

If data is the new oil, how do you make the most of your reserves

PHUSE Connect US 2023, Orlando, FL

Simon Luo, Data scientist
Ashley Mao, Data scientist

Agenda

- Data Reuse
- Introduction to Data Mart
- Example: Lupus Data Mart
- Demo
- Summary
- Q&A

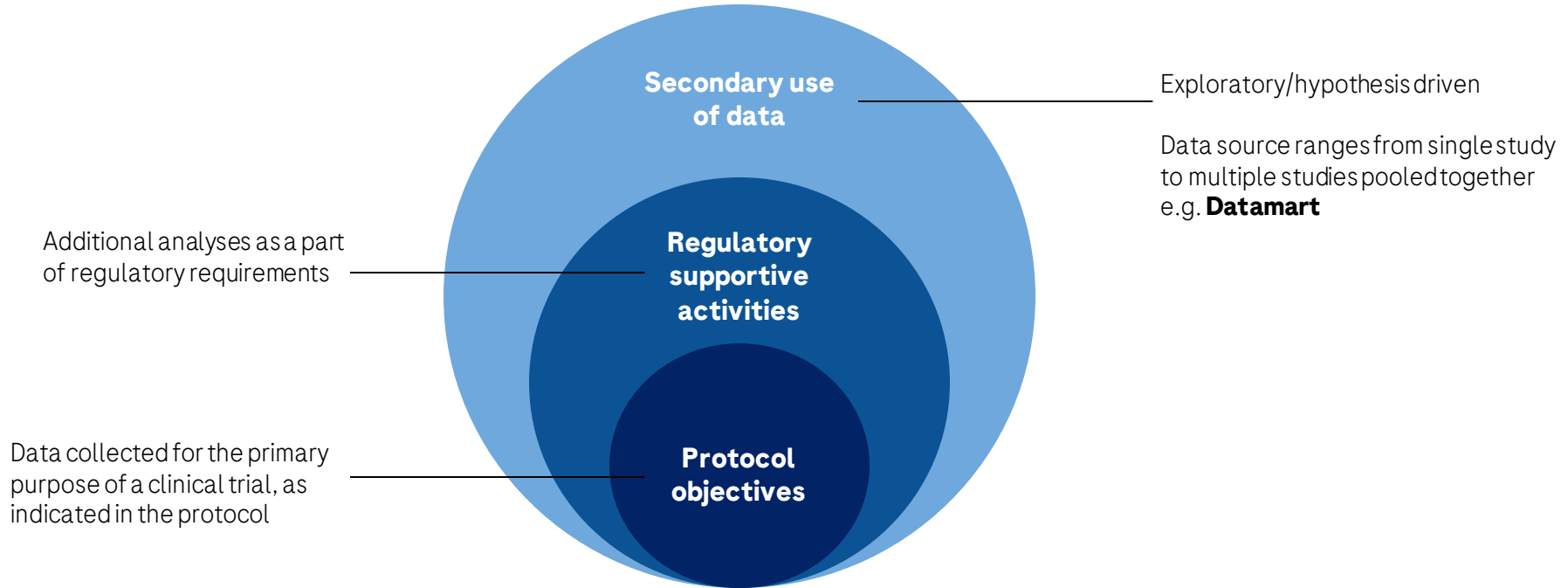


Data Reuse

The Data Economy



How Data is Used/Reused in Pharmaceutical Industry



FAIR Principles



Findable

"I know where data is"



Accessible

"I can access data I need"



Interoperable

"I use consistent standards & terms"



