

A PHUSE EU Connect 2021 Presentation

DS09



Engineering an
MDR of the future



INTRODUCTIONS:



Ali Dootson

**Regional Director, Consulting
d-wise**



Mike Grove

**Founder & SVP of Engineering
Stardog**

What will you learn today?

1

What's wrong with MDRs today

2

Key to unlocking value from Data

3

How Data Fabrics will change MDR

Challenges around MDR

- Implementations failing, metadata still not used effectively
- Expectations of what an MDR can do exceed what they can actually do
- This probably sounds familiar:
 - Just a repository of data elements and their definitions
 - No automation or intelligence
 - No onward use

A1 ⬆ ✕ ✓ f_x Version											
	A	B	C	D	E	F	G	H	I	J	K
1	Version	Variable Order	Class	Dataset Name	Variable Name	Variable Label	Type	Controlled Terms, Codelist or Form	Role	Role Description	Description
2	SDTM v1.7	1	General Observations		STUDYID	Study Identifier	Char		Identifier	Identifier	Unique identifier for a study.
3	SDTM v1.7	2	General Observations		DOMAIN	Domain Abbreviation	Char		Identifier	Identifier	2-character abbreviation for the domain most relevant to the observation. The domain abbreviation is also used as a prefix for variables to ensure uniqueness when datasets are merged.
4	SDTM v1.7	3	General Observations		USUBJID	Unique Subject Identifier	Char		Identifier	Identifier	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.
5	SDTM v1.7	4	General Observations		APID	Associated Persons Identifier	Char		Identifier	Identifier	Identifier for a single associated person, a group of associated persons, or a pool of associated persons. If APID identifies a pool, POOLDEF records must exist for each associated person. (See Table 4.1.3.1, Pool Definition Dataset and Section 6, Applying Model Fundamentals to Associated Persons).
6	SDTM v1.7	5	General Observations		POOLID	Pool Identifier	Char		Identifier	Identifier	An identifier used to identify a result from a group of subjects that is not assignable to a specific subject.
7	SDTM v1.7	6	General Observations		SPDEVID	Sponsor Device Identifier	Char		Identifier	Identifier	Sponsor-defined identifier for a device.
8	SDTM v1.7	7	General Observations		NHOID	Non-Host Organism Identifier	Char		Identifier	Identifier	Sponsor-defined identifier for a non-host organism. This variable should be populated with an intuitive name based on the identity of the non-host organism as reported by the lab. Example: "A/California/7/2009 (H1N1)".

Five Terminology Tables



Image source: Tim Williams, Phuse, from

Mars Climate Orbiter

- Second mission under the Small Spacecraft Technology initiative
- Designed to study Mars' climate
 - Importantly, distribution of water
- \$330mm to design, build, launch
- 10 month flight time



Data + Context = Knowledge

- Everything is Data
 - No artificial distinctions between data & metadata
 - Sources, their contents, their schemas, all connected, all data
- Key to effective use is not simply *having* data
- It's understanding *what* that data means
- When data is self describing it ...
 - Makes it easier to find & access relevant data
 - Designed for interoperability; Improves scale of processing
 - Data dynamically aligned to common format easier to reuse



What does the ideal solution look like?

- Not just a metadata repository, not just storage. Not just Excel
- FAIR; Machine readable to power eventual AI & DSML
- Support traceability and lineage. Auditable
 - As soon as you get an answer, first question is why
- Really, want something to drive insights. Help us do better than we've done before
 - Inform trial design
 - IT system design
 - Make it easier to access, trace and audit data later on

