

# The Next Wave of CMC Digitalization: Challenges and Opportunities

---

Dave Dorsett | Chief Digital Officer

April 4, 2025

# The Next Wave of CMC Digitalization: Challenges and Opportunities

Bio-IT (Boston, USA)

Mohit Agnihotri  
April 2025

Version 1

# CMC Digital Transformation @ Abbvie



## VISION

*Transform Dev Sci into a **digitally integrated organization** that leverages **data as an asset***



## WHY

We enable **better scientific decision-making** & realize **efficiency gains**



## HOW

By transforming the way we work, using **data & digital technologies**



## WHAT

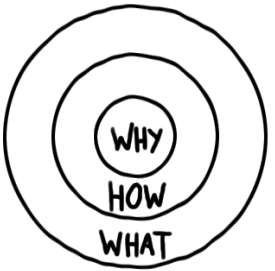
- ✓ Capture structured data
- ✓ Make data centrally available and reusable
- ✓ Minimize manual data transcription and verification
- ✓ Automate scientific report writing
- ✓ Create and leverage predictive models

abbvie



## INSTEAD OF

- ... Data managed in unstructured tools like SharePoint, Excel, OneNote
- ... Data stored in silos/personal drives, limited to the data producers
- ... Information transcribed over & over to multiple systems
- ... Manual creation of reports
- ... Generating datasets for every hypothesis