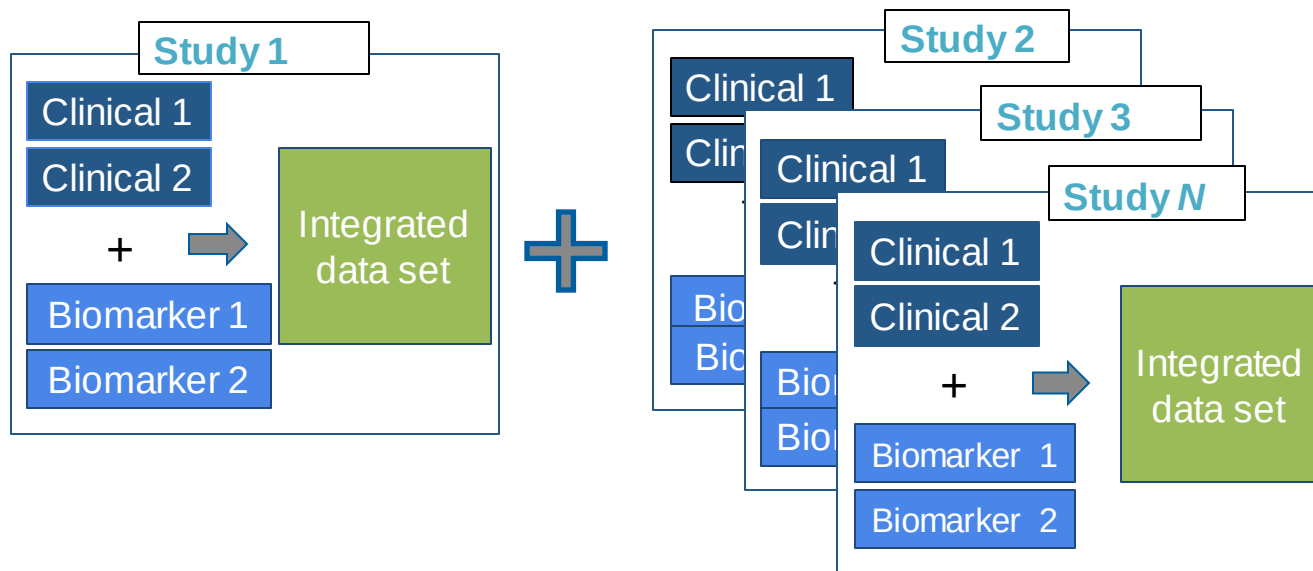
Reusable

"I use existing data to answer new questions"

Introduction to Data Mart

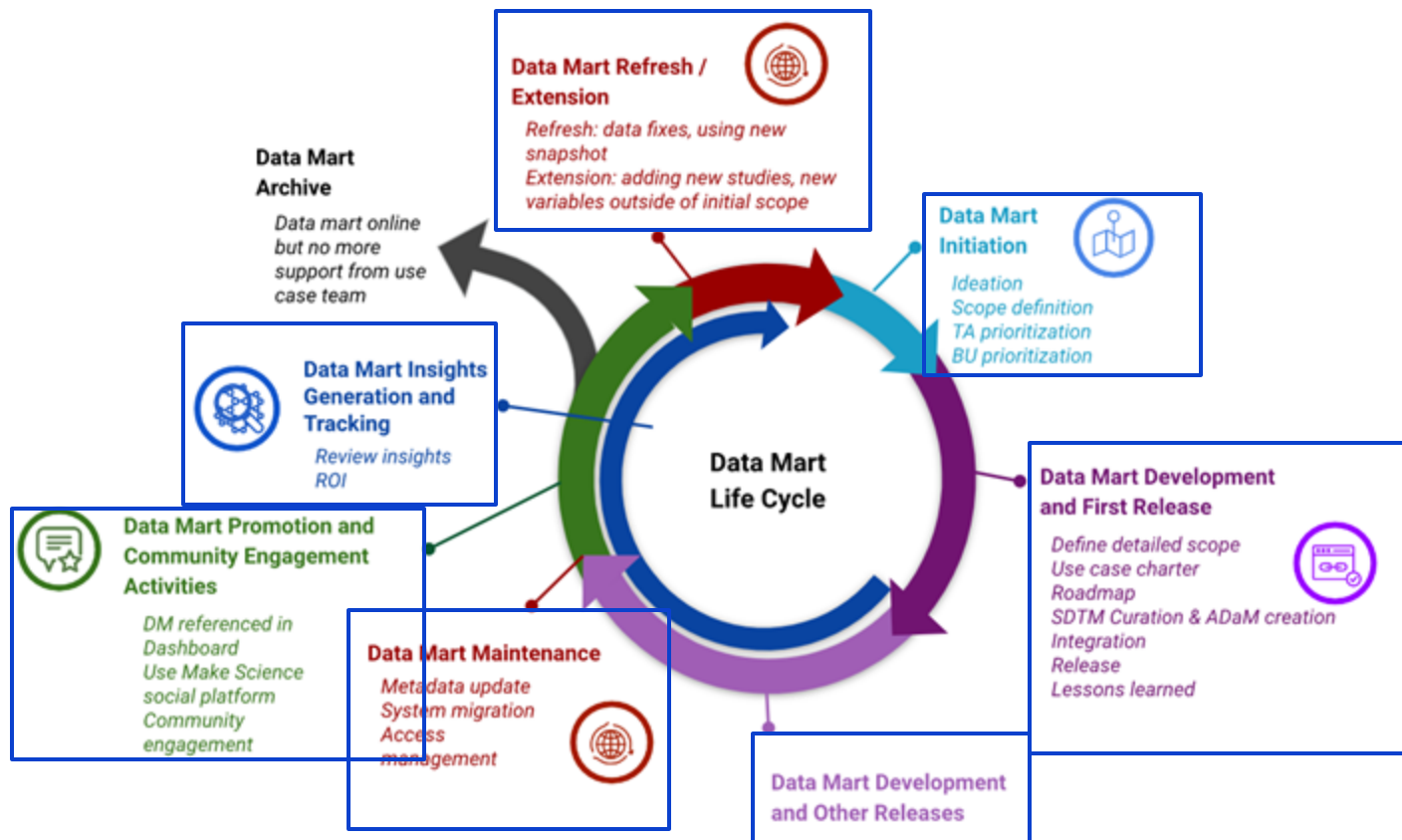
What is a Data Mart?