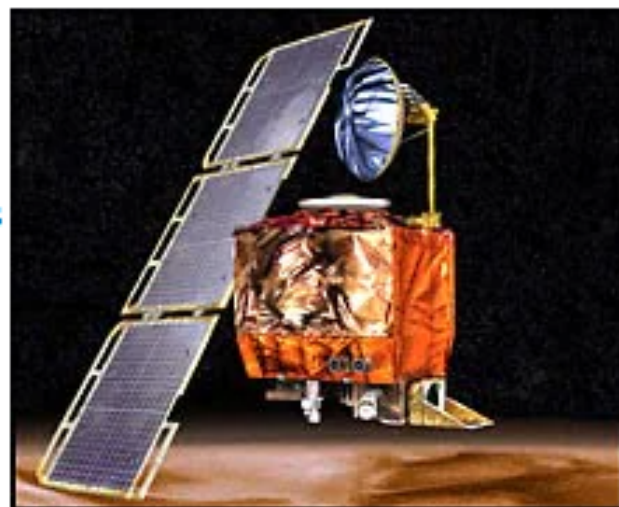
Mystery of Orbiter Crash Solved

By Kathy Sawyer

Washington Post Staff Writer

Friday, October 1, 1999; Page A1

NASA's Mars Climate Orbiter was lost in space last week because engineers failed to make a simple conversion from English units to metric, an embarrassing lapse that sent the \$125 million craft fatally close to the Martian surface, investigators said yesterday.



Scientists do not yet know what caused the Mars Orbiter to crash. (AP)

“

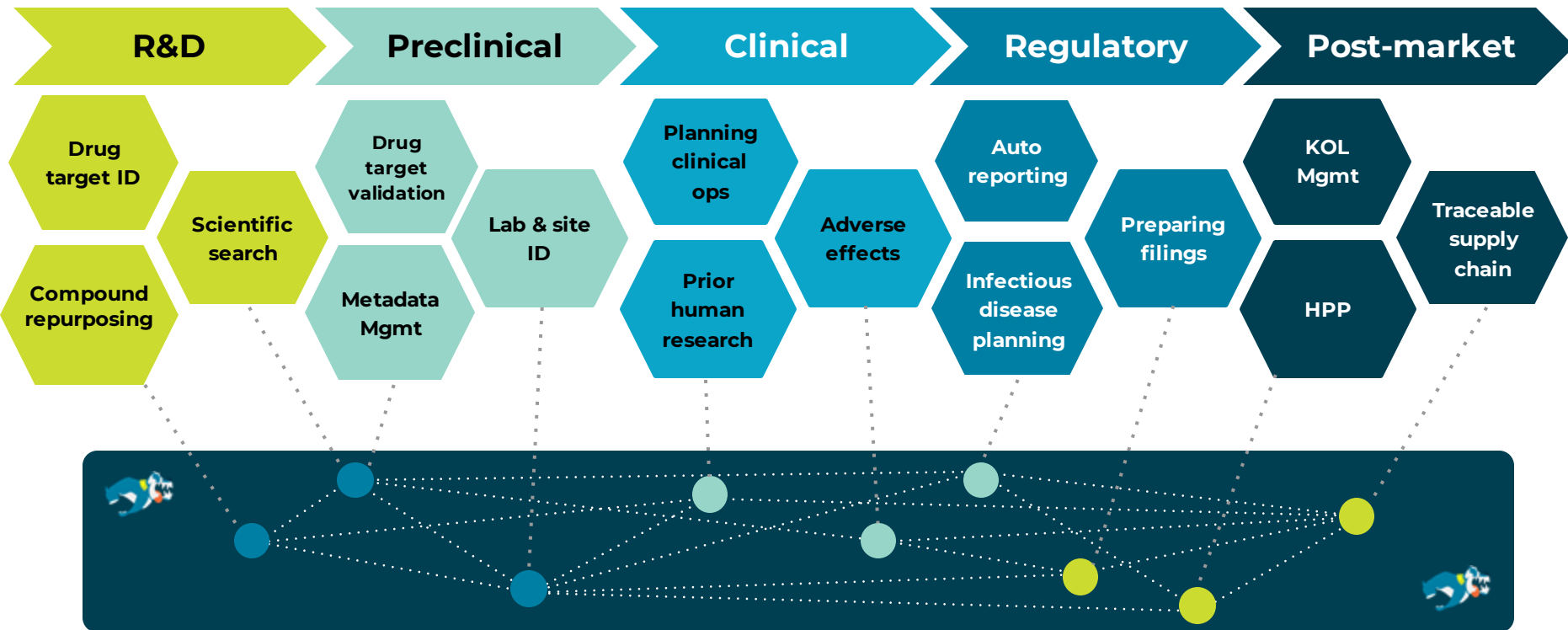
The Space Shuttle Program has a wealth of data tucked away in multiple databases without a convenient way to integrate and use the data for management, engineering, or safety decisions.