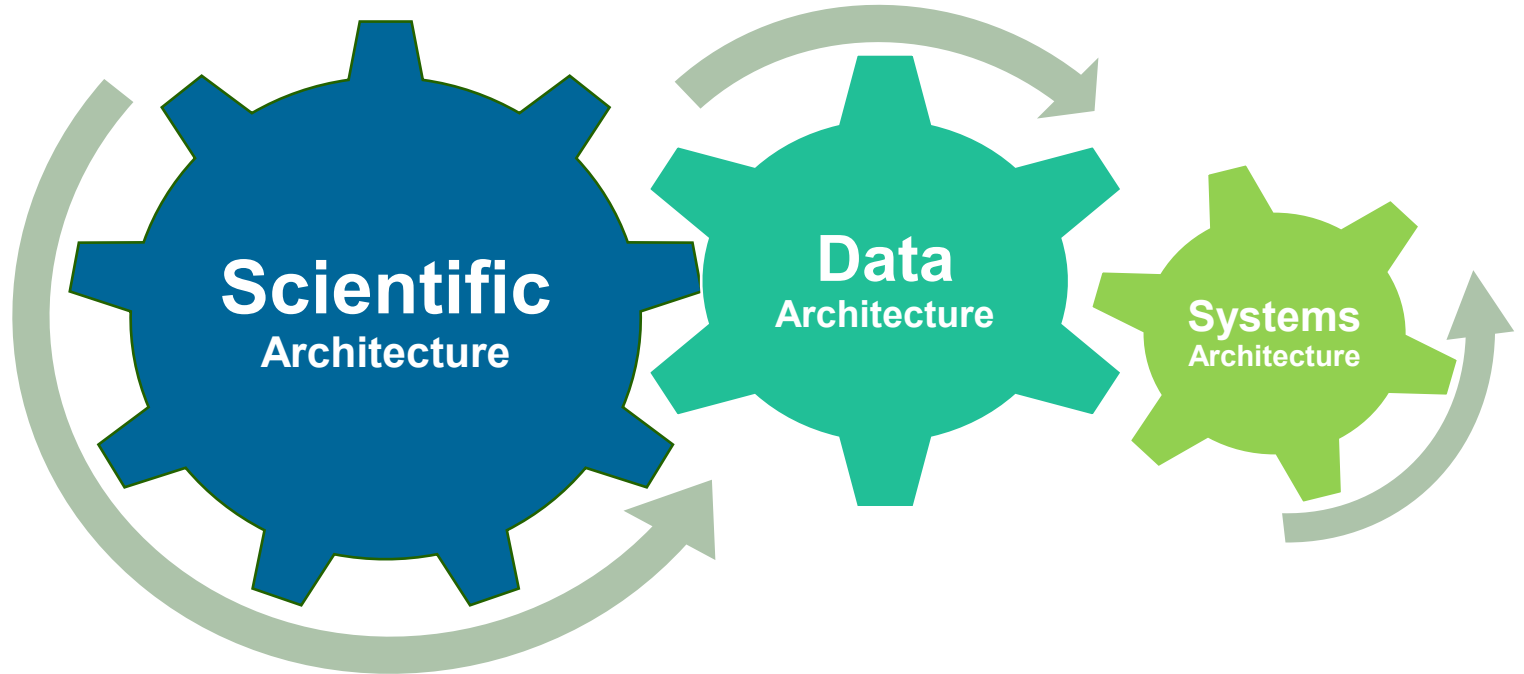


Simon Sinek, 2009

# Two Key Differentiating Principles:

## Three Gears Approach:

1. A deep understanding of the current state of a specific challenge or **Use Case**
2. An assessment of the data requirements and a future state **combining updates to data & business process**
3. Evaluation of multiple systems, each used for their intended purpose, to create a **comprehensive solution**



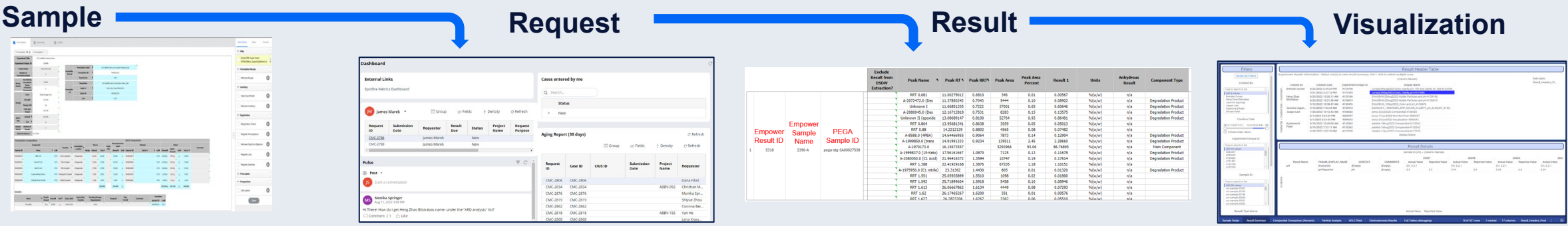
**Where** does a sample come from and **why**?

