

The 12th IABS Statistics Workshop

# The Statistician's Role in Shaping CMC Data Strategy and Driving Scalable Analytics

Ruojia Li, Dan (Cassie) Liu | October 20, 2026

# The Common Data Challenges Limit Value Realization



## Fragmented

Data sits in many systems, formats and reports.



## Manual & slow

Entry, cleaning and reformatting consume expert time.



## Hard to reuse

Naming and context vary, so each team starts over.



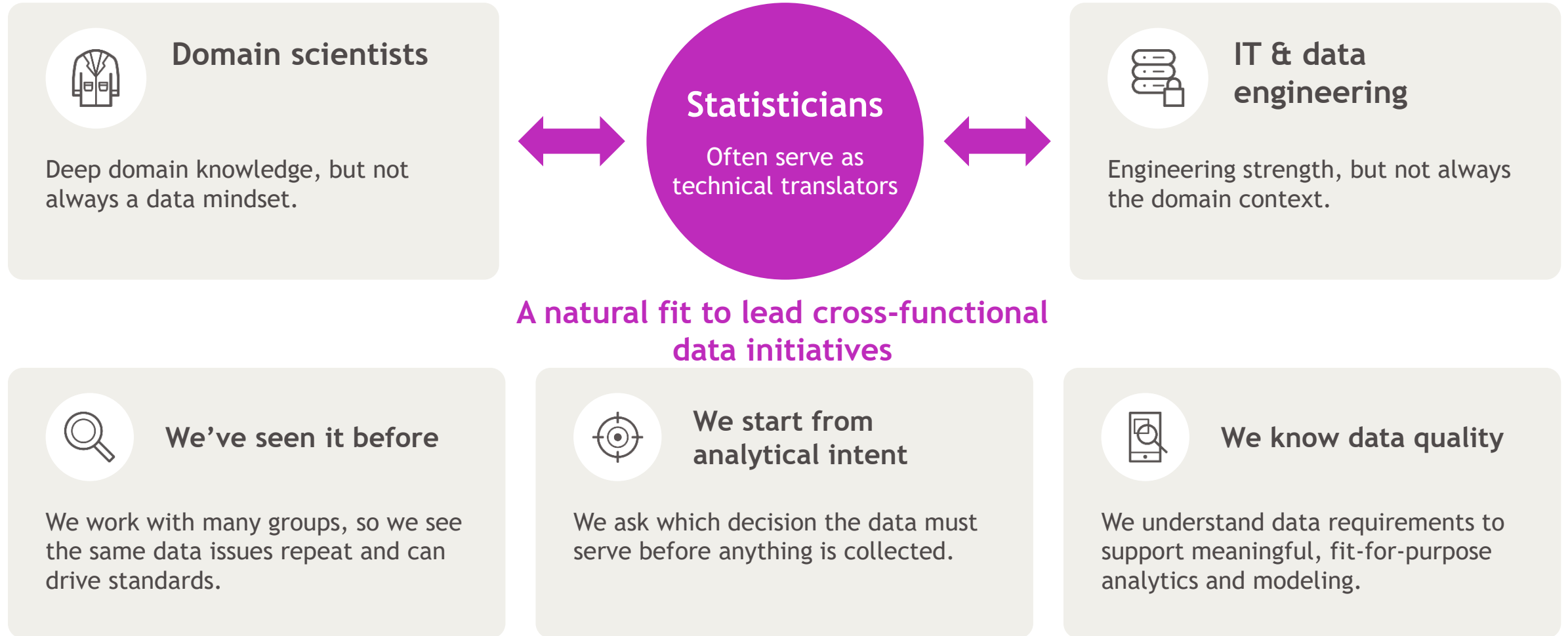
## Limiting to modeling

Without quality and structure, even great models add little.

