

A Pragmatic Approach to Data Cleaning – Decisions in Selecting When to Apply AI/ML



5–8 March
Renaissance Orlando
at SeaWorld Florida

US 2023

The Clinical Data
Science Conference



Abstract: When it comes to artificial intelligence (AI) and machine learning (ML), our industry is at the peak of interest, intent and expectations. But it's possible to prioritize overall AI/ML adoption instead of the problem where AI/ML will assist. It can be helpful to ask, "What are you trying to solve?" Data cleaning is a common first business case and an excellent candidate for AI/ML. However, because of the up-front work to train and validate the model, sometimes it's more efficient to take an intelligent approach to traditional, rules-based edit checks instead of one entirely AI/ML-driven. There is also a middle ground of a metadata-driven approach that is rules-based for greater efficiency. This session will discuss the continuum of decision-making for a pragmatic approach to data cleaning – from edit checks to AI/ML. Explore if/then scenarios, including considerations for reusability and lead times that help with selecting when to apply AI/ML.

Robert Musterer
VP Product Management
eClinical Solutions

- 
- Data sources
 - Complexity
 - Urgency



Or Is It?



Hype Cycle for Life Science Clinical Development, 2022



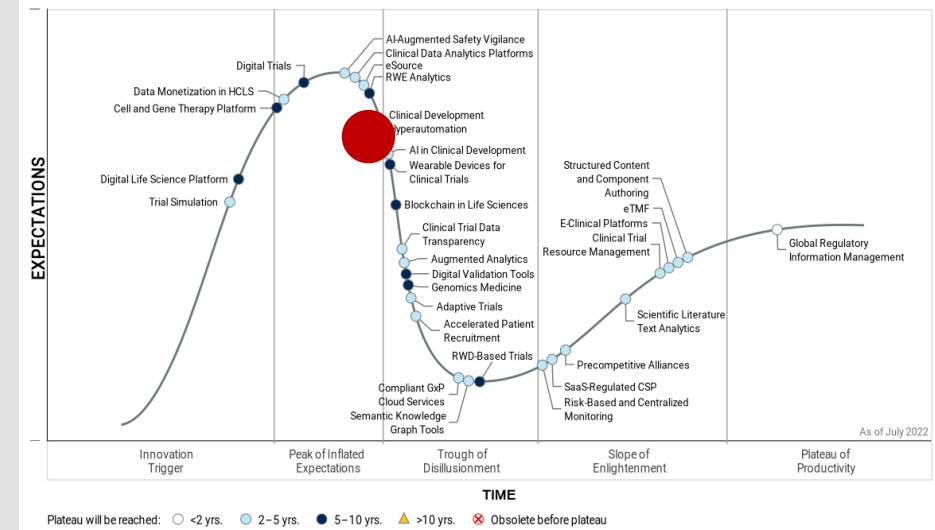
Plateau will be reached: ○ <2 yrs.

more plateau

Unrealistic Expectations:

- Just works **immediately** out of the box
 - Dismisses need for training data
 - Dismisses need for feature set definition
 - Dismisses need for ground truth data
 - Dismisses need to train models
- Assume all studies are equal
 - Ignores variance by therapeutic area
 - Ignores variability by drug class
 - Ignores variability by treatment arm (exposure duration and dose)
- Assumption of “Auto-Magic”
 - Ignores need for Human in the Loop
 - Ignores need to define use cases

Hype Cycle for Life Science Clinical Development, 2022



prag-'mat-ik\ also
law or business, fr. GK
fr. *prassein* to do — more at PR
USY (2) : OFFICIOUS b : OPINIONATED
practical affairs often to the exclusion
ters : practical as opposed to idealistic (the
the product of science and ~ effort — T. F.
have had no time or inclination to deal with
Clark) 3 : relating to or being in accordance
matism — pragmatic *n* — prag·mat·i·cal·ly \-i-
prag·mat·i·cism \prag-'mat-ə-,siz-əm\ *n* (1905)
trine of C. S. Peirce — prag·mat·i·cist \-səst\ *n*
prag·mat·ics \prag-'mat-iks\ *n* *pl* but *sing* or
branch of semiotic that deals with the relation
tic expressions and their users
pragmatic sanction *n* (1643) : a solemn decree
ter of primary importance and with the force of
prag·ma·tism \'prag-mə-,tiz-əm\ *n* (ca. 1864)
to problems and affairs (tried to strike a
and ~) 2 : an American movement in
Peirce and William James and marked by
of conceptions is to be sought in the
on of thought is to guide action
by the practical consequences
prag·ma·tic



What are we trying to achieve?

