Traceability and AI for Improved Understanding, Communication and Quality Control of Clinical Trials

Karl Brand & Tomas Sabat Stöfsel

2024-05-30

https://github.com/Bayer-Group/phuse-verisian

Disclaimer

Neither I nor my employer receive any payment or benefits from Verisian.

This presntation may contain forward-looking statements based on current assumptions and forecasts made by Bayer management. Various known and unknown risks, uncertainties and other factors could lead to material differences between the actual future results, financial situation, development or performance of the company and the estimates given here. These factors include those discussed in Bayer’s public reports which are available on the Bayer website at www.bayer.com. The company assumes no liability whatsoever to update these forward-looking statements or to conform them to future events or developments.

The written and expressed views in this presentation are entirely those of the presenter and are not intended to be representative of my employer, Bayer Aktiengesellschaft.

Contents

Clincial Statistical Programming

I. Introduction

II. Improving Today

III. Building the Future

I. Introduction

Why are we here?

To **communicate** our quantitative measures of therapy safety and efficacy in patients to third parties for evaluation of health value.

To change for the better how we do this.

Statistical progamming

[as-is]

a. People: statistical programmers / **[SAS ninjas]**

https://github.com/Bayer-Group/phuse-verisian

Statistical progamming

[as-is]

a. People: statistical programmers / **[SAS ninjas]**

b. Actions: work with a mountain of SAS code

Statistical progamming

[as-is]

a. People: statistical programmers / **[SAS ninjas]**

b. Actions: work with a mountain of SAS code

c. Deliverable: a **Clinical Study Report**

https://github.com/Bayer-Group/phuse-verisian

Challenges in 2024

Legacy Processes: 2024 or 1994? [as-is]

a. Data Storage

b. Data Transformation

c. Processes: the solo programmer & double programming

d. Silo everything: challenges unique to the Domain (data privacy): Data and code and tools are closed source

II. Improving Today

https://github.com/Bayer-Group/phuse-verisian

Learn from the best

The software dev / tech industry

[to-be]

a. Version control: code *and data*

b. **Agile development methods**: collaboration and iteration

c. Openish source: democratize the tools and processes that we use to make money with, not from. *SAS macros anyone?*

Impediments to change

a. Incumbents: legacy technologies and infrastructure

b. Mindset: lethargic and closed [source]

c. Generation change

d. Existant SAS code: **highly complex**

e. Motivation

“Change happens when the pain of staying the same is greater than the pain of change” -Tony Robbins

What can and should we do today?

Accept the **as-is**:

a. Work with what we've got

b. Commit what's need to change, i.e. time, money and people.

Objective judgement, now, at this very moment.
Unselfish action, now, at this very moment.
*Willing **acceptance** - now, at this very moment - of all external events. That's all you need.*