Columbia Accident Investigation Board Report

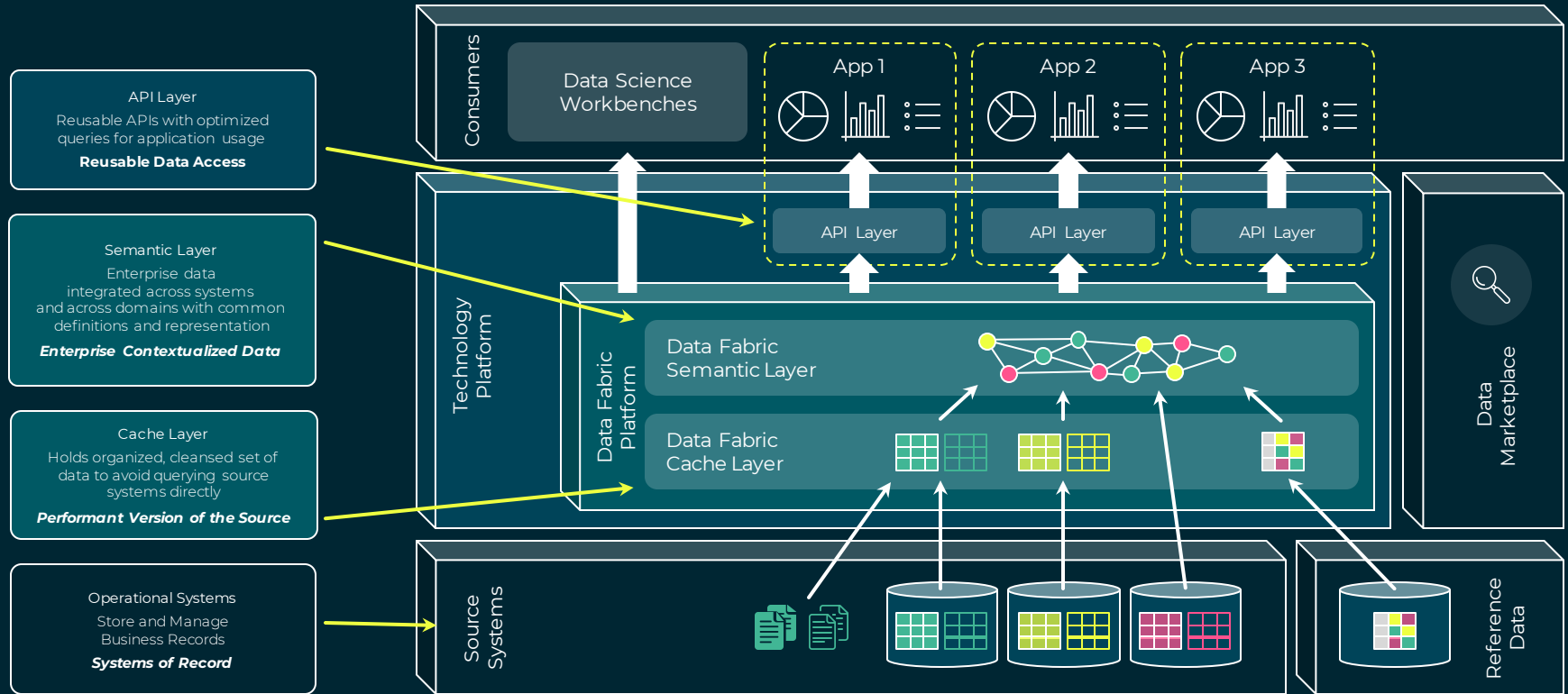
What does this mean for MDR?



