Medical Procedure Data Harmonization to SNOMED-CT for Observational Research

Ines Reinecke ¹, Michael Kallfelz ², Martin Sedlmayr ¹, Joscha Siebel ¹, Franziska Bathelt ¹

Affiliations — collapse

Affiliations

- ¹ Carl Gustav Carus Faculty of Medicine, Center for Medical Informatics, Institute for Medical Informatics and Biometry, Technische Universität Dresden, Dresden, Germany.
- ² Odysseus Data Services GmbH, Berlin, Germany.

PMID: 35612106 DOI: [10.3233/SHTI220484](#)

Abstract

The relevance of health data research on real world data (RWD) is increasing. To prepare national RWD for international research, harmonization with standard terminologies is required. In this paper, we evaluate to what extent the German OPS vocabulary in OHDSI covers codes present in RWD and mappings to SNOMED-CT. The evaluation identified a mapping gap of 21.1% in the RWD set.

概要

実世界データ（RWD）を用いたヘルスデータ研究の関連性が高まっている。国際的な研究のために国内のRWDを準備するためには、標準的な用語集との調和が必要である。本論文では、OHDSIのドイツ語OPS語彙がRWDに存在するコードをどの程度カバーしているか、またSNOMED-CTとのマッピングを評価した。その結果、RWDのセットでは21.1%のマッピングギャップがあることが確認されました。

Operationen- und Prozedurenschlüssel
Version 2022

Übersicht über die Kapitel

Kapitel	Gliederung	Titel
1	1-10...1-99	DIAGNOSTISCHE MASSNAHMEN
3	3-03...3-99	BILDGEBENDE DIAGNOSTIK
5	5-01...5-99	OPERATIONEN
6	6-00...6-00	MEDIKAMENTE
8	8-01...8-99	NICHT OPERATIVE THERAPEUTISCHE MASSNAHMEN
9	9-26...9-99	ERGÄNZENDE MASSNAHMEN

Kapitel 1

DIAGNOSTISCHE MASSNAHMEN ([1-10...1-99](#))

Dieses Kapitel gliedert sich in folgende Gruppen:

1-10...1-10	Klinische Untersuchung
1-20...1-33	Untersuchung einzelner Körpersysteme
1-40...1-49	Biopsie ohne Inzision
1-50...1-58	Biopsie durch Inzision
1-61...1-69	Diagnostische Endoskopie
1-70...1-79	Funktionstests
1-84...1-85	Explorative diagnostische Maßnahmen
1-90...1-99	Andere diagnostische Maßnahmen

Kapitel 3

BILDGEBENDE DIAGNOSTIK ([3-03...3-99](#))

Dieses Kapitel gliedert sich in folgende Gruppen:

3-03...3-05	Ultraschalluntersuchungen
3-10...3-13	Projektionsradiographie
3-20...3-26	Computertomographie [CT]
3-30...3-31	Optische Verfahren
3-63...3-69	Röntgenverfahren

OPS: German procedure classificaton (Operationen- und Prozedurenschlüssel - OPS)

Official classification for the encoding of operations, procedures and general medical measures.

OHDSIボキャブラリで

すでにOMOP標準にマッピングされているOPSコードは実際に使われているOPSコードのうち80%程度だった。



> Stud Health Technol Inform. 2022 May 25;294:413-414. doi: 10.3233/SHTI220488.

Beyond the Brain: MIDS Extends BIDS to Multiple Modalities and Anatomical Regions

Jose Manuel Saborit-Torres ¹, Silvia Nadal-Almela ¹, Joaquim Angel Montell-Serrano ¹,
Elena Oliver-Garcia ¹, Hector Carceller ^{1 2 3}, Jon Ander Gómez-Ádrian ⁴,
Marisa Caparrós-Redondo ¹, Francisco García-García ^{1 5}, Julio Domenech-Fernández ⁶,
Maria De La Iglesia-Vayá ^{1 2 7}

Affiliations

- 1 Unidad Mixta de Imagen Biomédica FISABIO-CIPF. Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana - València, Spain.
- 2 CIBERSAM, ISC III. Av. Blasco Ibáñez 15, 46010 - València, Spain.
- 3 Institute for Biotechnology and Biomedicine (BIOTECMED), Universitat de Valencia, Burjassot, 46100, Spain.
- 4 Department of Computer Systems and Computation, Universitat Politècnica de València - València, Spain.
- 5 Bioinformatics & Biostatistics Unit, Príncipe Felipe Research Center - Valencia, Spain.
- 6 Hospital Arnau de Vilanova - Valencia, Spain.
- 7 General Directorate of Research, Innovation, Technology and Quality. Subdirectorate General of Information Systems for Health - Valencia, Spain.

Abstract

Brain Imaging Data Structure (BIDS) provides a valuable tool to organise brain imaging data into a clear and easy standard directory structure. Moreover, BIDS is widely supported by the scientific community and has been established as a powerful standard for medical imaging management. Nonetheless, the original BIDS is restricted to magnetic resonance imaging (MRI) of the brain, limiting its implantation to other techniques and anatomical regions. We developed Medical Imaging Data Structure (MIDS), conceived to extend BIDS methodology to other anatomical regions and multiple imaging systems in these areas. The MIDS standard was developed to store and manage medical images as an extension of BIDS. It allows the user to handily save studies of multiple anatomical regions and imaging techniques. Besides, MIDS improves the classification of multiple images within the structure, allowing the possibility to unify them in a single study to apply on them preprocessing or artificial intelligence algorithms. Finally, the results generated are saved in the derivatives folder.

