



BEYOND SUBMISSION

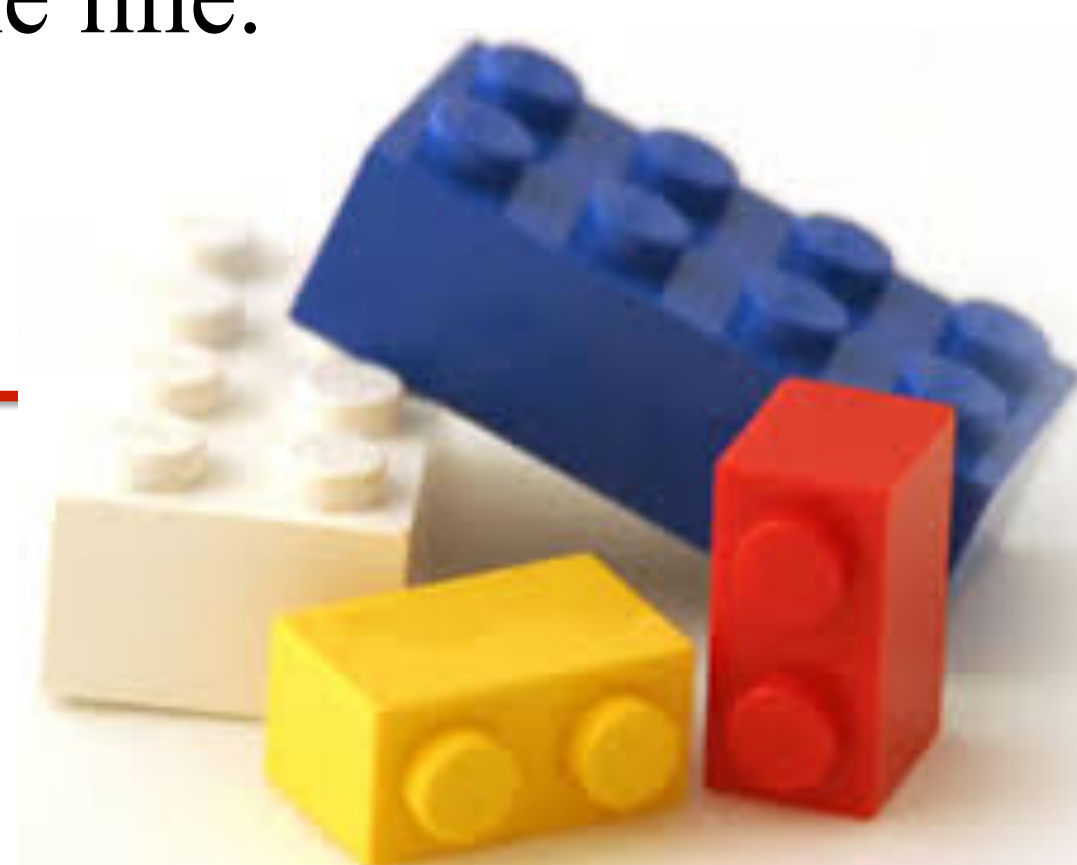


Business rules in legacy data conversion for better CTR integration

Helena Sviglin¹, Eileen Navarro¹, Bobbie Witzcak¹, Lilliam Rosario¹

BACKGROUND:

The Janus Clinical Trials Repository (CTR) stores standardized clinical trial data at FDA CDER. One use of the CTR is to aggregate data across studies and applications. While adherence to the SDTM standard is essential to ensure effective aggregation, limitations to the semantic interoperability* in this standard can impede aggregation – even when data is perfectly adherent. Some of these limitations came to light when FDA CDER converted legacy data from 51 diabetes pivotal trials into SDTM v3.1.2 to answer diabetes-related safety and efficacy questions. To overcome these limitations, over 50 business rules were developed and deployed during conversion. A few of these business rules are described below the line.



RESULTS AND DISCUSSION:

The following rules listed below are a subset of the 50+ rules developed during LDC efforts

Dates

All date comparisons are done to account for partial dates and times.

DY values (--STDY, --ENDY, --DY) are calculated independently to those submitted based on a reference start date that represents date of first exposure to the study drug. These new DY values allow data across studies to be stacked and pooled meaningfully (i.e., a 12-week visit is always 12 weeks from first exposure instead of randomization or onset of a titration phase).