DataMart = multiple *curated, harmonized and integrated analysis-ready* datasets from multiple studies



Part of the Enhanced Data & Insights Sharing (EDIS) program using FAIR data principles

Data Mart Life Cycle

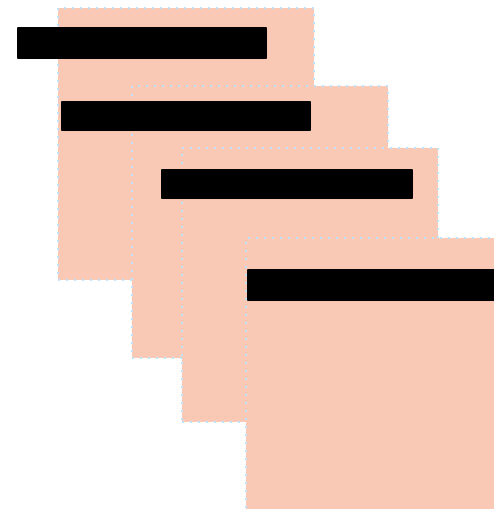


Lupus Data Mart

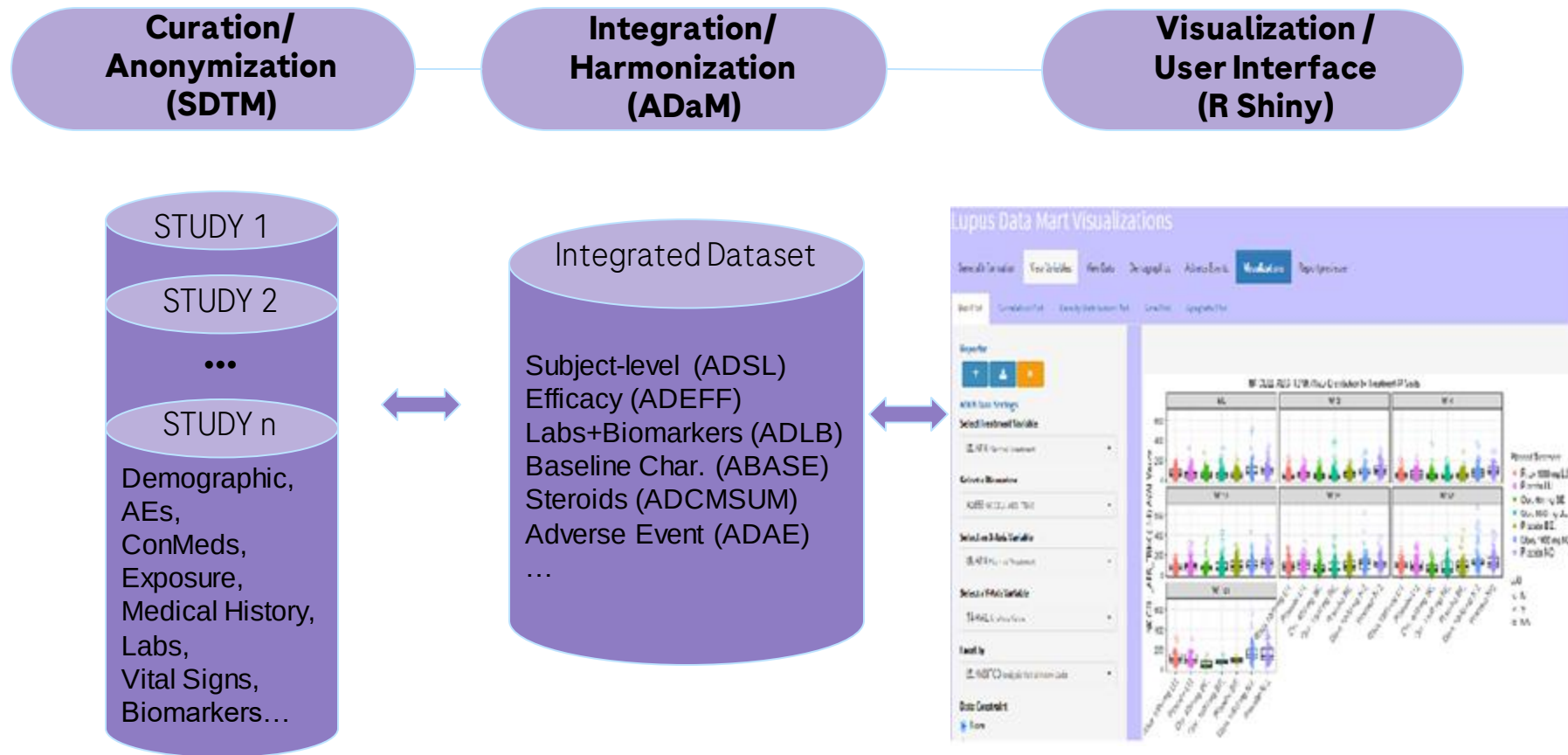
Why do we need a Lupus DataMart?