**Powerful AI can't fix weak data**

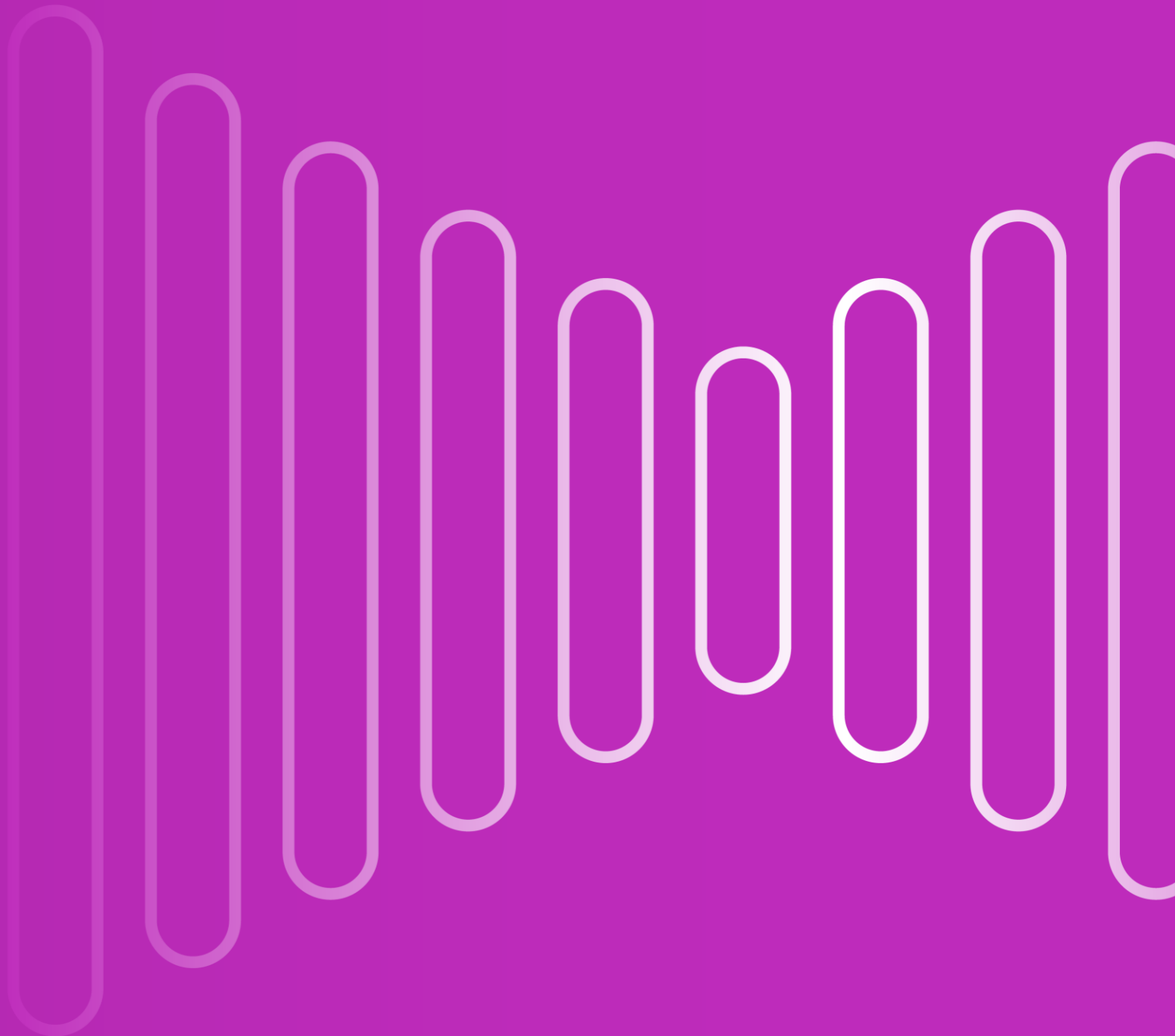
A robust data foundation is what unlocks AI and predictive modeling

