

Constructing a Simple Data Repository

Magnus Mengelbier
Managing Director, Limellogic

Topics

- Introduction
- Cheryl's Birthday
- Data Repository
- Early Development
- Central Data Repository
- Summary

Cheryl's Birthday

24. Albert and Bernard just become friends with Cheryl, and they want to know when her birthday is. Cheryl gives them a list of 10 possible dates.

May 15	May 16	May 19
June 17	June 18	
July 14	July 16	
August 14	August 15	August 17

Cheryl then tells Albert the month and the day of her birthday respectively.

Albert: I don't know when Cheryl's birthday is, but I know that Bernard does not know too.

Bernard: At first I don't know when Cheryl's birthday is, but I know now.

Albert: Then I also know when Cheryl's birthday is.

So when is Cheryl's birthday?

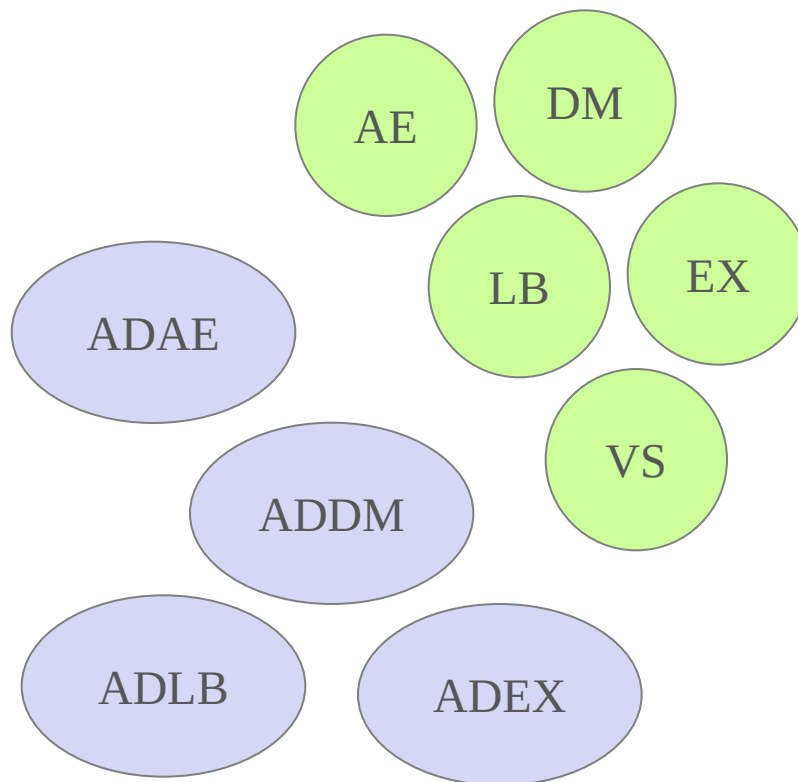
KENNETH KONG

Cheryl's Birthday: Singapore's maths puzzle baffles world

<http://www.bbc.com/news/world-asia-32297367>

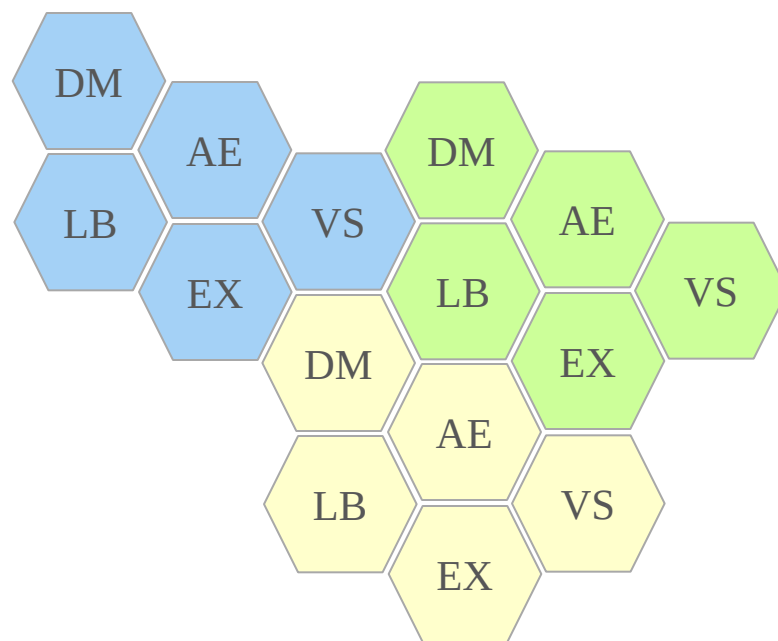
Origins

- Study data
 - Source data sets
 - Analysis data sets
- Data silos
 - Interim analysis
 - Data Monitoring Committees
 - Investigator Brochure
 - IND
 - IND update
 - Safety monitoring
 - That request
 - The poster
 - Clinical Study Report