## End-To-End Solution Analysis

*Never about a single group. Always look at the interfaces!*

**How** is the data used to make a decision and by **whom**?

# Capabilities in CMC Development, 2021-2025



CMC Request

Formulation Composition, Process Parameters

Analytical Results

Results Summary Optimization

Clinical In-Use Studies

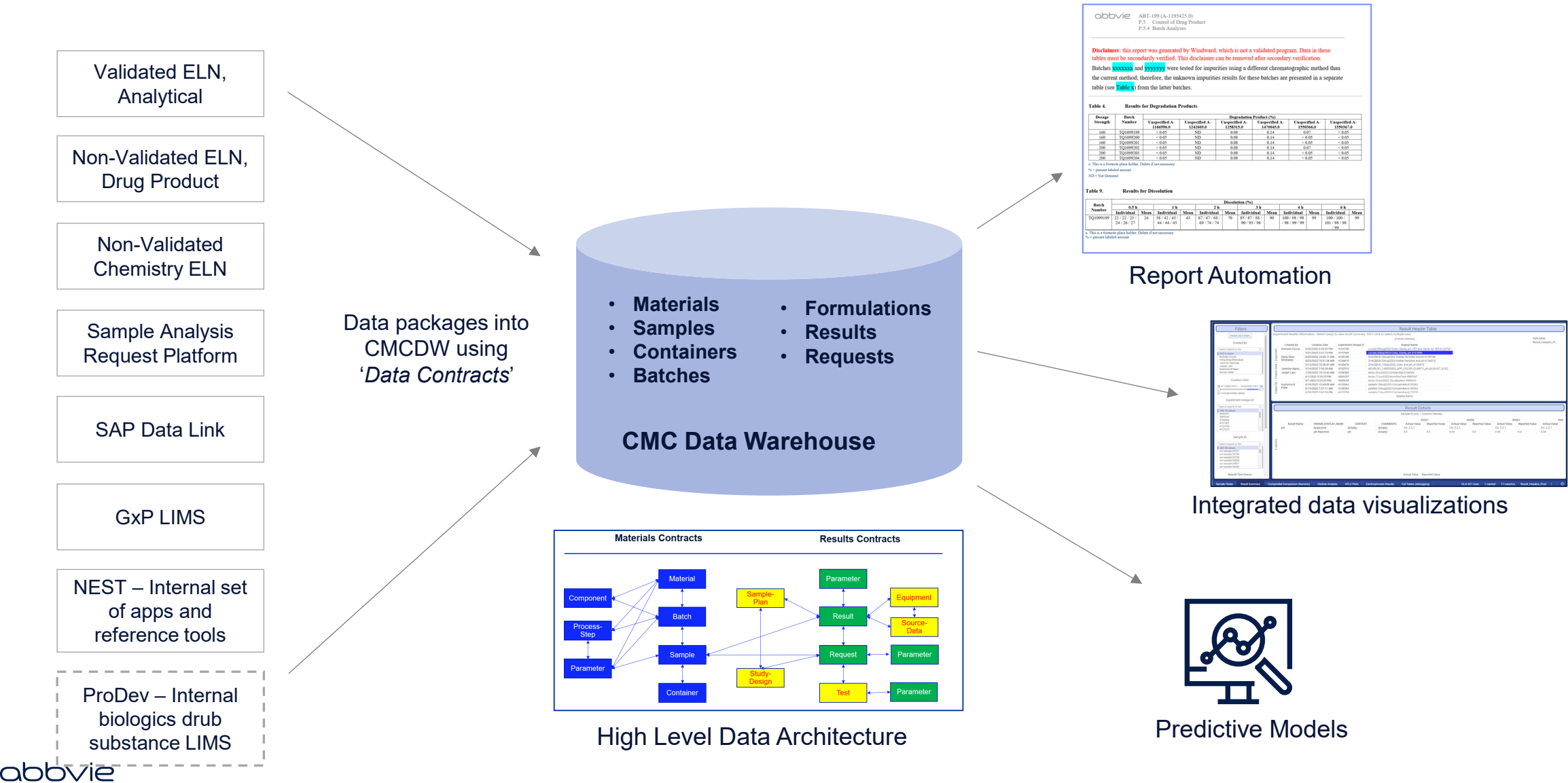
Biologics Developability

CMC Stability Design

## Technical platforms in use to support the capabilities:

| Experiment Execution                                                                                                                                                                        | Workflow Orchestration                                                                                                                                                                         | Data Aggregation                                                           | Visualization & Reporting                                                                                                   | Primary CDS                                                                                                                                | Reference Systems                                                                                                   |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"><li>IDBS ELN<ul style="list-style-type: none"><li>Non-validated</li><li>Validated</li></ul></li><li>Dotmatics ELN - Chem</li><li>Sample Manager</li></ul> | <ul style="list-style-type: none"><li>Pega<ul style="list-style-type: none"><li>CMC Request</li><li>Study Design</li><li>Stability Request</li></ul></li><li>Library of Data Parsers</li></ul> | <ul style="list-style-type: none"><li>CMCDW (CMC Data Warehouse)</li></ul> | <ul style="list-style-type: none"><li>SpotFire</li><li>Shiny - RStudio</li><li>Windward</li><li>ACD/Labs Luminata</li></ul> | <ul style="list-style-type: none"><li>Empower CDS<ul style="list-style-type: none"><li>Non-validated</li><li>Validated</li></ul></li></ul> | <ul style="list-style-type: none"><li>CENTree</li><li>EARS (NEST)</li><li>MALTA (NEST)</li><li>CPR (NEST)</li></ul> |