Weaving a Data fabric for life sciences



Data Fabric Architecture



Some common misconceptions

- Data fabric is synonymous with digital transformation – “This is going to take forever”
- I need to replace all of my current applications
- I need personnel with special skills

Breaking down the Silos

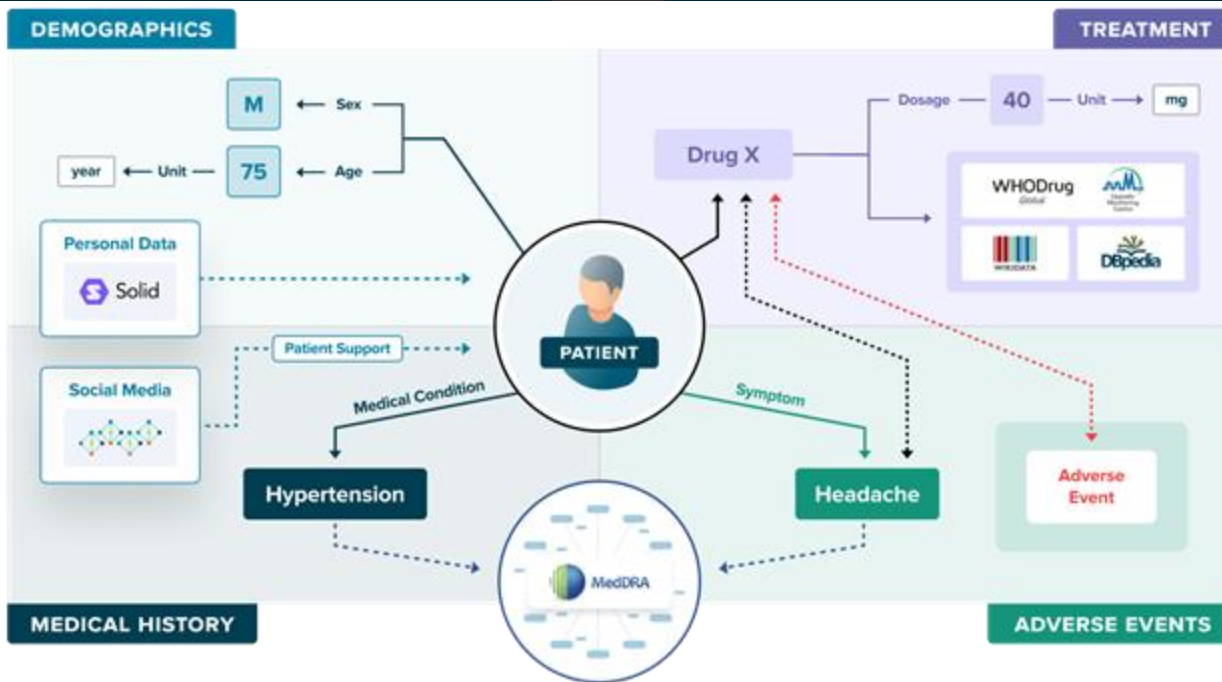


Image source: Tim Williams, Phuse, adapted from [webinar "How Knowledge Graphs will Transform the Pharmaceutical Industry"](#)

Summary

- MDRs have not found the success we'd hoped
 - Missed expectations, lots of Excel and hand work
 - Too much focus on only one piece of the puzzle
 - To meet their expectations, need to rethink the approach
- Data Fabrics cleanly align with major MDR objectives
 - Supporting a FAIR architecture
 - Native traceability of data
 - Easy integration with AI & DSML efforts
- Data Fabrics are the engine of the MDR of the Future

Thanks!



<https://www.linkedin.com/in/alistairdootson/>

@dootsy
@dwisetech



<https://www.linkedin.com/in/michaelhgrove/>

@mikegrovesoft
@StardogHQ