概要

Brain Imaging Data Structure (BIDS) は、脳画像データを明確かつ容易な標準ディレクトリ構造に整理するための貴重なツールを提供します。さらに、BIDSは科学界から広く支持されており、医用画像管理の強力な標準として確立しています。それにもかかわらず、オリジナルのBIDSは脳の磁気共鳴画像（MRI）に限定されており、他の技術や解剖学的領域への移植が制限されています。そこで、BIDSの手法を他の解剖学的領域や複数の画像システムに拡張するために、MIDS（Medical Imaging Data Structure）を開発しました。MIDSは、BIDSの拡張機能として、医用画像を保存・管理するために開発された規格です。これにより、ユーザーは複数の解剖学的領域や画像技術に関する研究を簡単に保存することができます。さらに、MIDSは構造内の複数の画像の分類を改善し、それらを1つの研究に統合して前処理や人工知能のアルゴリズムを適用する可能性を可能にします。最後に、生成された結果は、派生フォルダに保存されます。

- ・BIDSを拡張してMIDSという規格を作った。
- ・さらにOMOP CDMと互換性を持たせるよう拡張を続けている。



RWDに基づく臨床ガイドラインの改良にむけて

➤ [Stud Health Technol Inform.](#) 2022 May 25;294:480-484. doi: 10.3233/SHTI220505.

Towards the Improvement of Clinical Guidelines Based on Real World Data

[Elisa Henke](#)¹, [Ines Reinecke](#)¹, [Michele Zoch](#)¹, [Martin Sedlmayr](#)¹, [Franziska Bathelt](#)¹

Affiliations — collapse

Affiliation

¹ Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine,
Technische Universität Dresden, Dresden, Germany.

PMID: 35612126 DOI: [10.3233/SHTI220505](#)

Abstract

Computerized clinical guidelines (CCG) are effective instruments for standardizing, monitoring and optimizing medical treatment processes. Nevertheless, due to barriers in flexibility, transferability and acceptance, the widespread use of CCG in clinical practice is not yet common. To overcome those issues, we present a concept on how to use real world data to evaluate CCG and to recommend improvements. As a first result, we defined an algorithm to extract treatment processes based on the standardized Observational Medical Outcomes Partnership (OMOP) Common Data Model as well as their visualization using the graphical modeling language Business Process Model and Notation (BPMN).

概要

コンピュータ化された臨床ガイドライン（CCG）は、医療プロセスの標準化、モニタリング、最適化のための有効な手段である。しかしながら、柔軟性、移植性、受容性などの障壁により、臨床現場でのCCGの普及はまだ一般的でない。これらの問題を克服するために、我々は、実世界のデータを用いてCCGを評価し、改善を推奨する方法についての概念を提示する。第一の成果として、標準化されたOMO CDMに基づく治療プロセスの抽出と、グラフィカルモデリング言語BPMN（Business Process Model and Notation）によるその可視化アルゴリズムを定義した。

- ・CCGの妥当性を検証し改良するために、OMOP CDMのデータを使うという発想。
- ・OMOP CDMから治療プロセスを抽出する。
- ・まだ5例でしか評価していないプロトタイプ。



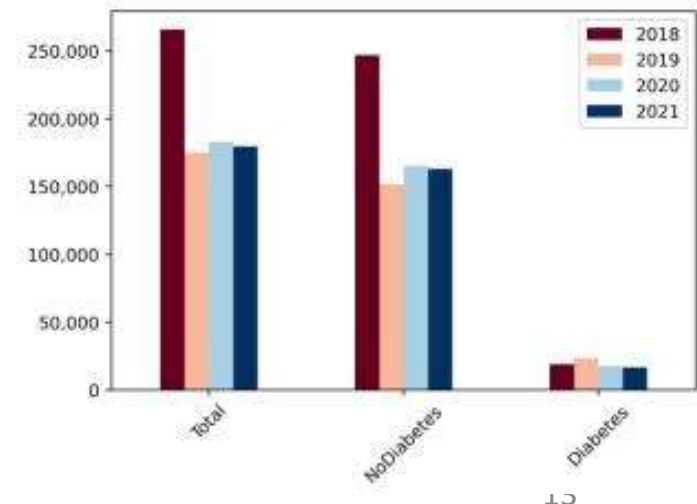
> [Nutrients](#). 2022 May 11;14(10):2016. doi: 10.3390/nu14102016.

Opportunities of Digital Infrastructures for Disease Management—Exemplified on COVID-19-Related Change in Diagnosis Counts for Diabetes-Related Eye Diseases

Franziska Bathelt ¹, Ines Reinecke ¹, Yuan Peng ¹, Elisa Henke ¹, Jens Weidner ¹, Martin Bartos ², Robert Gött ³, Dagmar Waltemath ³, Katrin Engelmann ⁴, Peter Eh Schwarz ⁵, Martin Sedlmayr ¹

Affiliations

- ¹ Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, 01307 Dresden, Germany.
- ² Department of Computer Science, Klinikum Chemnitz gGmbH, Flemmingstr. 2, 09116 Chemnitz, Germany.
- ³ Core Unit Datenintegrationszentrum, Universitätsmedizin Greifswald, Walther-Rathenau-Str. 48, 17475 Greifswald, Germany.
- ⁴ Department of Ophthalmology, Klinikum Chemnitz gGmbH, Flemmingstr. 2, 09116 Chemnitz, Germany.
- ⁵ Department of Medicine, University of Dresden, Carl Gustav Carus, 01307 Dresden, Germany.