# CMC Data Warehouse: Opportunity to integrate CMC data in one platform to drive data automation



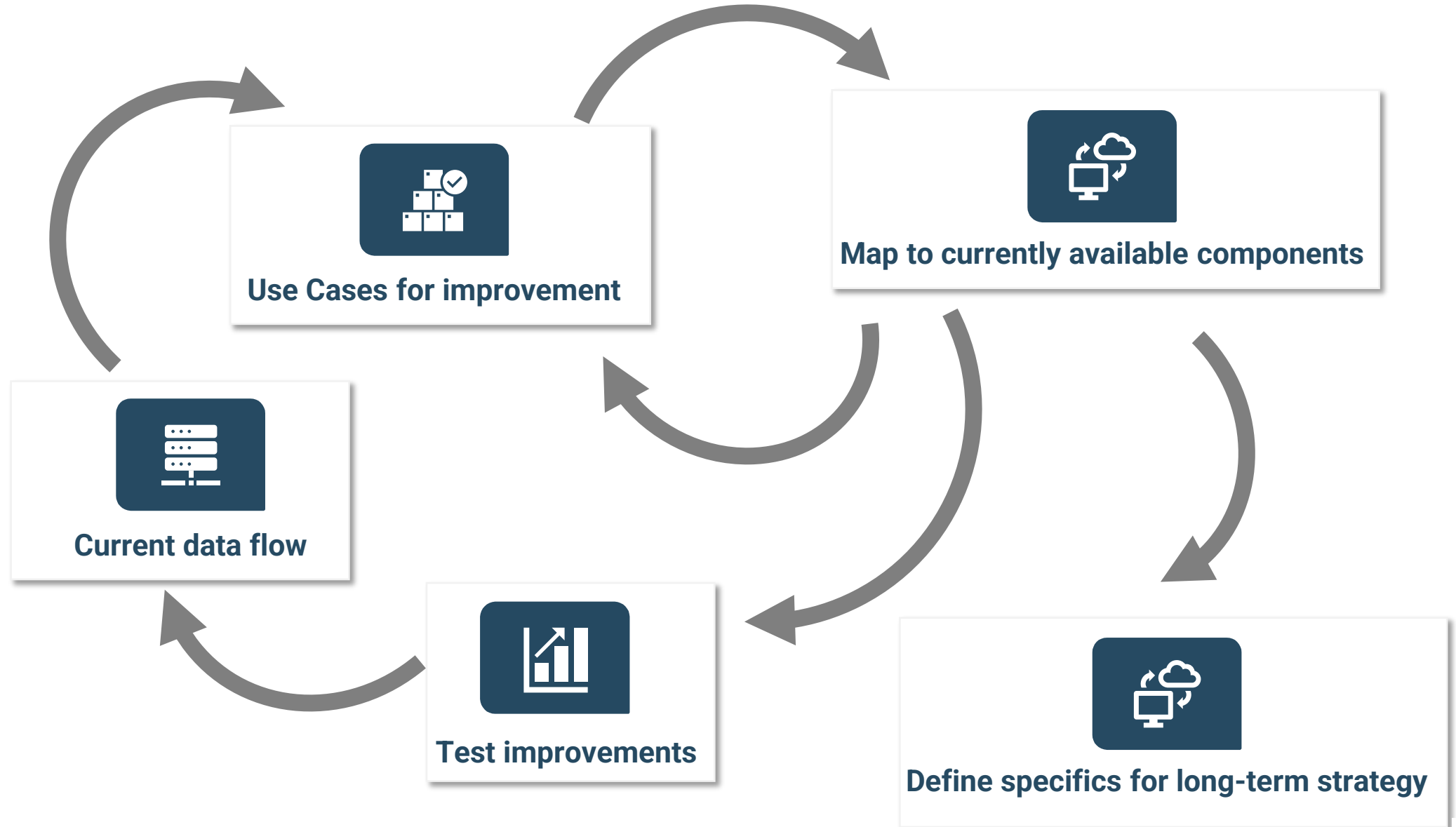
# Brief Background

- **“Digitalization”** is about changing practices to data-first methods
- **Changing practices** = transforming how work is done
  - It’s a people thing!
  - Technology is only a tool.
- **This is not new- excerpts from my personal history**
  - 1980s: chemical structures from paper into databases.
  - 1990s: *in-vitro* high throughput screening, liquid handling robotics, library design.
  - 2000s: robotic catalysis r&d and physical-mechanical property testing and modeling.
  - 2010s: paperless veterinary sciences and study design.