Currently, our institutional knowledge of lupus patients' data mainly comes from >23,000 pages of *non-queryable CSRs*

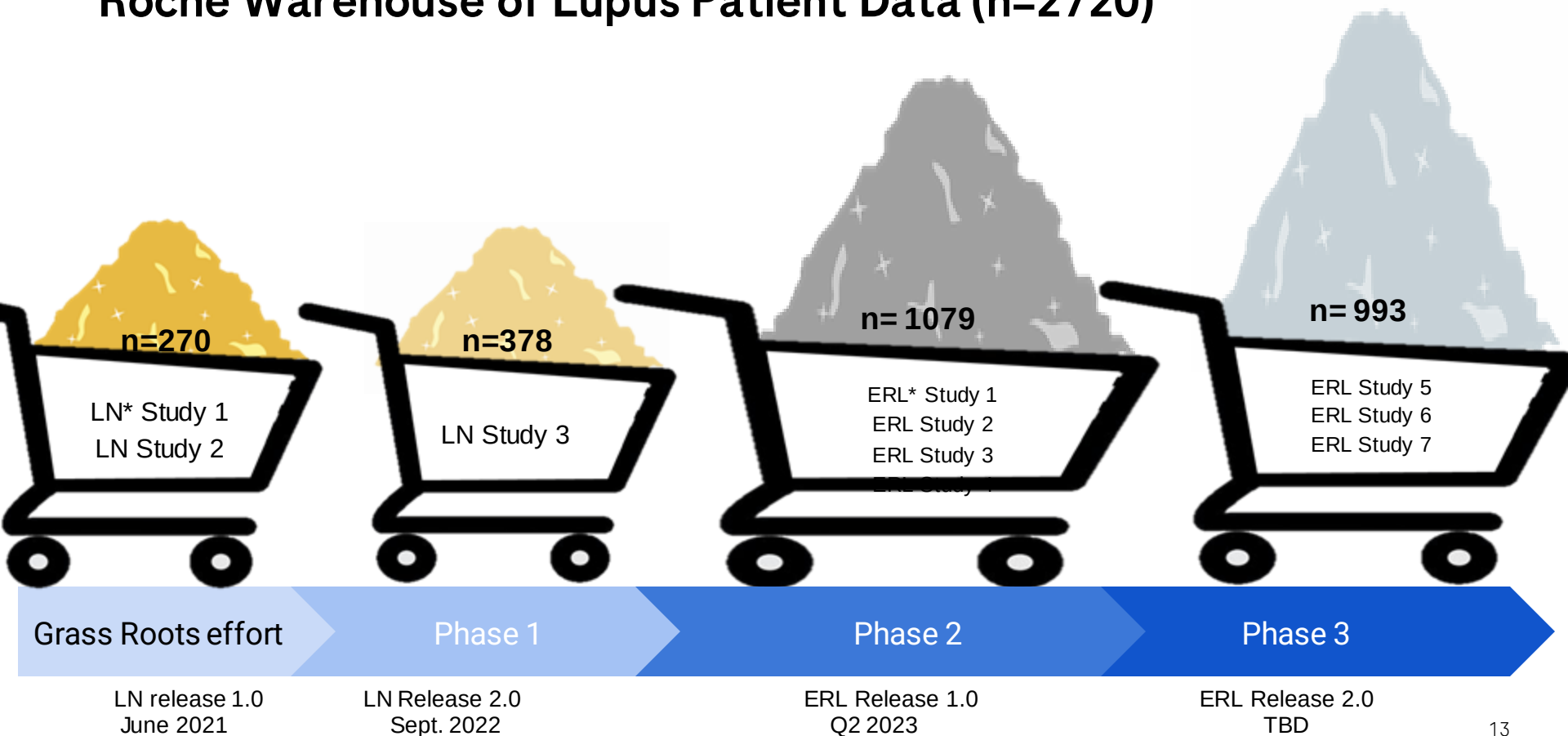
+ *Untapped Transcelerate Lupus Clinical Data (4 Ph3 studies)*