Abstract

Background: Retrospective research on real-world data provides the ability to gain evidence on specific topics especially when running across different sites in research networks. Those research networks have become increasingly relevant in recent years; not least due to the special situation caused by the COVID-19 pandemic. An important requirement for those networks is the data harmonization by ensuring the semantic interoperability. **Aims:** In this paper we demonstrate (1) how to facilitate digital infrastructures to run a retrospective study in a research network spread across university and non-university hospital sites; and (2) to answer a medical question on COVID-19 related change in diagnostic counts for diabetes-related eye diseases. **Materials and methods:** The study is retrospective and non-interventional and runs on medical case data documented in routine care at the participating sites. The technical infrastructure consists of the OMOP CDM and other OHDSI tools that is provided in a transferable format. An ETL process to transfer and harmonize the data to the OMOP CDM has been utilized. Cohort definitions for each year in observation have been created centrally and applied locally against medical case data of all participating sites and analyzed with descriptive statistics. **Results:** The analyses showed an expectable drop of the total number of diagnoses and the diagnoses for diabetes in general; whereas the number of diagnoses for diabetes-related eye diseases surprisingly decreased stronger compared to non-eye diseases. Differences in relative changes of diagnoses counts between sites show an urgent need to process multi-centric studies rather than single-site studies to reduce bias in the data. **Conclusions:** This study has demonstrated the ability to utilize an existing portable and standardized infrastructure and ETL process from a university hospital setting and transfer it to non-university sites. From a medical perspective further activity is needed to evaluate data quality of the utilized real-world data documented in routine care and to investigate its eligibility of this data for research.

背景: 実データを用いたレトロスペクティブな研究は、特に研究ネットワークにおいて異なる施設間で行われる場合に、特定のトピックに関するエビデンスを得る能力を提供するものである。近年、COVID-19の大流行により、研究ネットワークの重要性が増している。これらのネットワークの重要な要件は、セマンティックな相互運用性を確保することによって、データの調和を図ることである。本論文では、(1)大学病院と大学以外の病院にまたがる研究ネットワークにおいて、デジタルインフラを利用して後ろ向き研究を行う方法、および(2)COVID-19に関連した糖尿病関連眼疾患の診断回数の変化に関する医学的疑問への回答を示す。**材料と方法:** 本研究は、参加施設の日常診療で記録された医療症例データをもとに、レトロスペクティブかつ非介入的に実施されるものである。技術的基盤は、OMOP CDM と他の OHDSI ツールから構成され、転送可能な形式で提供される。OMOP CDM にデータを転送し、整合させるための ETL プロセスが利用されている。観察年ごとのコホート定義は中央で作成され、全参加施設の医療ケースデータに対してローカルに適用され、記述統計で分析されている。**結果:** 糖尿病に関連する眼科疾患の診断数は、眼科以外の疾患と比較して、驚くほど減少していた。診断数の相対的变化における施設間の差は、データの偏りを減らすために、単一施設での調査よりもむしろ多施設での調査を処理することが急務であることを示している。**結論:** 本研究は、大学病院の環境から、既存のポータブルで標準化されたインフラと ETL プロセスを利用し、大学以外の施設に転送する能力を実証したものである。医学的な観点からは、ルーチンケアで記録された利用された実世界のデータの品質を評価し、このデータの研究への適格性を調査するために、さらなる活動が必要である。



血栓性血小板減少症の背景発症率
今度は拡大して国際17DSで⇒バラツキがやばかった

Observational Study > Drug Saf. 2022 Jun;45(6):685-698. doi: 10.1007/s40264-022-01187-y.
Epub 2022 Jun 2.

Phenotype Algorithms for the Identification and Characterization of Vaccine-Induced Thrombotic Thrombocytopenia in Real World Data: A Multinational Network Cohort Study

Azza Shoaibi ^{1 2}, Gowtham A Rao ^{3 4}, Erica A Voss ^{3 4}, Anna Ostropelets ^{4 5},
Miguel Angel Mayer ⁶, Juan Manuel Ramírez-Anguita ⁶, Filip Maljković ⁷, Biljana Carević ⁸,
Scott Horban ⁹, Daniel R Morales ⁹, Talita Duarte-Salles ¹⁰, Clement Fraboulet ¹¹,
Tanguy Le Carrou ¹², Spiros Denaxas ¹³, Vaclav Papez ¹³, Luis H John ¹⁴, Peter R Rijnbeek ¹⁴,
Evan Minty ¹⁵, Thamir M Alshammari ^{4 16}, Rupa Makadia ^{3 4}, Clair Blacketer ^{3 4},
Frank DeFalco ^{3 4}, Anthony G Sena ^{3 4}, Marc A Suchard ^{4 17}, Daniel Prieto-Alhambra ¹⁸,
Patrick B Ryan ^{3 4}

Abstract

Introduction: **Vaccine-induced thrombotic thrombocytopenia** (VITT) has been identified as a rare but serious adverse event associated with coronavirus disease 2019 (COVID-19) vaccines.

