



# DISCOVERABLE DATA

*And the ontologies to facilitate it*



3DEXPERIENCE™



# THE COMPLEXITY OF SCIENTIFIC R&D



## Integration Data & Analytics Total Quality

- Disciplinary Silos
- Data Quality and Reliability
- Data Heterogeneity
- Data Volume and Complexity
- Complexity of Systems
- Data Privacy and Security
- Lack of Standardized Methods
- Data Analysis Techniques
- Resource Constraints
- Reproducibility
- Research Integrity and Misconduct
- Resource Constraints

- Cultural Barriers
- Collaboration and Communication
- Balancing Compliance with Innovation

## Globalized R&D

- Funding and Resource Constraints
- Intellectual Property Issues
- Time Zone Differences
- Infrastructure and Technology Differences
- Regulatory and Legal Differences
- Talent Acquisition and Retention
- Interdisciplinary Barriers
- Societal Resistance

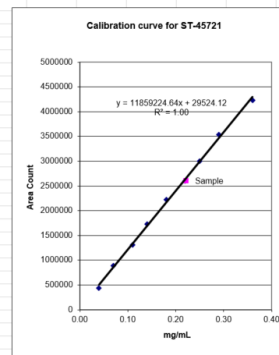
# LAB TESTING AND EXPERIMENTATION IS A BOTTLENECK



# PLACES DATA HIDES

- Spreadsheets
- Text Sections
- Batch Records
- Logbooks
- Sharepoint
- Fileshares

mmol/mL	ST-45721 mg/mL	Area Ct UV220
1.3288	0.36	4223364
1.0712	0.29	3539430
0.9235	0.25	3002637
0.8126	0.22	2603513
0.6649	0.18	2221230
0.5171	0.14	1731236
0.4063	0.11	1304217
0.2586	0.07	892778
0.1476	0.04	433625



Area: 2803513  
 Conc of HPLC sample: 0.2202439 mg/mL  
 Amount of sample in volumetric (mL): 0.0200  
 Amount of solvent in volumetric (mL): 10.0000  
 Dilution factor: 0.0020  
 Actual Concentration of material: 111.01 mg/mL  
 410.06 mmol/mL  
 volume (mL): 750.00  
 loss (g): 83.25914636  
 loss (mmol): 307.5471

**BATCH BAK COMPOUNDING FORMULA RECORD** 503384

GUANFENSIN EPICUTORANT 11,000 ml

QTY	INGREDIENTS	MANUFACTURER	LOT NO.	EXP. DATE	APPROVED	RECEIVED BY
1	Guafenesin 300 gm	Schering Plough	37676	May 88	5/10/95	CH
2	Simple Syrup, USP 3,400 ml	Abbott	GP 602	July 95	9/26/95	CH
3	Chloroform 45 ml	Alcon Chemical	900418	June 97	9/5/95	CH
4	Menthol 1.5ml	Mediatech Inc	0104447	Aug 95	1/5/96	CH
5	Mild Cherry Flavor 10.5ml	Team Products	J2-214	Sept 95	10/5/95	CH
6	Anarath, 1% Solution 50 ml	Conkey Corp	1373454	Sept 85	2/20/96	CH
7	Ethyl Alcohol, USP 750 ml	H. K. F. Inc	18A21	Feb 86	7/20/96	CH
8	Purified Water, used 11,000 ml	Final	37004	Mar 95	10/15/96	CH

**PREPARATION:**

1. Dissolve the guafenesin in 2,000 ml of hot purified water while the water is still on the hot plate.
2. In a separate container, dissolve the chloroform, menthol, and mild cherry flavor in the ethyl alcohol.
3. Add and mix both solutions together in the Alcop Mixer/Filter Unit and agitate thoroughly. Then add the simple syrup.
4. Add the anarath solution to the solution.
5. Add sufficient purified water to make the product measure 15,000 ml.
6. Filter the product using #51 filter pads.
7. Package and label the product.