Data Flow in Lupus Data Mart



Roche Warehouse of Lupus Patient Data (n=2720)



* LN: lupus nephritis; ERL: extrarenal lupus

Key Challenges & Learnings

Harmonizations

Endpoints:

- Complete renal response def.
- Rescue medications list
- Escape therapy def.
- Flare definitions

Biomarker data:

- B cell subset with different names, codes and units
- LLoQs

Legacy v.s CDISC

- Control terminology terms
- Lab tests & results



Documentations & Systems/Processes

Documentations:

- Key decisions documented for traceability
- ADaM Derivation rules (spec)
- User guidance & Access control documents
- Data dictionary

Systems/processes:

BEE/BAA/entimICE/R
Shiny app/CDSE/EDIS
Dashboard/Anonymization process...



Communications

Communications:

- Biostats/scientists to define the datasets/variables of interest
- Curator/Integrator/scientists to come up Biomarker/Lab tests harmonization rules per DSG¹
- Team discussion on the visualization tool with App developer
- Release communication with EDIS² group and the wider user groups



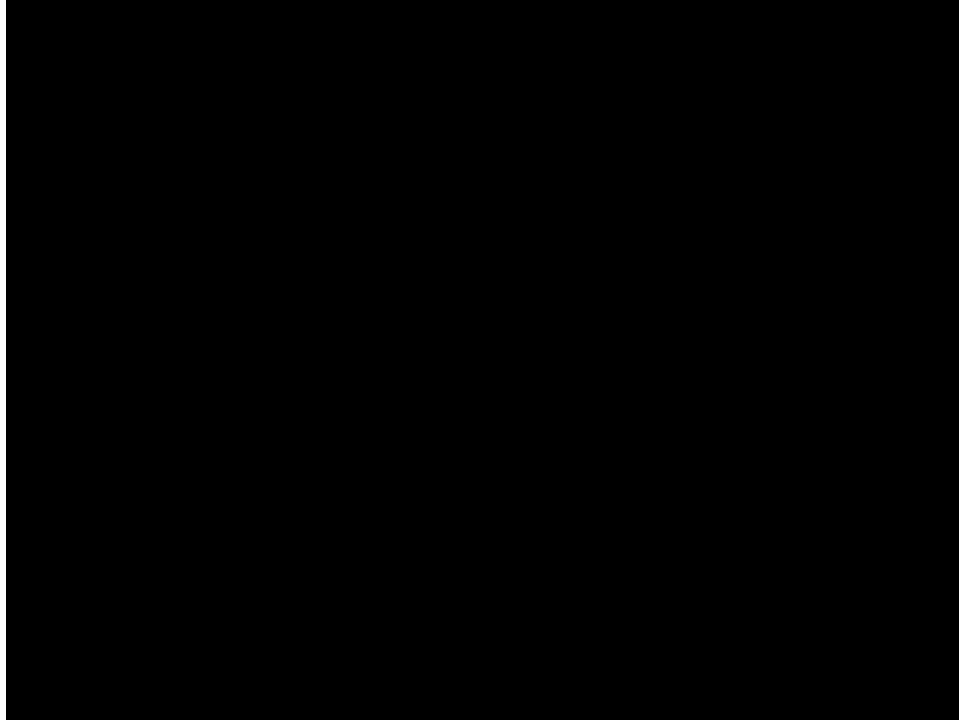
Anonymization

Data privacy / Sharing considerations

- Data linkage broken causing ADaM program malfunction
- Ongoing study access license