Objectives: In this study, we explored the pre-pandemic co-occurrence of thrombosis with thrombocytopenia (TWT) using 17 observational health data sources across the world. We applied multiple TWT definitions, estimated the background rate of TWT, characterized TWT patients, and explored the makeup of thrombosis types among TWT patients.

Methods: We conducted an international network retrospective cohort study using electronic health records and insurance claims data, estimating background rates of TWT amongst persons observed from 2017 to 2019. Following the principles of existing VITT clinical definitions, TWT was defined as patients with a diagnosis of embolic or thrombotic arterial or venous events and a diagnosis or measurement of thrombocytopenia within 7 days. Six TWT phenotypes were considered, which varied in the approach taken in defining thrombosis and thrombocytopenia in real world data.

Results: Overall TWT incidence rates ranged from 1.62 to 150.65 per 100,000 person-years. Substantial heterogeneity exists across data sources and by age, sex, and alternative TWT phenotypes. TWT patients were likely to be men of older age with various comorbidities. Among the thrombosis types, arterial thrombotic events were the most common.

Conclusion: Our findings suggest that identifying VITT in observational data presents a substantial challenge, as implementing VITT case definitions based on the co-occurrence of TWT results in large and heterogeneous incidence rate and in a cohort of patients with baseline characteristics that are inconsistent with the VITT cases reported to date.

はじめに: ワクチンによる血栓性血小板減少症（VITT）は、COVID-19ワクチンに関連したまれではあるが重篤な有害事象として確認されている。

目的: 本研究では、世界中の17の観察型健康データソースを用いて、パンデミック前の血小板減少症を伴う血栓症（TWT）の共発生を調査しました。複数のTWTの定義を適用し、TWTのバックグラウンド率を推定し、TWT患者の特徴を示し、TWT患者における血栓症のタイプの構成を調査した。

方法: 電子カルテと保険請求データを用いた国際ネットワークレトロスペクティブコホート研究を実施し、2017年から2019年に観察された人の中のTWTのバックグラウンド率を推定した。既存のVITT臨床定義の原則に従い、TWTは塞栓性または血栓性の動脈または静脈イベントの診断を受け、7日以内に血小板減少の診断または測定を受けた患者と定義した。6つのTWT Phenotyping法が検討され、それらは実際のデータにおける血栓症および血小板減少症の定義においてとられたアプローチに差異があった。

結果: 全体のTWT発生率は、10万人年当たり1.62から150.65の範囲であった。データソースや年齢、性別、TWTの表現型によって大きな異質性が存在する。TWT患者は、様々な併存疾患を持つ高齢の男性である可能性が高い。血栓症のタイプでは、動脈血栓塞栓症が最も一般的であった。

結論: TWT の併発に基づく VITT 症例定義を実施すると、発症率が大きく不均一になり、これまで報告された VITT 症例と一致しないベースライン特性を持つ患者コホートになるため、観察データから VITT を特定することはかなりの難題であることが示唆された。



> [Stud Health Technol Inform.](#) 2022 Jun 6;290:56-60. doi: 10.3233/SHTI220031.

A Multi-Omics Common Data Model for Primary Immunodeficiencies

Mélanie Buy¹, William Digan^{1 2}, Xiaoyi Chen², Julien Husson¹, Mickael Ménager³,
Frédéric Rieux-Laucat⁴, Nicolas Garcelon¹, ATRACtion Members

Affiliations — collapse

Affiliations

- 1 Université de Paris, Imagine Institute, Data Science Platform, INSERM UMR 1163, F-75015, Paris, France.
- 2 Centre de Recherche des Cordeliers, INSERM, Université de Paris, Sorbonne Université, F-75006, Paris, France.
- 3 Université de Paris, Imagine Institute, Laboratory of Inflammatory Responses and Transcriptomic Networks in Diseases, Atip-Avenir Team, INSERM UMR 1163, F-75015 Paris, France.
- 4 Université de Paris, Imagine Institute, Laboratory of Immunogenetics of Pediatric Autoimmune Diseases, INSERM UMR U1163, F-75015 Paris, France.

Abstract

Primary Immunodeficiencies (PIDs) are associated with more than 400 rare monogenic diseases affecting various biological functions (e.g., development, regulation of the immune response) with a heterogeneous clinical expression (from no symptom to severe manifestations). To better understand PIDs, the ATRACTion project aims to perform a multi-omics analysis of PIDs cases versus a control group patients, including single-cell transcriptomics, epigenetics, proteomics, metabolomics, metagenomics and lipidomics. In this study, our goal is to develop a common data model integrating clinical and omics data, which can be used to obtain standardized information necessary for characterization of PIDs patients and for further systematic analysis. For that purpose, we extend the OMOP Common Data Model (CDM) and propose a multi-omics ATRACTion OMOP-CDM to integrate multi-omics data. This model, available for the community, is customizable for other types of rare diseases (<https://framagit.org/imagine-plateforme-bdd/pub-rhu4-atraction>).