**Character Data:** 11,000 ml  
**Volume:** 125-4 ml amber bottles  
**Container:** 500 ml amber bottles  
**Label:** 125-4 ml amber bottles  
**Lot:** 503384  
**Control #:** 503384  
**Prep by:** SPG  
**Checked by:** SPG  
**Released by:** SPG  
**Signature:** SPG  
**Date:** 10/15/96



# THE AI FALLACY

- AI is the same as Machine Learning
- AI can magically find the answer
- AI learns by itself
- AI is intelligent
- AI is infallible

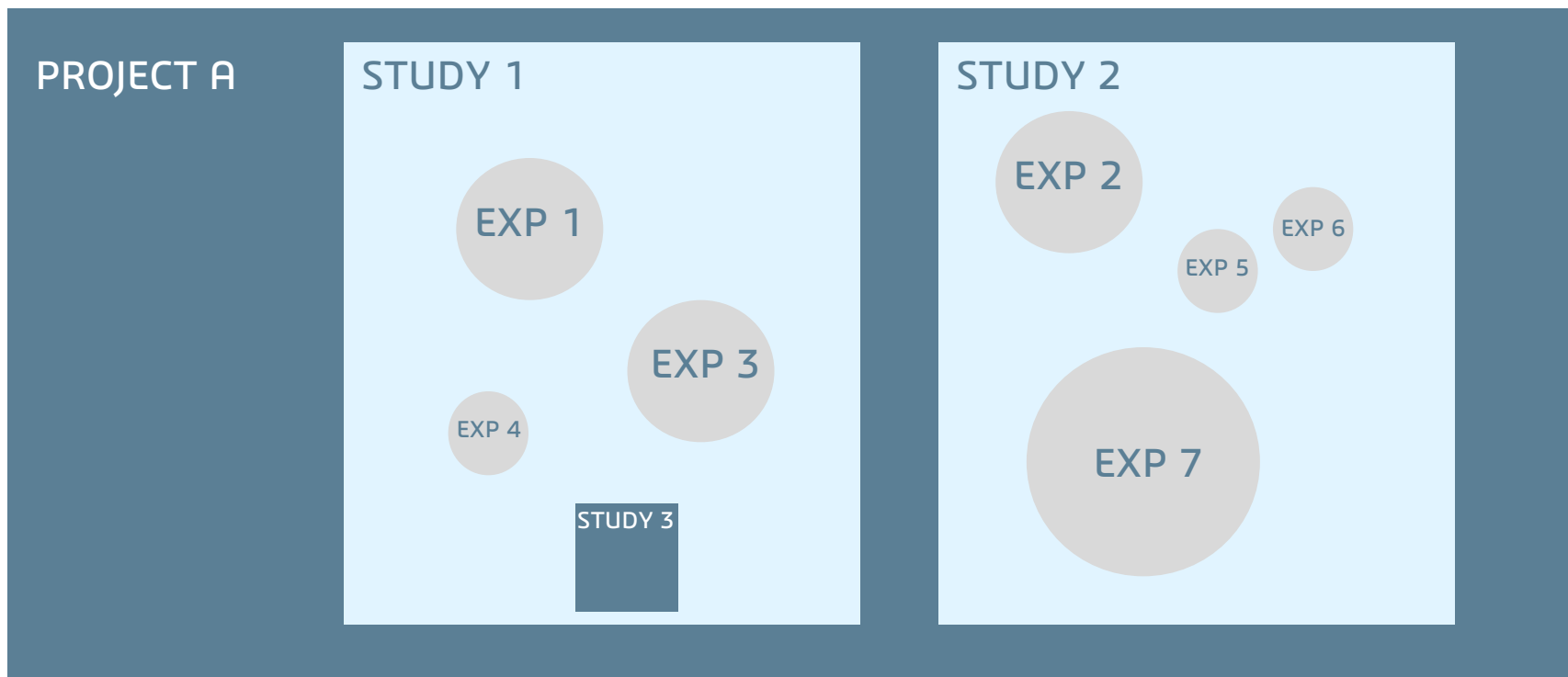


# GARBAGE IN, GARBAGE OUT

AI is only as good as its source data



# DATA HIERARCHY FOR CONTEXT



# USABILITY VS STRUCTURE

Flexible  
Intuitive  
Not Clicky  
Attractive  
Fast

*but....*

*"data" is hard to find  
and use*



**VS.**



Rigid  
Complicated  
Information Dense  
Lots of  
Configuration  
Slow

*but....*

*Data can be analyzed  
and reused*



# ...BIOVIA OUT-OF-THE-BOX ONTOLOGIES

## Measurements



Data from an observation, instrument or calculation

140

## Activities



A category of measurement from characterization

52

## Unit Operations



A type of work in a laboratory that involves materials

43

## Equipment Class



Data structure to integrate a laboratory instrument

<150

# ...COMBINED WITH INDUSTRY STANDARDS



...and others

# PARAMETERS AND QUALITY ATTRIBUTES

*The quality attribute of a parameter defines comparability of results across activities*

## Water Content

Parameter	Parameter Type	Scientific Data Type	Quality Attribute
sample	input	sample	sample
method name	task	string	assay.watercontent.method
water content	output	mass concentration	assay.watercontent.watercontent