# Summary of First Round of Digitalization

- **Putting stakes in the ground based on prioritized R&D outcome generally worked**
  - Iteration required was challenging for IT/system owners
  - Challenging sometimes to maintain end-to-end focus (dropping into the rabbit holes)
- **Got pretty good at being good enough (avoiding the enemy of perfect or finding the “best”)**
  - Challenging sometimes to put enough attention on the important unhappy paths
- **(Re-)confirmed that this is not a “buy and deploy it” project**
- **Did not invest sufficiently in “run”**
- **Found a mission to focus on adoption**

# A Little More about Iteration

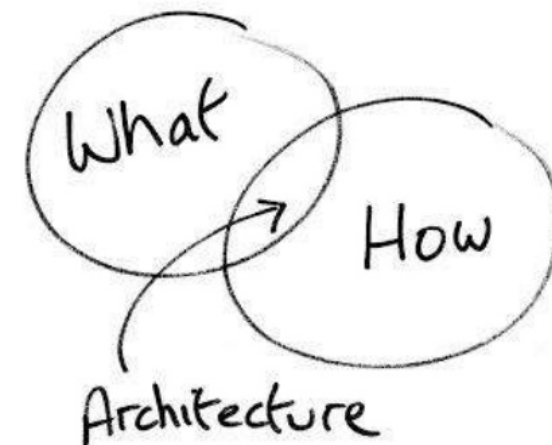


# Notable First Round Successes

## Data Products Architecture

- Getting good at capture and harmonization in a broad sense
- Able overall to balance normalization of data with localized workflow requirements
- Approach to the hierarchy of data from source to aggregate
- Succeeded in system independent definitions and “snap-on” triumphs

**Found a sweet spot of benefits with Study Orchestration**



# Some Next Round Challenges

- **Process Data**

- Started/attempted during the first round (batch genealogy, Process capture)
- Very rich opportunities

- **Extending the hierarchy of data**

- Further summarization with annotation
- Practical handling of provenance relationships

- **Demand management**

- Road-mapping for such a large initiative has been challenging
- Speed and flexibility of data governance to keep up with demand



## Future: Assembly and Review of Submission Documents

> **The initial set of data resources are in place**

> **Future work**

- Authoring and review processes (usability!)
- Incrementally extending the data hierarchy

# Future: Preparing for HL7 FHIR

- **First round of data resources aligned with FHIR Phase 1, Stage 1 and 2**
  - Drug Product, Control of Materials, Specification
  - Substance and Product Characterization, Batch Formula
- **Current work**
  - Re-visiting the test/attribute ontology based on 4 years of data
  - Iterating on Stability
- **Future work**
  - Manufacturing Process



# Future: Developing Models from CMC Data

- **Prerequisite of clean normalized data and a logical data strategy**
  - Harmonized data sets are reaching interesting sizes
  - Have tested looping handling model output into the experimental data architecture
- **Next step is to develop usable (non-technical) “self-service” data access**
  - Targeting through specific workflows problems (e.g. Developability)
- **Model governance (e.g. Model naming, versioning) will need attention**

# Vignette 1: Biologics Developability

## Empowered decision-making for selection of viable candidates

| Attributes                   |                          | Experimental results values*                                                       |                                                                             |                                                                  |                                                                  | Predicted re values AI/ML model |
|------------------------------|--------------------------|------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|------------------------------------------------------------------|------------------------------------------------------------------|---------------------------------|
|                              |                          | mAb 1                                                                              | mAb 2                                                                       | mAb 3                                                            | Peer 1                                                           | Modeled mAb                     |
| Thermal Stability            | DSF                      | <div></div>                                                                        | <div></div>                                                                 | <div></div>                                                      | <div></div>                                                      | <div></div>                     |
|                              | DSC                      | <div></div>                                                                        | <div></div>                                                                 | <div></div>                                                      | <div></div>                                                      | <div></div>                     |
| Chemical Stability           | Forced Degradation LC-MS | <div></div>                                                                        | <div></div>                                                                 | <div></div>                                                      | <div></div>                                                      | <div></div>                     |
| Viscosity                    | iBeacon                  | <div></div>                                                                        | <div></div>                                                                 | <div></div>                                                      | <div></div>                                                      | <div></div>                     |
|                              | Viscosity                | <div></div>                                                                        | <div></div>                                                                 | <div></div>                                                      | <div></div>                                                      | <div></div>                     |
| Aggregation                  | SEC                      | <div></div>                                                                        | <div></div>                                                                 | <div></div>                                                      | <div></div>                                                      | <div></div>                     |
| Serum Stability              | SEC                      | <div></div>                                                                        | <div></div>                                                                 | <div></div>                                                      | <div></div>                                                      | <div></div>                     |
| Sequence Alignment           |                          | ASGYTFTSYNMHWVRQAPCGCLEI                                                           |                                                                             | ASGYTFTSYNMHWVRQAPCGCLEWVGAIYPCEGDTSY                            |                                                                  |                                 |
| CUMULATIVE RISK (rule based) |                          | <div></div>                                                                        | <div></div>                                                                 | <div></div>                                                      | <div></div>                                                      | <div></div>                     |
| COMMENTS (AI generated)      |                          | ~5% HMW species increase per day in mouse and rat plasma. Has unfolding below 50C. | Sequence liabilities not manifesting. However, high viscosity from 50 mg/mL | Liabilities aren't surface exposed and not manifesting. Too good | Liabilities aren't surface exposed and not manifesting. Too good | Aggregation pCQA                |