# Statisticians Bridge Domains to Turn Data into Business Value



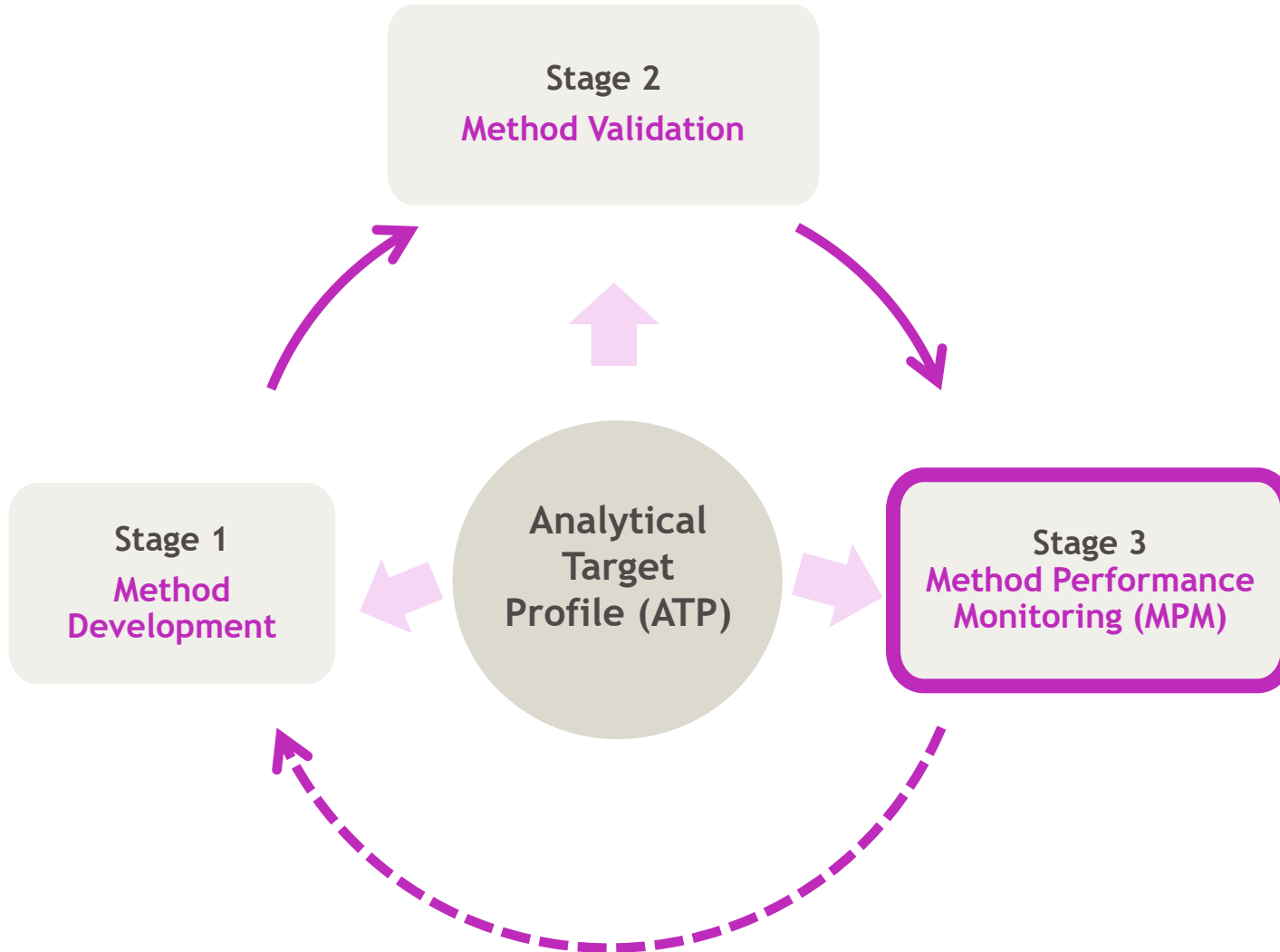
# Two Journeys

How statisticians lead cross-functional initiatives and drive data strategy



# Case 1: Building Method Performance Monitoring (MPM)

MPM keeps analytical methods fit for purpose across their lifecycle



**MPM is an integral part of the method lifecycle**

- Understanding long-term performance
- Detect unusual behaviors early
- Method trouble shooting and improvement
- Support product quality investigations

**Our 10-year journey began with a clear need, but no routine MPM.**

# Our Journey: from Scratch to Sustainable MPM

## 1 Foundation



### Established routine MPM

- Started with no routine MPM and manual, inconsistent entry
- Defined MPM objectives, parameters and data analyses
- Standardized Excel data entry
- Automated dashboards

**Role: program leader & SME**

## 2 Automation



### Automated the data pipeline

- Instrument software → structured ELN database → analysis platform
- Standard templates and a central database
- Designed to cut duplicate entry, errors and effort

