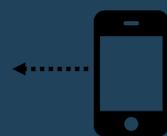


Standardization, automation and monitoring system for ETL process to improve CDM data quality

Dahye Shin, Seol Paik, Sehee Chang, Junghyun Do, Youmi Lee, Hye jin Kam

'EvidFormer' has standardized on ETL tasks and is an all-in-one solution that enables data validation and monitoring, increasing work efficiency, simplifying management and reducing work time.



Take a picture to download the full paper

Conclusion

The end users can get answers and explanations for their fundamental questions whether the CDM data is reliable or how the local source data enters the CDM from our proposed solution.

Layer definition in conversion process

Layer step	Definition
Interface table	This table is most similar to the source data and contains all the meta information needed to convert to CDM.
MT_rule table	Table that classifies data to be removed from interface table and applies conversion rule.
Meta table	This table is the most similar to the CDM table after all the conversion rules are applied, the data to be excluded is deleted.
CDM table	This table joins the vocabulary mapping table from the meta table and converts it into standard vocabularies and the CDMv5.3.

Data example

Data filter list and example

- 1) Classification of canceled data
ex) Discharge order, cancel_yn
- 2) Classification of data that has not been enforced
ex) PRN(pro re nata), acting_yn
- 3) Classification of test data
ex) test_yn, valid_yn, not person_id
- 4) Classification of data with date errors
ex) data after the observation period
startdate > enddate
startdate < birthdate
startdate > deathdate

Apply conversion rules list and example

- 1) Prioritize data selection
ex) provider_id -> medical_dr, admission_dr, acting_dr
- 2) Duplicate data unification
ex) One patient may be assigned multiple patient_id
One provider may be assigned multiple provider_id
- 3) Calculated date or quantity

Acknowledgement

This work was supported by the Bio Industrial Strategic Technology Development Program (20001234) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea)

INTRODUCTION

- Evidnet is a governmental project participating company that is responsible to deliver the CDM conversions for 43 hospitals in Korea.
- In order to deliver proper data quality management, standardized process and the efficient conversion, we proposed a DQ(Data quality) check system.

MAJOR CHALLENGE

1. Conversion process standardization

We have standardized the process so that any hospital data can be converted through the same conversion process.

A layer was constructed according to the purpose and characteristics of the conversion path from source data to CDM data.

2. Data validation

We has developed a rule check package for each domain by layer. If the package finds an error data during the conversion process, it corrects the error and then continues to convert it to the final CDM. This eliminates unnecessary processes such as re-conversion and satisfies data quality and time efficiency at the same time.

3. Process automation and visualization

Scheduling and automation service of ETL process was developed based on standardized conversion path. This eliminates human errors that can occur during the conversion process. And we are developing a monitoring service to visualize the conversion and the data verification results for work management