## Color

Parameter	Parameter Type	Scientific Data Type	Quality Attribute
sample	input	sample	sample
color	output	color	appearance.color
preparation	input	string	appearance.color.preparation
method name	task	string	appearance.color.method
reference	output	string	appearance.color.referencestandard
assessment	output	pass/fail	appearance.color.assessment

## Visible Particles

Parameter	Parameter Type	Scientific Data Type	Quality Attribute
sample	input	sample	sample
method name	task	string	appearance.visibleparticles.method
sample size	input	volume	appearance.visibleparticles.sampleamount
particle count	output	count	appearance.visibleparticles.count
assessment	output	pass/fail	appearance.visibleparticles.assessment

# GENERIC V SPECIFIC

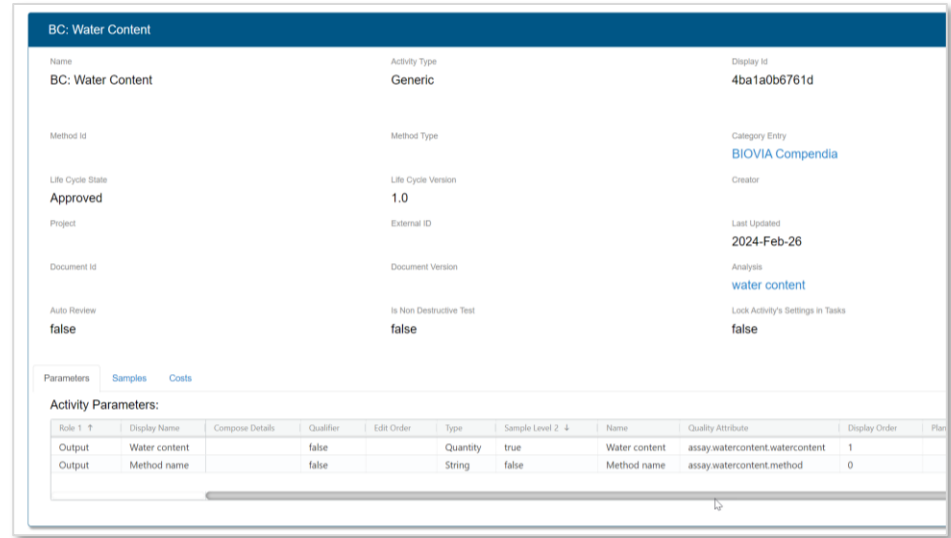
*Generics Analyses are intended to relate specific techniques together*

## Example | Water Content

Possible specific analyses (with or without equipment integration)

- Karl Fisher
- Loss on Drying
- NIR
- Electrical
- Microwave

All methods produce the same final Result.  
Specific methods may produce additional results.



Name	Activity Type	Display ID
BC: Water Content	Generic	4ba1a0b6761d
Method ID	Method Type	Category Entry
		BIOVIA Compendia
Life Cycle State	Life Cycle Version	Creator
Approved	1.0	
Project	External ID	Last Updated
		2024-Feb-26
Document ID	Document Version	Analysis
		water content
Auto Review	Is Non Destructive Test	Lock Activity's Settings in Tasks
false	false	false

Role 1 ↑	Display Name	Compose Details	Qualifier	Edit Order	Type	Sample Level 2 ↓ 4	Name	Quality Attribute	Display Order	Plan
Output	Water content		false		Quantity	true	Water content	assay.watercontent.watercontent	1	
Output	Method name		false		String	false	Method name	assay.watercontent.method	0	