Business Impact:

1. Empowered decision-making for selection of viable candidates via:

- Native-like developability comparative assessment
- Evaluation of molecular drug optimization cost on drug-like-properties (DLP)

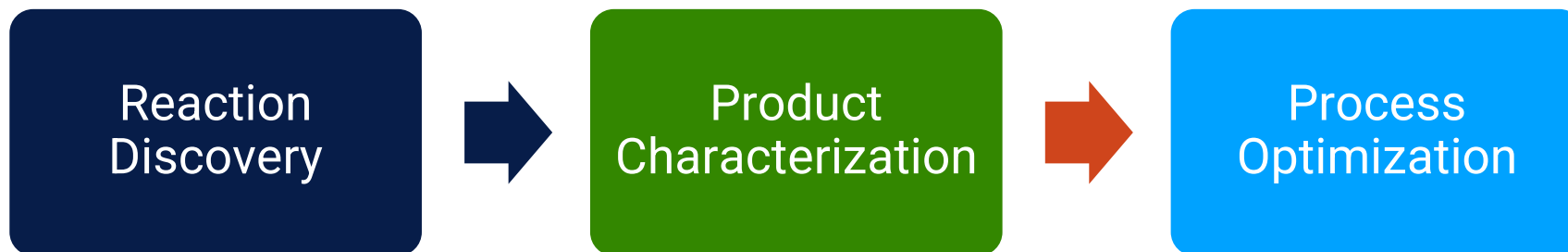
2. Downstream platforms fitness scores *via* comparison to analogs

3. Molecular diversity *via* Sequence alignment projection on DLP landscape

4. Success rate *via* comparison to relevant peer products

5. Accelerate R&D *via* AI/ML by leveraging experimental data to project DLP attributes and guide DLP experimental design

## Vignette 2: AI/ML in Process Development



- Accelerate the pipeline by significantly **reducing the amount of material and time** required for process development across all modalities
- **Reduce cost** and **environmental impact** of reaction optimization campaigns by significantly lowering the number of 'wet' experiments than traditionally required
- Deliver **novel solutions** for the synthesis and characterization of APIs of ever-increasing complexity

## What is needed to achieve these outcomes?

## Alignment

- Problem statement
- Data required
- Terminology

## Consistent Utilization of Tooling

- Adapt business processes to technology
- Everyone is accountable!

## Continuous Improvement

- Step 1: Train and adopt workflows to new tech
- Step 2: Provide constructive feedback and engage to iterate!
- Step 3: Ensure the resource model can flex and adapt to a changing landscape

**An understanding by everyone that the problems we tackle are ALWAYS larger than they initially appear**

[illegible]

## Report Automation



## Predictive Models

[illegible]

## Integrated data visualizations



# Acknowledgements

## DevSci Data Strategy

- Christian Pilger
- Dana Filoti
- Florie Le Priault
- Tod Davies
- Christian Metz
- Lance Heinle
- Mathew Mulhern
- Gerald Green
- Drashti Vasant

## Information Research

- Chris Butler
- Ashley Elliott
- Jared Epstein
- Alex Imdahl
- Kelsey Johnson
- Amy Lambrecht
- Xiaofeng Li
- Katherine Lopez
- Enrique Martinez
- Abhishek Pandey
- Swetal Parikh
- Julie Ring-Smith
- Jim Williams

## DevSci Use Case Leads & Subject Matter Experts

- Alex Ruggles
- Leon van Haandel
- Axel Wilbertz
- Corianne Randstrom
- Yemin Liu
- Julie Heflin
- Shashank Shekhar
- ...and dozens more!

## Partners

- Astrix
- IDBS
- Dotmatics
- ACD/Labs
- Zifo
- Pega Systems