Pooling

- Study data
 - Source data sets
 - Analysis data sets
- Integrated Analysis
 - ISE
 - ISS
 - Regulatory submissions
- Meta-Analysis
 - Safety
 - Epidemiology
 - Other



Pooling is Data

View: Column names
🔍
📄
🔄
📊
🔗
🔧
Filter: (none)

Columns

☒ Select all

☒ A Make

☒ A Model

☒ A Type

☒ A Origin

☒ A DriveTrain

☒ 123 MSRP

☒ 123 Invoice

☒ 123 EngineSize

☒ 123 Cylinders

☒ 123 Horsepower

☒ 123 MPG_City

☒ 123 MPG_Highway

☒ 123 Weight

☒ 123 Wheelbase

☒ 123 Length

Property	Value
Label	MSRP
Name	MSRP
Length	8
Type	Numeric
Format	DOLLAR8.
Informat	

Total rows: 428 Total columns: 15

⏪
⏩

	Make	Model	Type	Origin	DriveTrain	MSRP	Invoice
1	Acura	MDX	SUV	Asia	All	\$36,945	\$33,337
2	Acura	RSX Type S 2dr	Sedan	Asia	Front	\$23,820	\$21,761
3	Acura	TSX 4dr	Sedan	Asia	Front	\$26,990	\$24,647
4	Acura	TL 4dr	Sedan	Asia	Front	\$33,195	\$30,299
5	Acura	3.5 RL 4dr	Sedan	Asia	Front	\$43,755	\$39,014
6	Acura	3.5 RL w/Navigation 4dr	Sedan	Asia	Front	\$46,100	\$41,100
7	Acura	NSX coupe 2dr manual S	Sports	Asia	Rear	\$89,765	\$79,978
8	Audi	A4 1.8T 4dr	Sedan	Europe	Front	\$25,940	\$23,508
9	Audi	A41.8T convertible 2dr	Sedan	Europe	Front	\$35,940	\$32,506
10	Audi	A4 3.0 4dr	Sedan	Europe	Front	\$31,840	\$28,846
11	Audi	A4 3.0 Quattro 4dr manual	Sedan	Europe	All	\$33,430	\$30,366
12	Audi	A4 3.0 Quattro 4dr auto	Sedan	Europe	All	\$34,480	\$31,388
13	Audi	A6 3.0 4dr	Sedan	Europe	Front	\$36,640	\$33,129
14	Audi	A6 3.0 Quattro 4dr	Sedan	Europe	All	\$39,640	\$35,992
15	Audi	A4 3.0 convertible 2dr	Sedan	Europe	Front	\$42,490	\$38,325
16	Audi	A4 3.0 Quattro convertible 2dr	Sedan	Europe	All	\$44,240	\$40,075
17	Audi	A6 2.7 Turbo Quattro 4dr	Sedan	Europe	All	\$42,840	\$38,840
18	Audi	A6 4.2 Quattro 4dr	Sedan	Europe	All	\$49,690	\$44,936
19	Audi	A8 L Quattro 4dr	Sedan	Europe	All	\$69,190	\$64,740
20	Audi	S4 Quattro 4dr	Sedan	Europe	All	\$48,040	\$43,556
21	Audi	RS 6 4dr	Sports	Europe	Front	\$84,600	\$76,417
22	Audi	TT 1.8 convertible 2dr (coupe)	Sports	Europe	Front	\$35,940	\$32,512
23	Audi	TT 1.8 Quattro 2dr (convertible)	Sports	Europe	All	\$37,390	\$33,891
24	Audi	TT 3.2 coupe 2dr (convertible)	Sports	Europe	All	\$40,590	\$36,739
25	Audi	A6 3.0 Avant Quattro	Wagon	Europe	All	\$40,840	\$37,060
26	Audi	S4 Avant Quattro	Wagon	Europe	All	\$49,990	\$44,446