Study Population

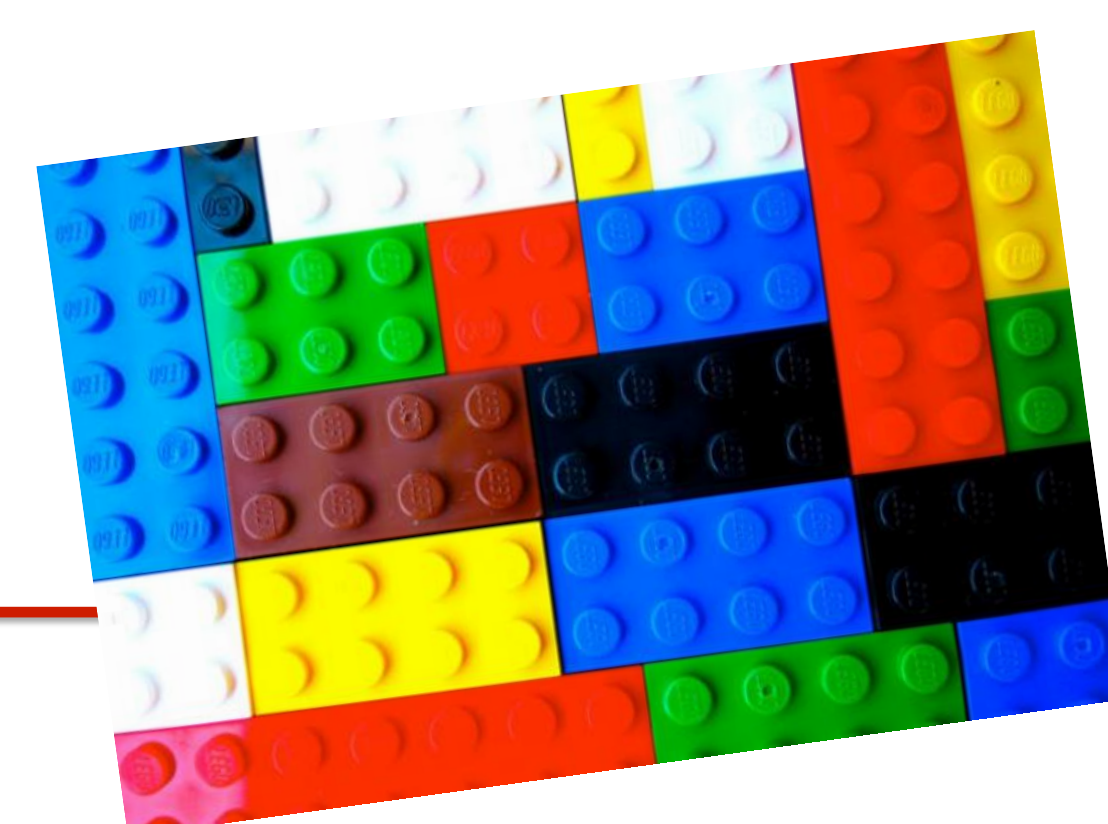
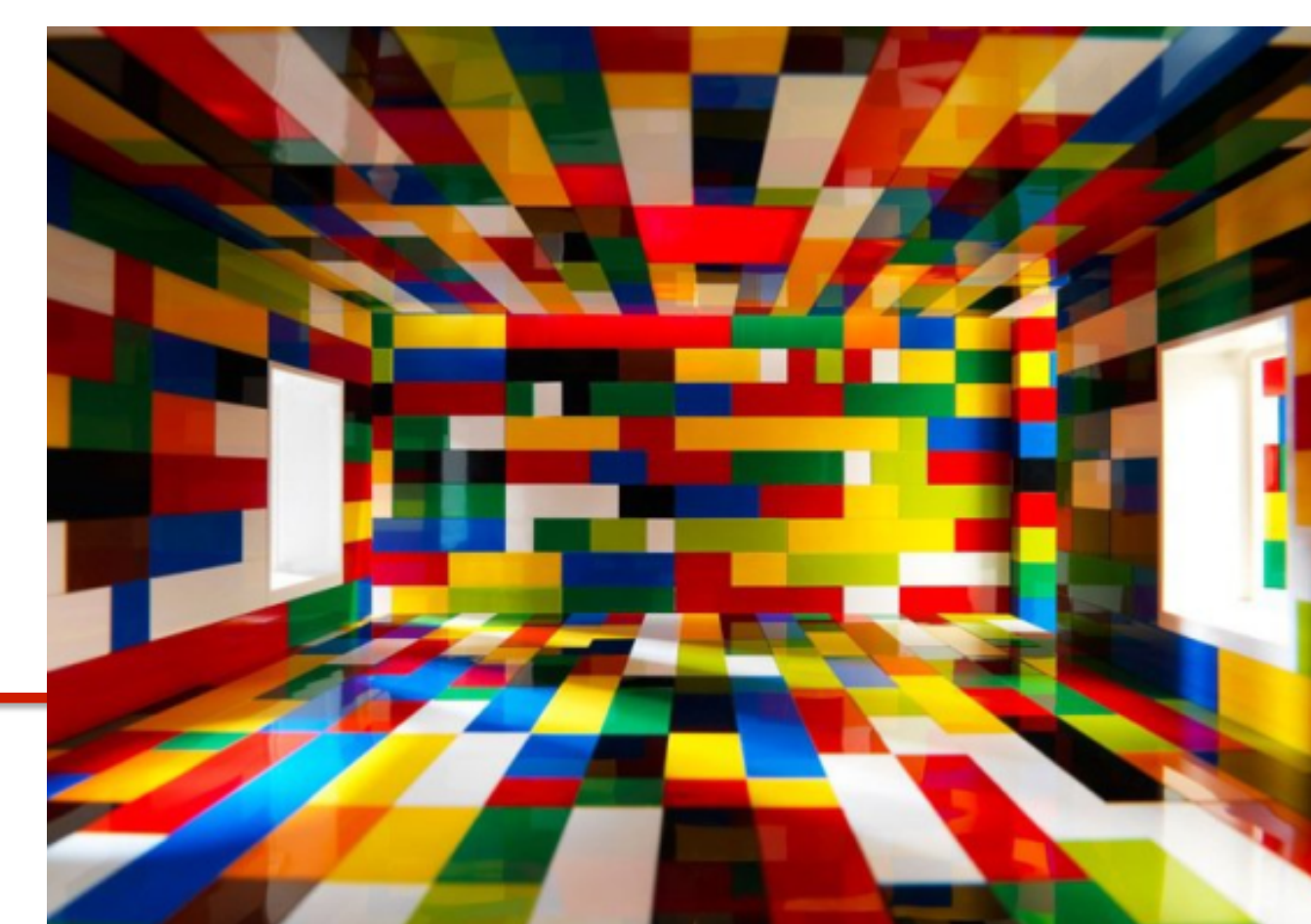
Population flags extant in the legacy data are moved to SUPPDM to enable reconciliation of clinical study reports and legacy data against SDTM (i.e., Completers, Intent-to-Treat, Per-Protocol, Safety.)