概要

原発性免疫不全症（PIDs）は、様々な生物学的機能（例えば、発生、免疫反応の制御）に影響を及ぼす400以上の希少な単発性疾患であり、臨床症状は不均一（無症状から重症まで）である。ATRACtionプロジェクトでは、PIDsの理解を深めるために、PIDs症例と対照群患者のシングルセル・トランスクリプトミクス、エピジェネティクス、プロテオミクス、メタボロミクス、メタゲノミクス、リピドミクスなどのマルチオミクス解析を実施することを目的としている。

本研究では、臨床データとオミックスデータを統合した共通のデータモデルを開発し、PIDs患者の特徴づけとさらなる系統的な解析に必要な標準的な情報を得ることを目的としています。そのために、OMOP CDM を拡張し、マルチオミックスデータを統合するためのマルチオミックスATRACtion OMOP-CDMを提案する。

このモデルはコミュニティ向けに公開されており、他のタイプの希少疾患にもカスタマイズ可能である。
(<https://framagit.org/imagine-plateforme-bdd/pub-rhu4-atraction>)。



> Stud Health Technol Inform. 2022 Jun 6;290:76-80. doi: 10.3233/SHTI220035.

Translating the Observational Medical Outcomes Partnership – Common Data Model (OMOP-CDM) Electronic Health Records to an OWL Ontology

Lamy Jean-Baptiste ¹, Abdelmalek Mouazer ¹, Karima Sedki ¹, Rosy Tsopra ^{2 3 4}

Affiliations – collapse

Affiliations

- ¹ Université Sorbonne Paris Nord, LIMICS, Sorbonne Université, INSERM, UMR 1142, F-93000, Bobigny, France.
- ² INSERM, Université de Paris, Sorbonne Université, Centre de Recherche des Cordeliers, Information Sciences to support Personalized Medicine, F-75006 Paris, France.
- ³ Department of Medical Informatics, Hôpital Européen Georges-Pompidou, AP-HP, Paris, France.
- ⁴ INRIA Paris, 75012 Paris, France.

PMID: 35672974 DOI: 10.3233/SHTI220035

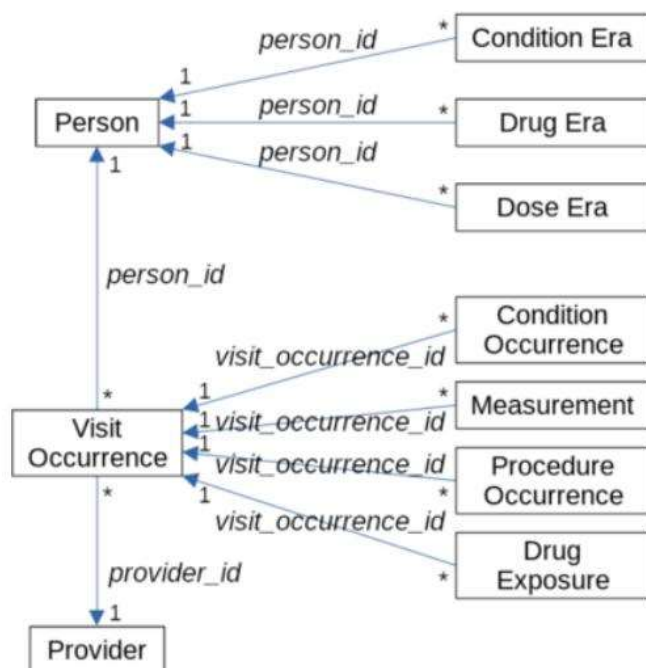
Abstract

The heterogeneity of electronic health records model is a major problem: it is necessary to gather data from various models for clinical research, but also for clinical decision support. The Observational Medical Outcomes Partnership - Common Data Model (OMOP-CDM) has emerged as a standard model for structuring health records populated from various other sources. This model is proposed as a relational database schema. However, in the field of decision support, formal ontologies are commonly used. In this paper, we propose a translation of OMOP-CDM into an ontology, and we explore the utility of the semantic web for structuring EHR in a clinical decision support perspective, and the use of the SPARQL language for querying health records. The resulting ontology is available online.