Standards

- SAS Data Sets

- Character
- Numeric

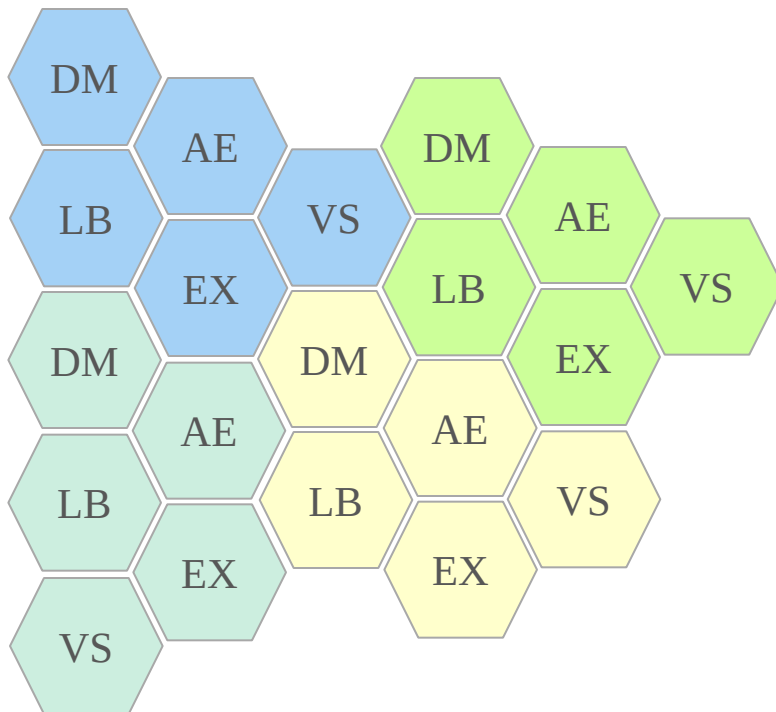
- (SAS) Data Sets

- Identifier(s)
- Text
- Catagorical Result
- Continuous Result
- Date/Time

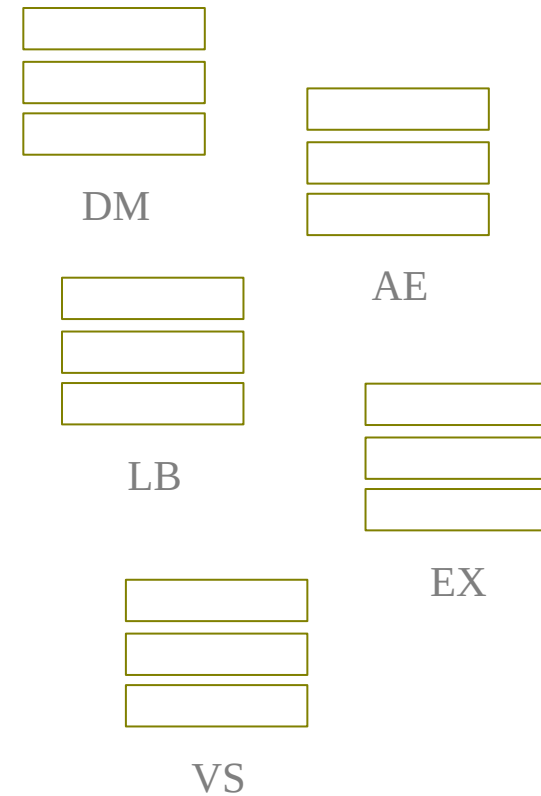
- Standards

- SDTM 1.4
- SDTM IG 3.2
- ADaM 2.1
- ADaM IG 1
- ADaM Results Metadata 1
- ADaM Validation Checks 1.3
- ADaM ADAE 1
- CDASH 1.1
- CDASH SAE Supplement 1
- CDASH UG 1

Real Life



≠



Information Standards

- Identifier(s)
 - General
 - Intervention(s)
 - Event(s)
- Content
 - Text  Text
 - Catagorical Result  Text
 - Catagorical Result  Number
 - Continuous Result  Whole number
 - Continuous Result  Number with decimals
 - Date/Time  ISO 8501