Data Mart - Demo

Demo

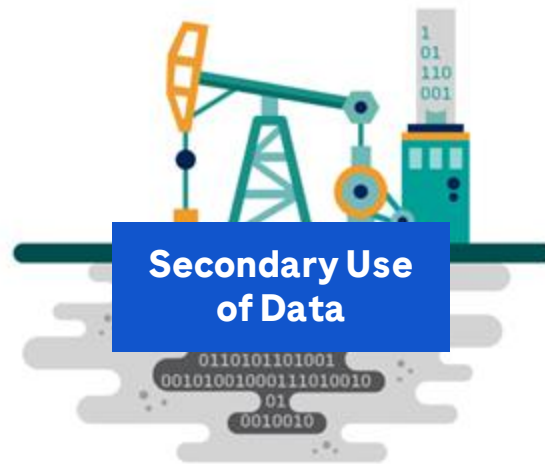


Link: <https://insightsengineering.github.io/teal>

Summary

F.A.I.R. principles

- Findable
- Accessible
- Interoperable
- Reusable



Learnings

- Harmonization, documentation, communication, anonymization
- Improvements to be implemented in data collection and SDTM mapping in clinical trials



Goal

- Accelerate scientific & medical advances
- Deliver treatments to patients



Application

- Understand MoA & differentiation
- Clinical development program
- Clinical response to therapeutics and PBO/SoC



Acknowledgement

Lupus Data Mart Core Team

Core Team			
OMNI-Biomarker	I2O Clinical	Data & Statistical Sciences	Data Strategy & Delivery
Julie Rae (use case lead) Harini Raghu <i>*Alyssa Morimoto</i> <i>*Mike Looney</i>	Justine Maller <i>*Matt Cascino</i>	Nick Paszty Simon Luo Sydney Lee Ashley Mao Vlad Khudov <i>*Alex Rybalko</i> <i>*Bali Toth</i> <i>*Betty Wang</i>	Annie Guevara Sudhamsh Dandamraj Ann Lam Daniel Sana Hanzal Muhusin Bonita Viegas Monteiro Xiaodong Peng Francois Sadok Brian Bedlington <i>*Adetola Odupitan(PM)</i> <i>*Josh Bernal</i> <i>*Stella Banjo</i> <i>*Lisa Thompson</i> <i>*Dony Unardi</i> <i>*Li Ma</i> <i>*Jiaowei Tang</i>
Key Collaborators			
Data Sharing & Privacy	Data Sciences Acceleration	External Controls Working Group	
Rebecca Sudlow Katherine Tucker	NEST Team	Melanie Poulin-Costello	