Data Cleaning

Assuring data is sufficiently ready for safety review and statistical analysis

In an efficient manner

In an Efficient Manner

Minimize:

Effort
Time to value
Expertise required

Maximize:

Value
Reusability
Scope of coverage

Hybrid approach

Keep the Simple
Things Simple



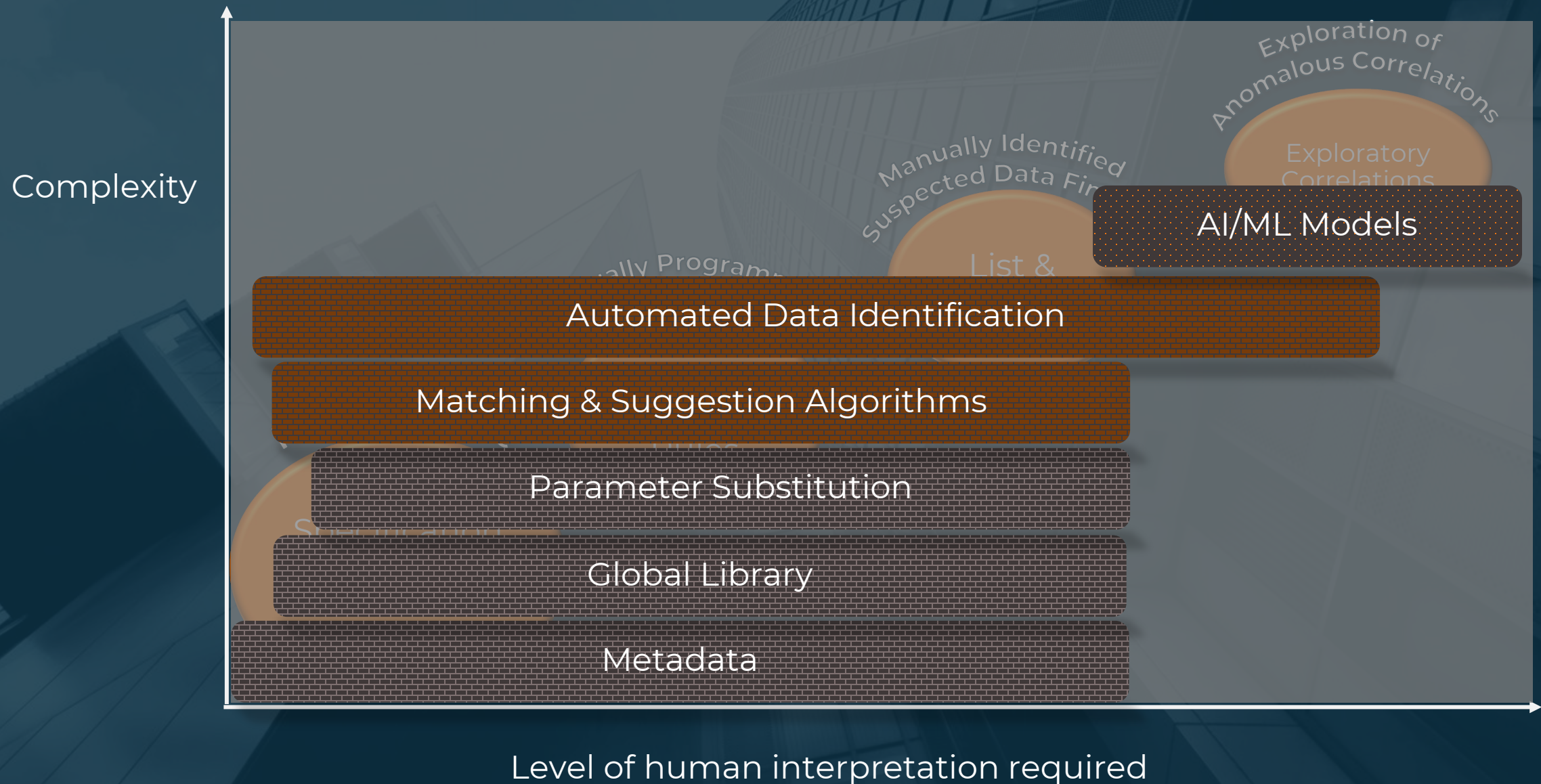
Hybrid Approach Building Blocks

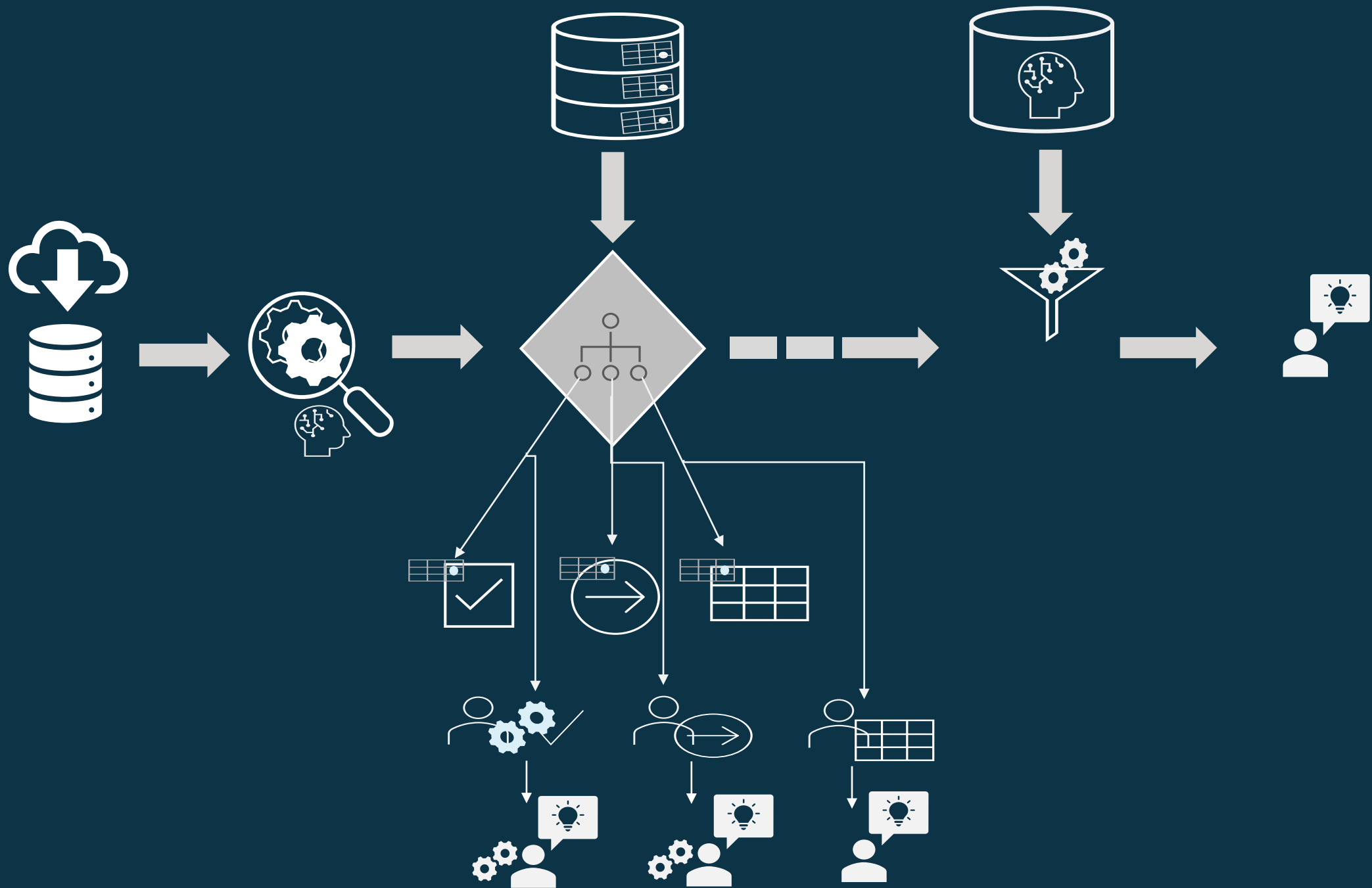
- Metadata
- Global library, with metadata
- Parameter substitution
- AI/ML assistance:
 - Data identification
 - Lookup matching and suggestion algorithms
 - Taming the volume funnel
 - Directing where to spend efforts
- AI/ML models
- Human in the Loop

In Practice



In Practice





Contact details

Robert Musterer
VP Product Management
eClinical Solutions
RMusterer@eclinicalsolutions.com