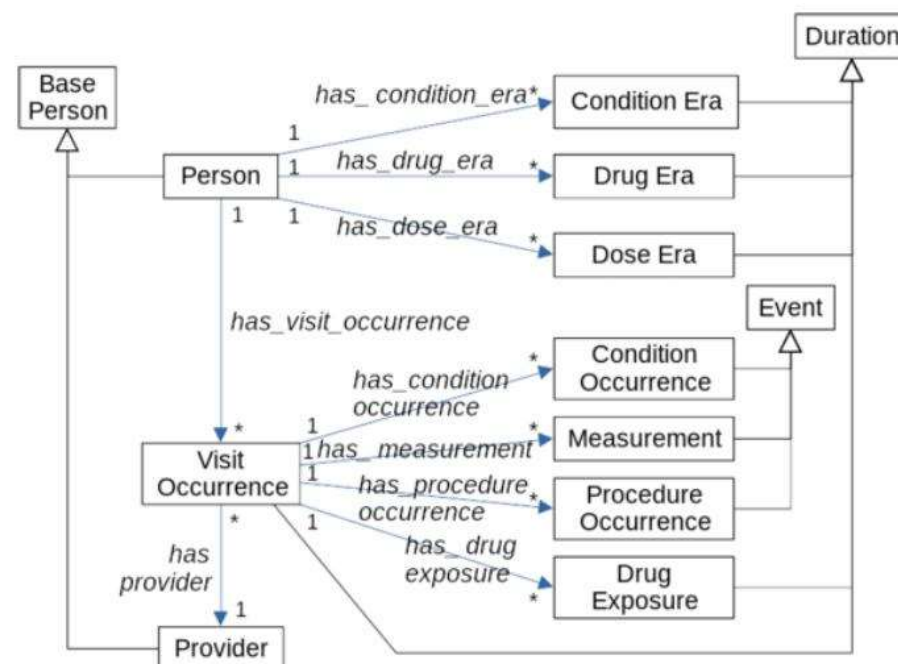
概要

電子カルテのモデルの異質性は大きな問題である。臨床研究のために様々なモデルからデータを収集する必要があるが、臨床的な意思決定支援のためにも必要である。様々なソースから収集された健康記録を構造化するための標準モデルとして、OMOP-CDM が登場した。このモデルは、リレーショナルデータベースのスキーマとして提案されている。しかし、意思決定支援の分野では、形式的なオントロジーを用いるのが一般的である。本論文では、OMOP-CDMのオントロジーへの翻訳を提案し、臨床的意思決定支援の観点からEHRを構造化するためのセマンティックウェブの有用性と、健康記録の問い合わせのためのSPARQL言語の利用を検討する。出来上がったオントロジーはオンラインで公開されている。

- ・OMOP 1000患者が、約1000万のRDF triplesになった(SynPUF)。
- ・検索速度(SQL vs SPARQL)はほぼ同じだった。



OMOP CDM



オントロジー

- ☐ 1 [Transforming Thyroid Cancer Diagnosis and Staging Information from Unstructured Reports to the Observational Medical Outcome Partnership Common Data Model.](#)
Cite Yoo S, Yoon E, Boo D, Kim B, Kim S, Paeng JC, Yoo IR, Choi IY, Kim K, Ryoo HG, Lee SJ, Song E, Joo YH, Kim J, Lee HY.
Share Appl Clin Inform. 2022 May;13(3):521-531. doi: 10.1055/s-0042-1748144. Epub 2022 Jun 15.
PMID: 35705182 [Free PMC article.](#)
To the best of our knowledge, there have been no studies on **OMOP** CDM transformation and natural language processing (NLP) for thyroid cancer to date. ...METHODS: Thyroid cancer diagnosis and stage-related modifiers were extracted with rule-based NLP from 63,795 thyroid can ...
- ☐ 2 [Large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus \(LEGEND-T2DM\): a protocol for a series of multinational, real-world comparative cardiovascular effectiveness and safety studies.](#)
Cite Khera R, Schuemie MJ, Lu Y, Ostropolets A, Chen R, Hripcsak G, Ryan PB, Krumholz HM, Suchard MA.
Share BMJ Open. 2022 Jun 9;12(6):e057977. doi: 10.1136/bmjopen-2021-057977.
PMID: 35680274 [Free PMC article.](#)
LEGEND-T2DM will leverage the Observational Health Data Sciences and Informatics (**OHDSI**) community that provides access to a global network of administrative claims and electronic health record data sources, representing 190 million patients in the USA and about 50 million ...
- ☐ 3 [A Natural Language Processing Tool Offering Data Extraction for COVID-19 Related Information \(DECOVRI\).](#)
Cite Heider PM, Pipaliya RM, Meystre SM.
Share Stud Health Technol Inform. 2022 Jun 6;290:1062-1063. doi: 10.3233/SHTI220268.
PMID: 35673206
This NLP application called DECOVRI (Data Extraction for COVID-19 Related Information) will be released as a free and open source tool to convert unstructured notes into structured data within an **OMOP** CDM-based ecosystem. The DECOVRI prototype is being continuously improve ...



最近のGlobal OHDSI話題から

■ 2022年のOHDSI Symposium

※今年からリアル開催に戻った

- 6月 24-26日 European Symposium
- 10月14-16日 Global Symposium
 - Oct. 14 – Main Conference
 - Oct. 15 – Tutorial
 - Oct. 15-16 – Workgroup Activities
- APAC Symposiumは11月12-13日に台北で開催予定
リアル+ Webの予定



最近のGlobal OHDSI話題から

6月のCommunity Call

●APAC Community Call テーマ

June 2 Mini Tutorial#2

Incidence and prevalence: Concepts and implementation in OHDSI

June 16 Chapter Mid-year Updates

●Global Community Call テーマ

May 31 Around The Workgroups

June 7 Welcome to OHDSI

June 14 OHDSI Publications in 2022

June 21 10-Minute Tutorials : PheValuator



最近のGlobal OHDSI話題から

重要なアナウンス

**SNOMED CTが、OHDSIでは地域を問わず
個別ライセンス不要になりました！**

条件

- OMOP CDMを含むOHDSIツール(ATLAS等)を利用するとき。
- SNOMED CTを外に取り出して使ってはならない。

「OHDSIツール」の範囲：

GitHubでOHDSIとして登録されていれば良さそう。



疑似データほしい

<https://synthea.mitre.org/>

Synthea

- 疑似患者データを生成してくれるソフト(java)
- 人数・性別・年齢範囲などを指定できる

SyntheticMass

- マサチューセッツ州を想定した疑似患者（生成データ）
- 100万人分。 1000人(1K)のサンプルもある。
- データ形式はHL7 FHIR, C-CDA, CSV。
- FHIR APIでのオンラインアクセスもできる。