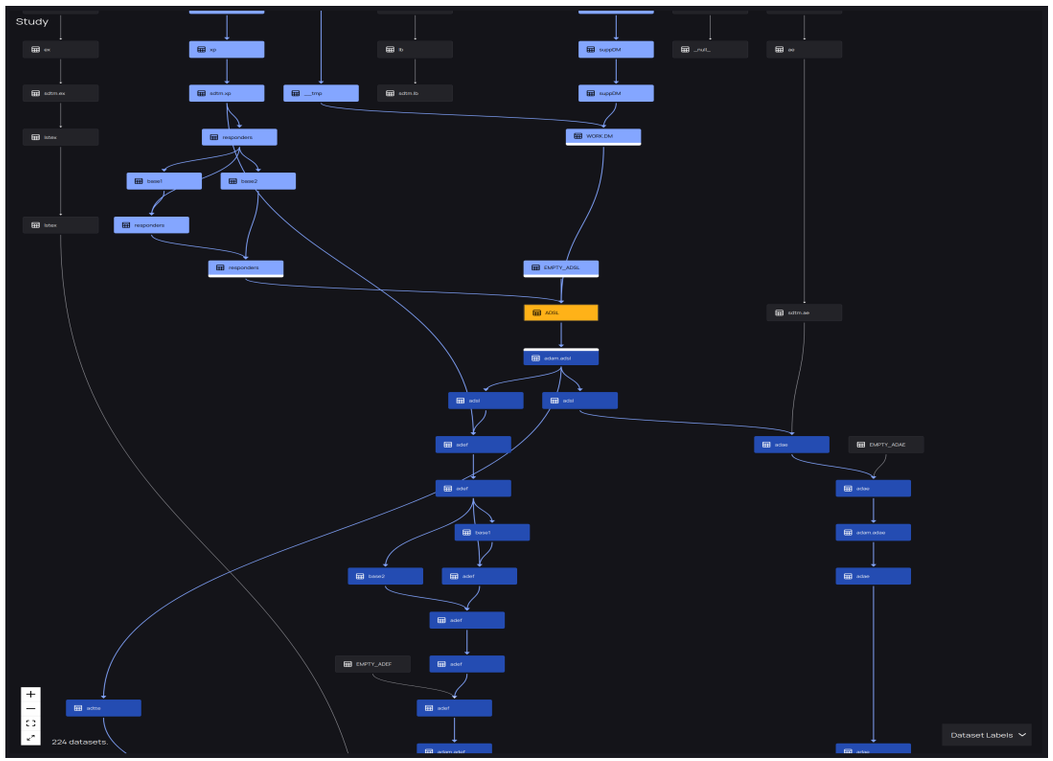
-Marcus Aurelius

III. Building the Future

https://github.com/Bayer-Group/phuse-verisian

Clinical trials as knowledge graphs

overlay a semantic data model on existing data and code

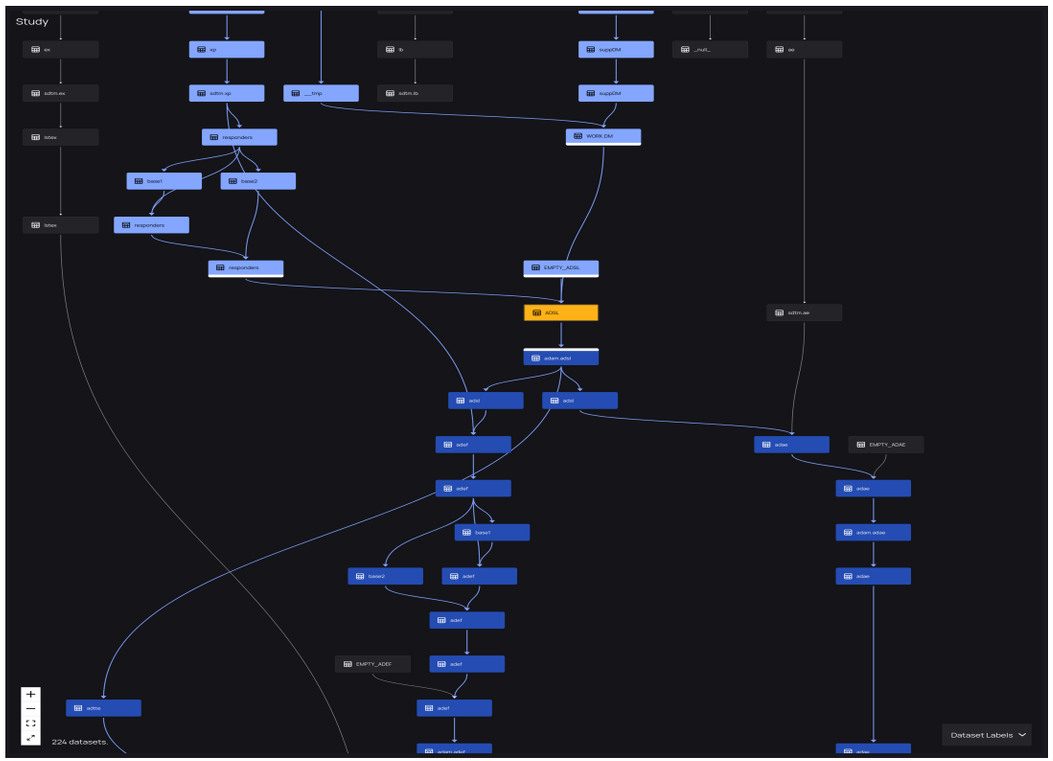


- Knowledge graphs add relationships or *edges* to data points or *nodes*
- This enables one to *traverse* the *topology* of the database

https://github.com/Bayer-Group/phuse-verisian

Clinical trials as knowledge graphs

overlay a semantic data model on existing data and code



i. end-to-end tracability/ full dependency map

ii. instant root-cause-analysis

iii. trivial data model mapping

iv. trivial study pooling

v. navigate studies and pools *in toto*

https://github.com/Bayer-Group/phuse-verisian

Clinical trials as knowledge graphs

overlay a semantic data model on existing data and code

Benefits previously demonstarted at PHUSE:

- PHUSE/Stardog – How Knowledge Graphs Will Transform the Pharmaceutical Industry
  Tim Williams, PHUSE & Guy Vorster, Stardog
- How will knowledge graphs improve clinical reporting workflows?
  Alexey Kuznetsov & Jorine Putter, GSK

No free lunch

Building an ontology and semantic model for clinical trials is highly specialized and **enormously labor intensive.**