Repository Data Standard

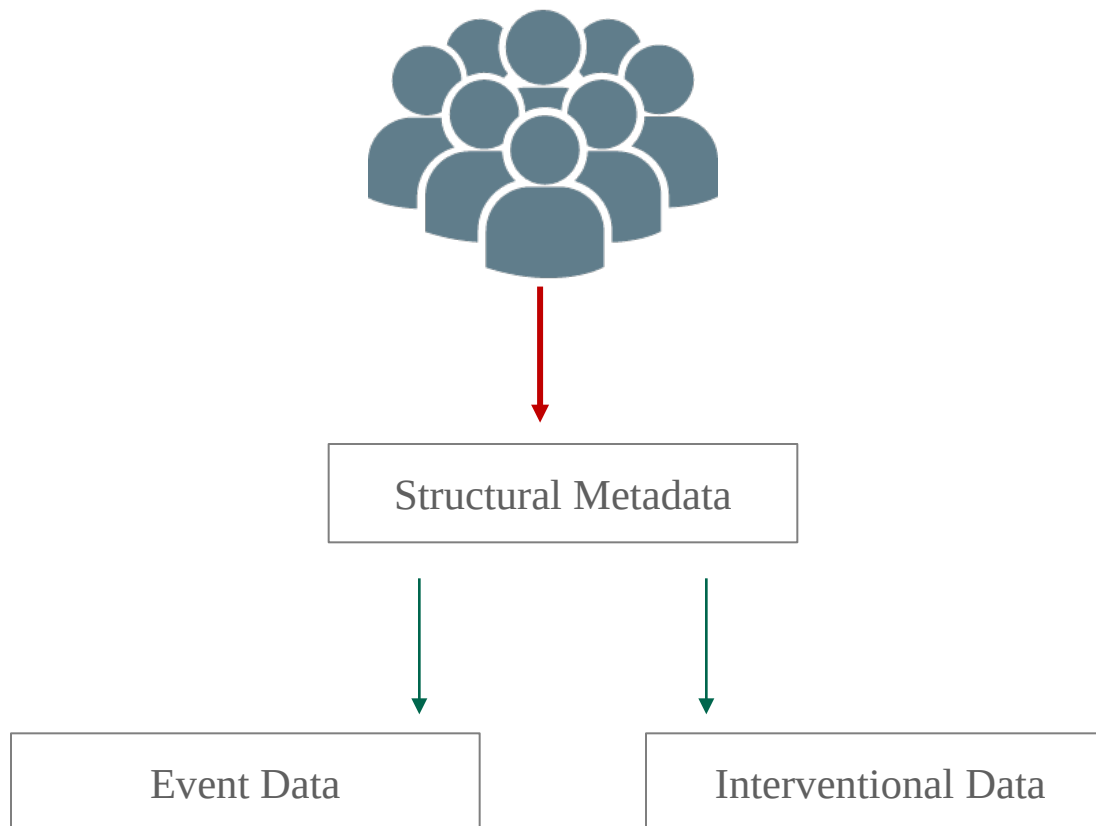
- Identifying information

- Context
- Domain
- Subject Identity
- Event identifiers
- Intervention identifiers
- Item reference

- Observations

- Text
- Numbers
 - ▣ Default format
 - ▣ Application specific format
- Date / Time

Data Repository Structure



Data Repository Structure (Example)

[illegible]

First Attempt

- Study level repository

- Store Data

```
%sdr_data_load( data = DM AE LB EX VS,
                repository = sdrlib );
```

- Retrieve Data

```
%sdr_data_read( data = DM AE EX,
                 out = WORK,
                 format = native,
                 repository = sdrlib );
```

Second Attempt

- Project / Compound level
- "Smarter" loading data
 - Loading data

```
%sdr_data_load( path = /path/to/my/study,
                 data = DM, type = subjectlevel,
                 repository = sdrlib );

%sdr_data_load( path = /path/to/my/study,
                 data = AE, type = event,
                 repository = sdrlib );

%sdr_data_load( path = /path/to/my/study,
                 data = EX LB VS, type = intervention,
                 repository = sdrlib );
```

- Reading data from repository unchanged

Central Data Repository

- Repository storage as an Oracle database
- Loading from SAS unchanged

```
%sdr_data_load( path = /path/to/my/study,
                data = DM, type = subjectlevel,
                repository = sdrdb );
```

- Suddenly it works with R

```
library( sdr )

sdrcon <- sdrconnect( ... )

sdrdataread( sdrcon, path = /path/to/my/study, data = DM
EX PK VS )

sdrdisconnect( sdrcon )
```

Central Data Repository Features

- Load data
- Read data
- Transactions
- Snapshots
- Version control
- Access control
 - Roles
 - Groups
 - Grant / Revoke permission on an individual item
- Views
 - Snapshots of data
 - De-identified data

Additional Features and Improvements

- Load
 - Data set specifications
 - Validating/Enforcing data set structure
 - Transactions
- Read
 - Select specific items (variables)
 - Complex query / filter
- Improve access from BI tools
- Concurrent access to data sets
- Scalability

Summary

- A simple data repository
- Start small and scale
- Allow multiple concurrent standards
- Standards managed by the business process
- Access from different applications
- Does not have to be a Clinical Data Warehouse

Simple Data Repository will be released as an
Open Source package

code.phuse.com

December 2015

Contact Information

- Introduction
- Cheryl's Birthday
- Repository
- Pools, Lakes and Oceans
- Solutions
- Summary

Magnus Mengelbier

Limelogic

+46 40 685 1481

mmr@limelogic.com