その他の疑似データもDLできる

- COVID-19患者データ(1万人、10万人)
- 小児疑似データ(6357人)
- 乳がん疑似データ
- カナダを想定した疑似データ



Syntheaソフトウェア

■synthea-with-dependencies.jarをダウンロード

■実行

```
java -jar synthea-with-dependencies.jar [-h]
                                     [-s seed]
                                     [-r referenceDate as YYYYMMDD]
                                     [-cs clinician seed]
                                     [-p populationSize]
                                     [-g gender]
                                     [-a minAge-maxAge]
                                     [-c localConfigFilePath]
                                     [-d localModulesDirPath]
                                     [state [city]]
```

これだけ



ETL-Synthea

<https://github.com/OHDSI/ETL-Synthea>

■ SyntheaデータをOMOPに変換

Utility to Load Synthea CSV data to OMOP CDM

Currently supports CDM v5.3 and v5.4

Follow the steps on the [synthea wiki](#) to run the program and generate the files. This builder works off of the csv files, not the fhir files. To do this the `exporter.csv.export` option in the `./src/main/resources/synthea.properties` file needs to be set to TRUE.



どんなデータなのか

SyntheticMass 1K サンプルデータ

含まれるファイル

72 allergies.csv
660 careplans.csv
1040 conditions.csv
20 devices.csv
16092 encounters.csv
204 imaging_studies.csv
2176 immunizations.csv
10632 medications.csv
42088 observations.csv
168 organizations.csv
328 patients.csv
4 payers.csv
336 payer_transitions.csv
5564 procedures.csv
1024 providers.csv
4 supplies.csv
(KB)

中を見るとID類にはUUIDが使われていて、とっつきにくいいため、UUIDをわかりやすいIDに置き換えた。

例: observations.csv最初の2行

```
DATE,PATIENT,ENCOUNTER,CODE,DESCRIPTION,VALUE,UNITS,TYPE  
2012-01-23T17:45:28Z,034e9e3b-2def-4559-bb2a-7850888ae060,e88bc3a9-  
007c-405e-aabc-792a38f4aa2b,8302-2,Body Height,193.3,cm,numeric
```



```
DATE,PATIENT,ENCOUNTER,CODE,DESCRIPTION,VALUE,UNITS,TYPE  
2012-01-23T17:45:28Z,pt0014,en48440,8302-2,Body Height,  
193.3,cm,numeric
```

Jason Walonoski, Mark Kramer, Joseph Nichols, Andre Quina, Chris Moesel, Dylan Hall, Carlton Duffett, Kudakwashe Dube, Thomas Gallagher, Scott McLachlan, Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record, *Journal of the American Medical Informatics Association*, Volume 25, Issue 3, March 2018, Pages 230–238, <https://doi.org/10.1093/jamia/ocx079>

ちなみに
UUIDの数は

Careplan = 3483
ENCOUNTER = 53346
Imaging = 855
ORGANIZATION = 1119
PATIENT = 1171
PAYER = 10
PROVIDER = 5855



SyntheticMass 1K サンプルデータ

中はこんなデータ#1/4

allergies.csv↓								
START	STOP	PATIENT	ENCOUNTER	CODE	DESCRIPTION↓			
1982-10-25		pt0499	en38608	300916003	Latex allergy↓			
1982-10-25		pt0499	en38608	300913006	Shellfish allergy↓			
↓								
careplans.csv↓								
Id	START	STOP	PATIENT	ENCOUNTER	CODE	DESCRIPTION	REASONCODE	REASONDESCRIPTION
cp2862	2010-01-23	2012-01-23	pt0014	en43503	53950000	Respiratory the	10509002	Acute bronchitis (
cp0121	2011-05-13	2011-08-02	pt0074	en47065	53950000	Respiratory the	10509002	Acute bronchitis (
↓								
conditions.csv↓								
START	STOP	PATIENT	ENCOUNTER	CODE	DESCRIPTION↓			
2001-05-01		pt0141	en29815	40055000	Chronic sinusitis (disord			
2011-08-09	2011-08-16	pt0633	en32706	444814009	Viral sinusitis (disorder)↓			
↓								
devices.csv↓								
START	STOP	PATIENT	ENCOUNTER	CODE	DESCRIPTION	UDI↓		
2001-07-04T08:42:44Z		pt0969	en07678	72506001	Implantable defibrillator	(01)6767798860		
2010-10-24T05:28:51Z		pt1013	en27446	72506001	Implantable defibrillator	(01)2018059525		
.↓								



SyntheticMass 1K サンプルデータ

中はこんなデータ#2/4

encounters.csv↓									
Id	START	STOP	PATIENT	ORGANIZATION	PROVIDER	PAYER	ENCOUNTERCLASS	CODE	DESCRIPTION
en43503	2010-01-23T17:45:22	2010-01-23T18:10:28Z	pt0014	org0960	pro5266	pay05	ambulatory	185345009	Encounter for s
en48440	2012-01-23T17:45:22	2012-01-23T18:00:28Z	pt0014	org0515	pro2587	pay05	wellness	162673000	General examin
↓									
imaging_studies.csv↓									
Id	DATE	PATIENT	ENCOUNTER	BODYSITE_CODE	BODYSITE_DESCRIPTION	MODALITY_CODE	MODALITY_DESCRIPTION	SOP_CODE	SOP_DESCRIPTION
im724	2014-07-08T15:35:3	pt0820	en12072	40983000	Arm	DX	Digital Radiography	1.2.840.10008.5.1.4	Digital X-Ray Im
im229	2014-01-22T18:58:3	pt0219	en10745	40983000	Arm	DX	Digital Radiography	1.2.840.10008.5.1.4	Digital X-Ray Im
↓									
immunizations.csv↓									
DATE	PATIENT	ENCOUNTER	CODE	DESCRIPTION	BASE_COST↓				
2010-07-27T12:58:08Z	pt0074	en45669	140	Influenza seasonal inj	140.52↓				
2010-11-20T03:04:34Z	pt1125	en26634	140	Influenza seasonal inj	140.52↓				
↓									
medications.csv↓									
START	STOP	PATIENT	PAYER	ENCOUNTER	CODE	DESCRIPTION	BASE_COST	PAYER_COVERAGE	DISPENSES
2010-05-05T00:26:23Z	2011-04-30T00:26:23Z	pt0633	pay08	en06249	389221	Etonogestrel 68	677.08	0.00	12
2011-04-30T00:26:23Z	2012-04-24T00:26:23Z	pt0633	pay08	en22211	389221	Etonogestrel 68	624.09	0.00	12
,									