# THE VISION

## Virtual Twins

Homogeneity Model

Cost Model

Texture Model

Hardness Model



Particle Size Model

Mathematical solvers that experts use to advise the business

# VIRTUAL TWINS

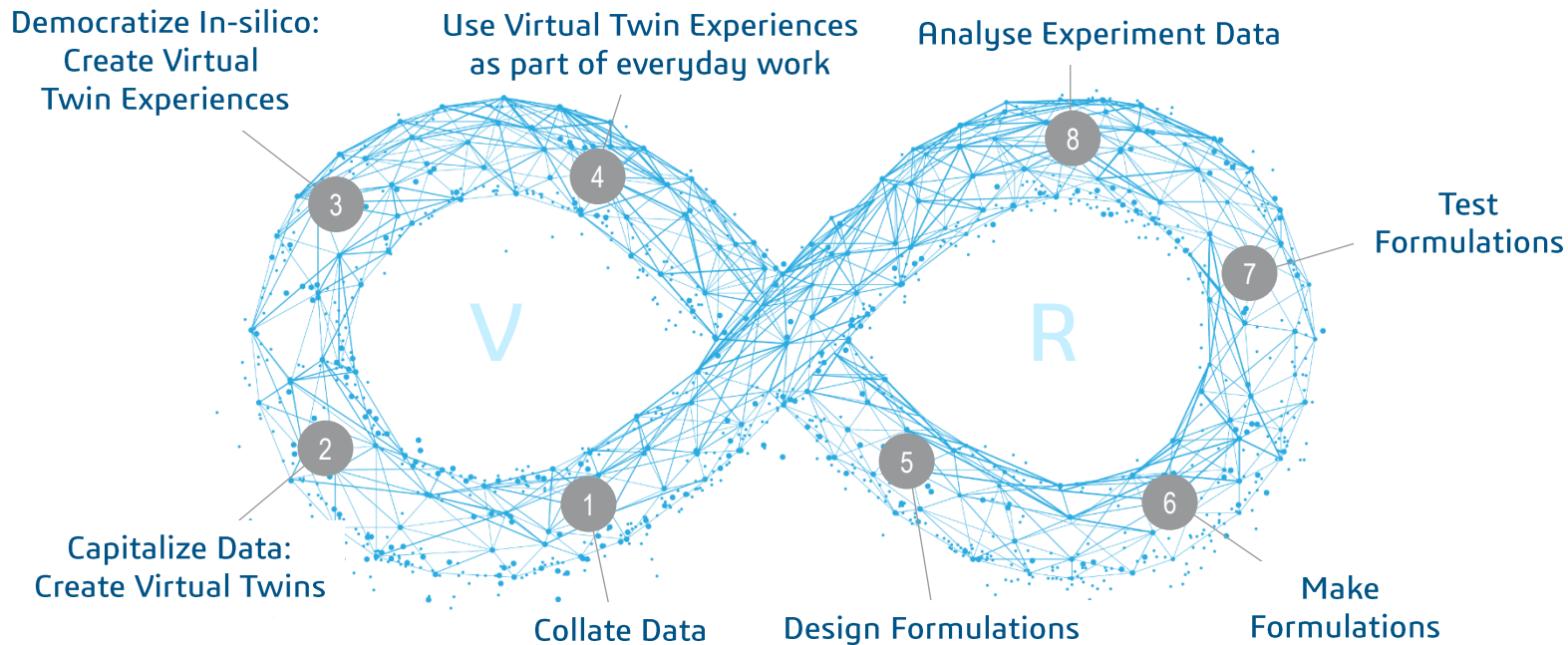


User experiences anyone can use to solve business problems

# THE DATA IS THE MODEL IS THE VIRTUAL TWIN

- ▶ **DATA** This is the foundation of a virtual twin. It's the real-time information gathered from the entity or system. This can include sensor readings, performance metrics, and other relevant data points & measurements.
- ▶ **MODEL** A model is a representation of the entity, process or system. It's created based on the data collected and incorporates mathematical equations, algorithms, and simulations to predict behavior, identify potential issues, and optimize performance.
- ▶ **VIRTUAL TWIN** The virtual twin is the combination of the data and the model. It's a virtual replica of the entity, capable of mirroring its behavior, history, and future states.

# VIRTUAL + REAL

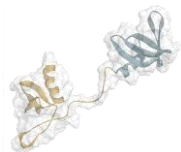




# TRACEABILITY THROUGHOUT PRODUCT LIFECYCLE

Combining real world data and evidence with models to assess outcomes

## Product Virtual Twin

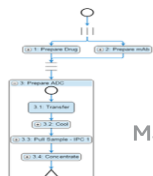


Drug Product / Substance



Device & Packaging

## Process Virtual Twin



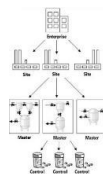
Materials



Equipment Parameters



S88 / S95 Recipe



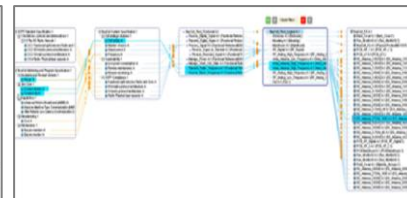
Specifications & Test Methods



CAD/PDM

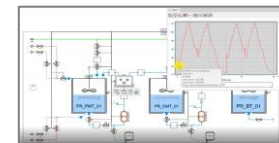


Issues & Change

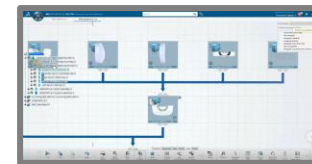


Requirements

## Plant Virtual Twin



Engineering Definition



Manufacturing Planning

# END-TO-END QUALITY FOR BIOPHARMA

