
Correspondence to:
Ferdinand Dhombres, MD, PhD
UMICS - INSERM U1142
Campus des Cordeliers
5, rue de l’école de médecine
75006 Paris, France
E-mail: ferdinand.dhombres@inserm.fr

Appendix: Content Summaries of Selected Best Papers for the 2017 IMIA Yearbook, Section Knowledge Representation and Management

Banda JM, Evans L, Vanguri RS, Tatonetti NP, Ryan PB, Shah NH
A curated and standardized adverse drug event resource to accelerate drug safety research
Sci Data 2016;3:160026

This open science paper introduces a large, curated, and publicly available resource for adverse drug event (ADE) research: it includes an ADE dataset with the source code and documentation to be used by the research community. This resource (AEO-LUS) derives from the publicly available US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) dataset, covering spontaneous reports of adverse drug events since 2012. The resource also integrates a legacy dataset covering 2004-2012. The final dataset results from a pipeline consolidating all relevant data for ADE research, normalizing different term usage, de-duplicating cases, mapping drugs to RxNorm, mapping drug indications and reactions to MedDRA, and generating drug-outcome pairs with associated statistics. The provided documentation and code is expected to support the evolution of this resource when updates of the FAERS dataset are released.

Integrated Data Repository Toolkit (IDRT). A Suite of Programs to Facilitate Health Analytics on Heterogeneous Medical Data

This paper presents a set of tools developed within the Integrated Data Repository Toolkit (IDRT) German project in order to facilitate heterogeneous data integration in the i2b2 framework. Among various applications, this toolset efficiently allows researchers to design their own analyses. For example, the Mapping Editor of the IRDT Import and Mapping Tool helps to import different formats of data into the current i2b2 terminology. The originality at this step is to allow users to integrate data described in different terminologies (e.g., ICD-10-GM, MedDRA, LOINC, ICP-O, and others) and to build a new repository, with a unique model able to support the representation of these data. The new target ontology is a dedicated model of the data of the project which can be saved for new analyses—with a prior translation of the data from the original terminological model to a new target hierarchy. As a result, IDRT appears as a step forward to the semantization of i2b2.

Greene D, NIHR BioResource, Richardson S, Turro E
Phenotype Similarity Regression for Identifying the Genetic Determinants of Rare Diseases
Am J Hum Genet 2016;98(3):490-9

This paper is dedicated to model the association between multiple phenotypic traits described with the Human Phenotype Ontology (HPO) and the corresponding genotype, in the specific context of rare disease (rare variants). HPO allows composite phenotypes to be represented systematically but association methods accounting for the ontological relationship between HPO terms do not exist. The authors propose a Bayesian method to model the association between HPO-coded phenotypes and genotypes. The method uncovers associations between rare genotypes and the similarities between patients’ phenotypes and a latent characteristic phenotype. The effectiveness of the approach is demonstrated on a simulation study and on a real dataset from the BRIDGE project.