SyntheticMass 1K サンプルデータ

中はこんなデータ#3/4

observations.csv↓

DATE	PATIENT	ENCOUNTER	CODE	DESCRIPTION	VALUE	UNITS	TYPE↓
2012-01-23T17:45:28Z	pt0014	en48440	8302-2	Body Height	193.3	cm	numeric↓
2012-01-23T17:45:28Z	pt0014	en48440	72514-3	Pain severity - 0-10 ver	2.0	{score}	numeric↓

↓

organizations.csv↓

Id	NAME	ADDRESS	CITY	STATE	ZIP	LAT	LON	PHONE
org1039	HEALTHALLIANCE	60 HOSPITAL ROAD	LEOMINSTER	MA	01453	42.520838	-71.770876	9784662000
org0454	MOUNT AUBURN H	330 MOUNT AUBURN ST	CAMBRIDGE	MA	02138	42.375967	-71.118275	6174923500

↓

patients.csv↓

Id	BIRTHDATE	DEATHDATE	SSN	DRIVERS	PASSPORT	PREFIX	FIRST	LAST
pt0141	1989-05-25		999-76-6866	S99984236	X19277260X	Mr.	Jose Eduardo181	Gomez206
pt0014	1983-11-14		999-73-5361	S99962402	X88275464X	Mr.	Milo271	Feil794

↓

payers.csv↓

Id	NAME	ADDRESS	CITY	STATE_HEADQUARTER	ZIP	PHONE	AMOUNT_COVERED	AMOUNT_UNCOVERED
pay09	Dual Eligible	7500 Security Blvd	Baltimore	MD	21244	1-877-267-2323	141676.87	119449.83
pay07	Medicare	7500 Security Blvd	Baltimore	MD	21244	1-800-633-4227	4756720.13	890394.37

↓



SyntheticMass 1K サンプルデータ

中はこんなデータ#4/4

payer_transitions.csv↓									
PATIENT	START_YEAR	END_YEAR	PAYER	OWNERSHIP↓					
pt0141	1989	1998	pay08	Guardian↓					
pt0014	1983	2001	pay05	Guardian↓					
↓									
procedures.csv↓									
DATE	PATIENT	ENCOUNTER	CODE	DESCRIPTION	BASE_COST	REASONCODE	REASONDESCRIPTION		
2011-04-30T00:26:23Z	pt0633	en22211	169553002	Insertion of subcutane	14896.56↓				
2010-07-27T12:58:08Z	pt0074	en45669	430193006	Medication Reconciliat	726.51↓				
↓									
providers.csv↓									
Id	ORGANIZATION	NAME	GENDER	SPECIALITY	ADDRESS	CITY	STATE	ZIP	LAT
pro1215	org1039	Tomas436 Sauer652	M	GENERAL PRACTICE	60 HOSPITAL ROAD	LEOMINSTER	MA	01453	42.5208
pro4645	org0454	Suzette512 Monahan736	F	GENERAL PRACTICE	330 MOUNT AUBURN S	CAMBRIDGE	MA	02138	42.3759
↓									
supplies.csv↓									
DATE	PATIENT	ENCOUNTER	CODE	DESCRIPTION	QUANTITY↓				



SyntheticMass 1K サンプルデータ

中身を変換(翻訳)して日本疑似データに見せかけたい計画。

- 日本疑似データ化する項目の出現用語を洗い出し、各カウントすると⇒

- SNOMED対策(案)
 - SNOMEDで入っている用語は、CIELの英語表記等に置き換える。(SNOMEDの翻訳は禁止)

※ CIEL: OpenMRS(オープンソースの電カル)で使われる用語集

allergies	DESCRIPTION	15
conditions	DESCRIPTION	129
devices	DESCRIPTION	3
encounters	ENCOUNTERCLASS	6
encounters	DESCRIPTION	50
imaging_studies	BODYSITE_DESCRIPTION	12
immunizations	DESCRIPTION	18
medications	DESCRIPTION	134
observations	DESCRIPTION	129
procedures	DESCRIPTION	144
All	REASONDESCRIPTION	100

- 正確な置き換えである必要はない。それらしければ良い。

- 日本用語・コードに 翻訳・マッピングする。
- その他データ内容を日本らしくする。名前・住所・etc
- ファイル形式を変更する。(案)レセプト風ファイル+検査結果CSVファイル

最終的にはSyntheticMassではなくSyntheaを使うかも