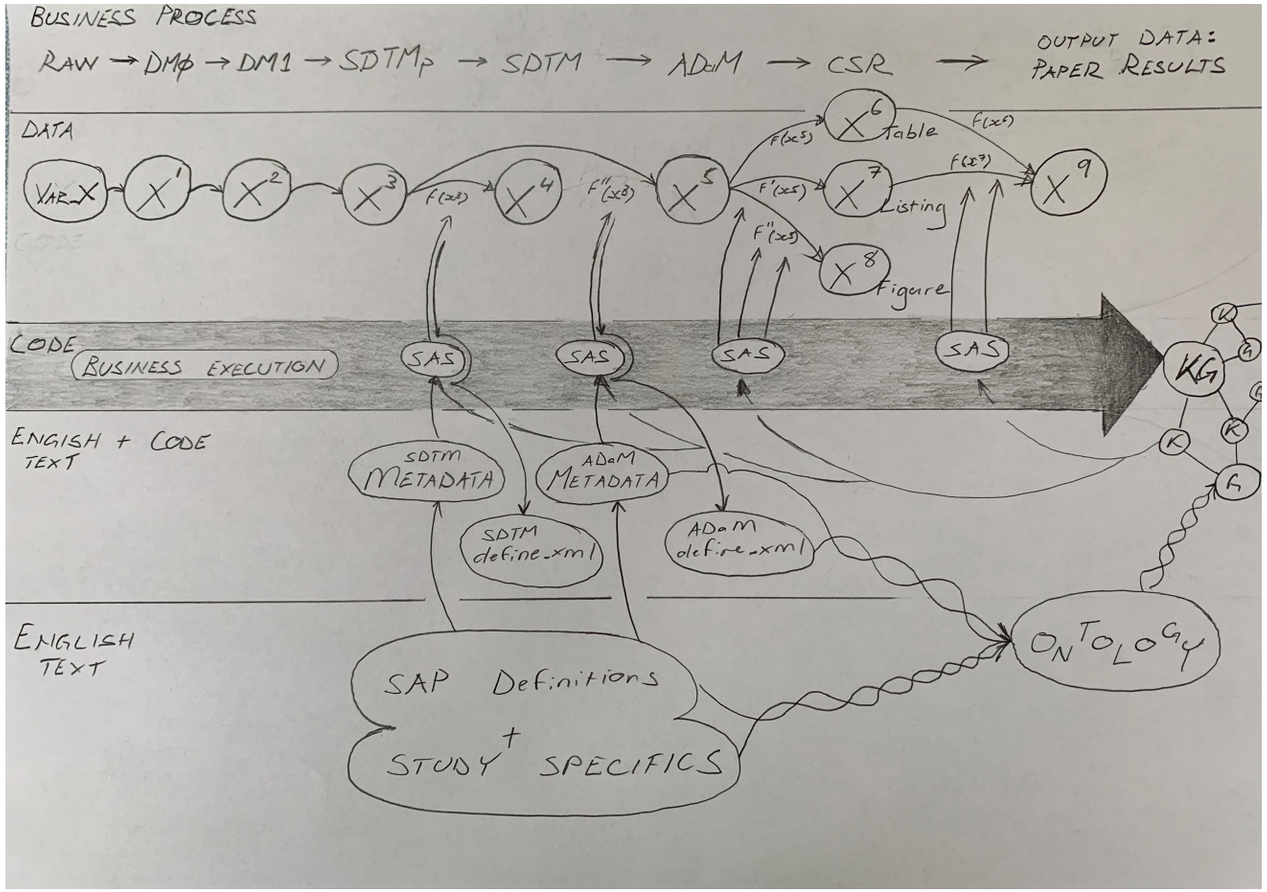
https://github.com/Bayer-Group/phuse-verisian

No free lunch

Building an ontology and semantic model for clinical trials is highly specialized and enormously labor intensive.

If we want to be faster we'll need a different approach:
Build it *bottom up*, from the code.

How bottom up?



- *top down* places human text at the top of the process, defining standard ontologies etc.
- *bottom up* places executed machine code at the top, of the process, form which standards are defined

https://github.com/Bayer-Group/phuse-verisian

Bottom up *Knowledge Graphs*: **under**pin a sematic data model on existing data and code

a. Not lost in translation: **code is *the* business execution** that delivers value (not the so much the other stuff)

b. Machines parse the code and deterministically extract all process logic

c. Works with *all* previous and current code
Change the *wheels* while the car is driving

Deterministically read machine parsed code

a. Deterministic vs probablistic

b. 100% explainable, no halucinations

c. Universal code logic extraction

d. Fully automate the graph creation **[upgrading while driving]**

e. Single source of truth feeds other processes

   i. Documentation creation: definitions and descriptions for **humans**

   ii. Translate to other languages: R, Python (SAS 👋)

And what of LLMs?

Can they replace stat programmers?

a. In short **no**, because they are probabilistic; determinism is at the heart of drug development especially for back translation

b. But they will be valuable for accelerating support tasks

Summary

- The challenges communicating clinical trial outcomes to authorities have exceeded our industries existing capabilities
- These challenges can be overcome by weaving clinical trials into knowledge graphs
- Automated knowledge-graph generation from code offers important benefits over top-down ontologically built graphs
- LLMs are finding many use-cases outside those requiring full explainability

Verisian Demo

https://github.com/Bayer-Group/phuse-verisian

Acknowledgements

Tomas Sabat Stöfsel & Henning Kuich **Verisian**
Oliver Wirtz, Cornelia Fulgenzi, Peter Bonata & Richardus Vonk **Bayer**
Paul Vervuren & Jules van der Zalm **Phuse**

Q&A

https://www.linkedin.com/in/karlbrand/

karl.brand@bayer.com