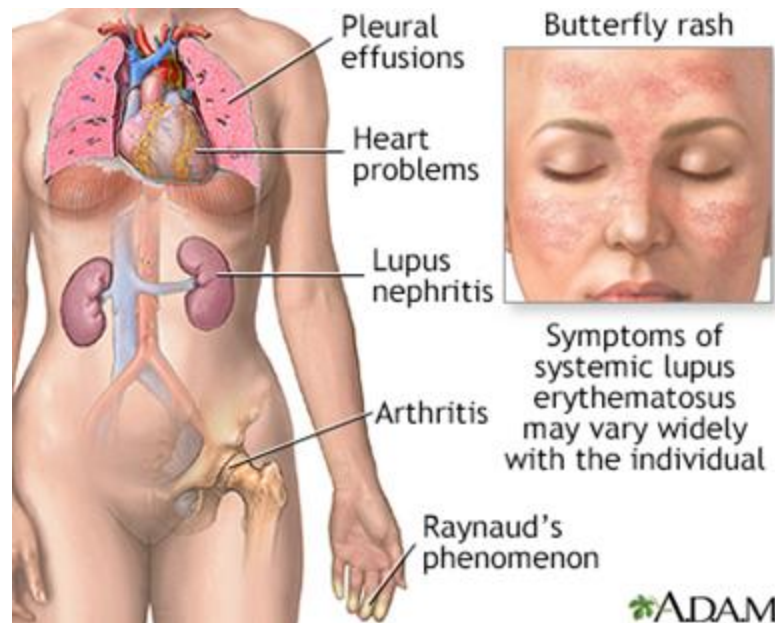
Doing now what patients need next

Backup slides



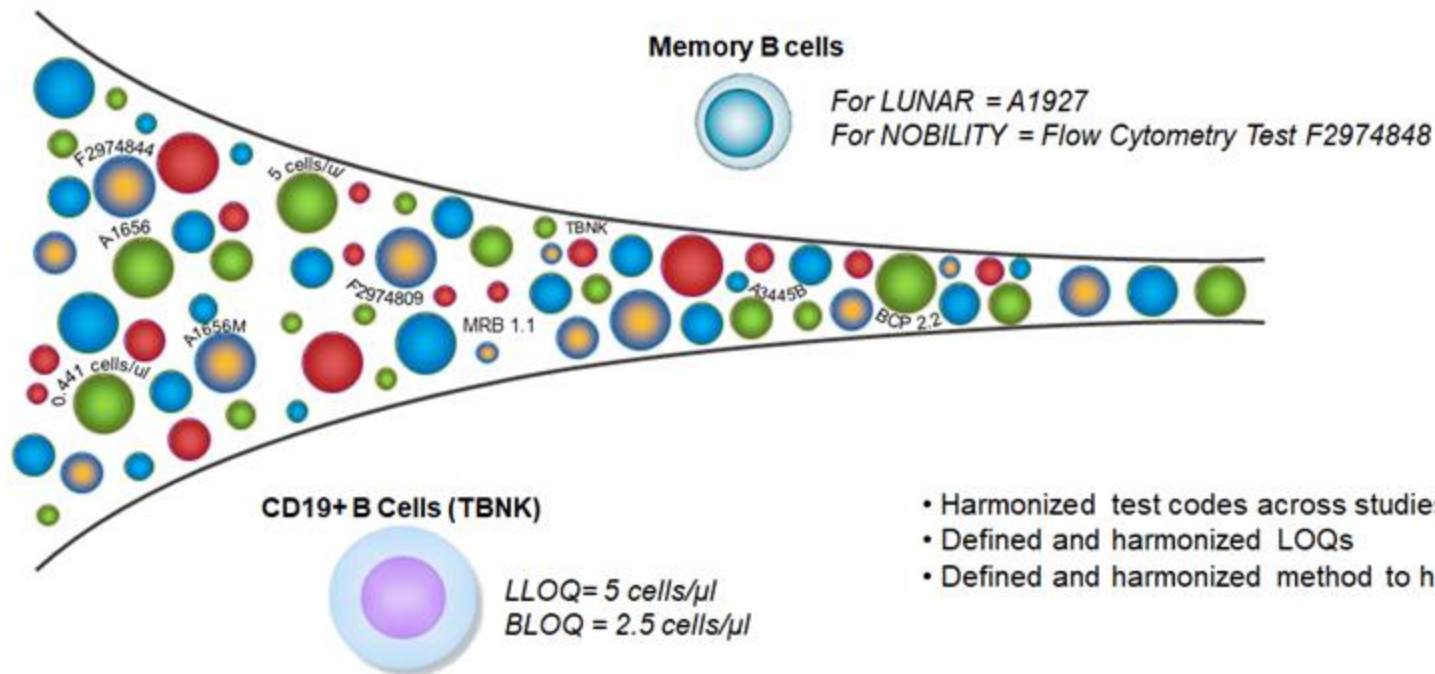
LUPUS Disease Introduction

- An autoimmune rare disease with unknown cause
 - It can affect any organs/tissues, but mostly skin, joints and Internal organs like kidneys and heart
- Female to Male ratio $\geq 9:1$
- Highest prevalence among ethnic groups.
 - Black to Caucasian ratio = 3:1



Lab tests harmonization- flow cytometry was complex

- 3 possible flow assays per study with up to 44 reportables
- Different test codes and variable names used for the same cell population between studies
- LLOQ values and handling of BLOQ values not consistent or not defined



- Harmonized test codes across studies
- Defined and harmonized LOQs
- Defined and harmonized method to handle BLOQ values

Systems and Data Flow