Whenever a waiver was granted for a subject to be enrolled in a study, a SUPPDM record is created.

Subject initials present in legacy data are never stored in the CTR.

METHODS:

Prior to conversion, the legacy data for each study was manually inventoried and each data element identified. Each element was mapped to an SDTM domain. The data elements in each domain were then examined and overlapping concepts were unified to a single representation. Each unification under a single concept was considered a potential business rule. Each potential business rule was further defined. Legacy data was mapped to SDTM according to these defined business rules.



Supplemental Qualifiers

All values of QLABEL are self-explanatory and do not depend on context or content of the parent record to understand the contents of the supplemental record. (For example, if a SUPPAE record contains information about whether an event was medically assisted, the QLABEL is ‘Was AE medically assisted?’ rather than ‘Medically Assisted?’)

RACE

Legacy values for race other than an accepted category (i.e., a nationality like ‘AUSTRALIAN’, or poorly defined terms like ‘GYPSY’) are changed to the value of ‘UNKNOWN’ and the original value is not stored in SUPPDM.

ARM and ACTARM values

A consistent convention applies to all ARM values across studies:

study drug generic name + dosage amount + frequency of dose

study drug generic name + titration (range from min to max) + frequency of dose

Where study drug is titrated, the word ‘titration’ appears in the ARM value. Where dosage is titrated prior to randomization, ‘Pre-treatment titration’ is used. Where a study drug is background therapy and no specific dosage is given, ‘OT’ (Optimized Therapy) is used in the ARM value.

CONCLUSION:

Conversion of legacy data is technically challenging and resource intensive as trials may differ in design and data may be in disparate formats. Data curation, trial comparison for common endpoints and time points, and harmonization of data elements to enable cross-trial examination is needed to fully realize the potential of the trial data repository. Data that is adherent to standards is the necessary first step in data pooling. Business rules like these presented here drive data toward semantic interoperability* and should at least be considered as the SDTM standard evolves.



Adverse Events

If an adverse event is adjudicated, this data is added as a unique row to AE and signified by a value of ‘ADJUDICATORx’ in the AEEVAL variable. AEGRPID is utilized to link the row with the AE to the row(s) with its related adjudication(s). The values in AEEVAL and AEGRPID can be used to prevent over-counting. Any data element from adjudication which cannot be added to the adjudication record in AE is migrated to SUPPAE.

Records in AE indicating withdrawal, interruption, or change in drug dosage should have an equivalent record in DS.

Death

Deaths in the DS domain should be one per subject.

A row with --TERM equal to ‘DEATH’ should be in both AE and DS.

The number of unique deaths should be equal in both AE and DS.

TEST

Where CRF questions have a prespecified response of ‘Other, specify’ the response is represented in the data as ‘Other, specify: xxxxxxxx’.

* For more on what the Agency has to say on Semantic Interoperability see:
<http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf>

¹ U.S. Food and Drug Administration/CDER/OTS/CSC