**Role: connector**

## 3 Sustainability



### Handed ownership to the business

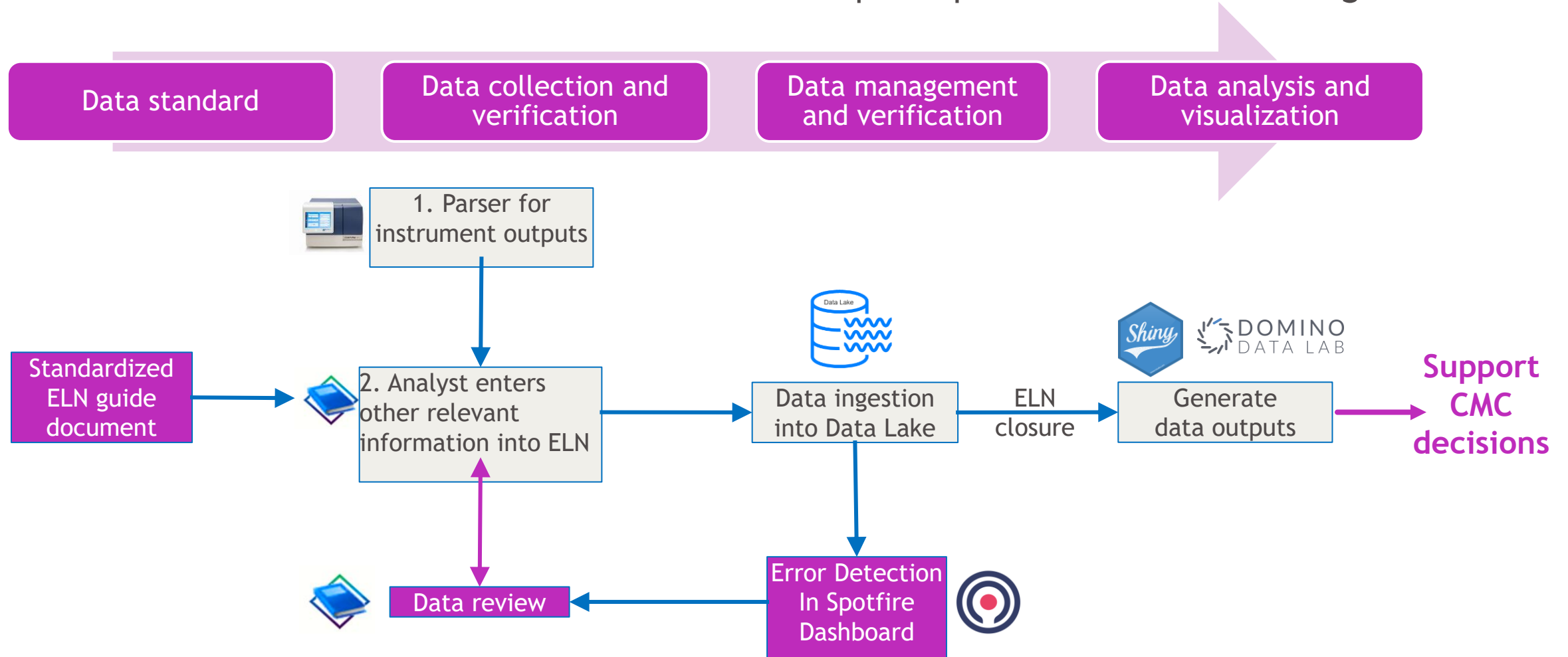
- Onboarded and upskilled an engaged scientist
- The business took over day-to-day MPM
- Consistently deliver business value and support CMC decisions

**Role: capability builder**

**Grow influence by stretching beyond expertise and building lasting capabilities**

# From Data Collection to Decision: An End-to-End Automated MPM Workflow

## Automated data workflow enables business ownership and portfolio decision-making



# The Lesson: Automation Reveals the Next Challenge



## The gap

Reviewing ELN data one experiment at a time made formatting issues that matter for analysis harder to spot.



## The fix

A scientist on the MPM team built an automated helper to flag data-quality issues.



## Lesson learned

Reassess quality checks whenever the workflow changes.

The screenshot shows a data table titled "ELN Trending Data" with columns including NB STATUS, Experiment, EXPERIMENT, GxP, D., T., Method, Me., PLA\_file\_name, S., Sample, N., E., LO., COA, QC, CoA, Ref, CoA, Sample\_Type, and Data table: ELN\_DATA. The table contains multiple rows of data, with some cells highlighted in red to indicate data quality issues. On the right side, there is a "Filters" panel with various filter options, including "Method\_Number", "EXPERIMENTCREATIONDATE", "Experiment", "GxP", "Sample\_Plate\_Position", "LOTNUMBER", "Coating\_Lot", "Detection\_Antibody\_Lot", and "Enzyme\_Lot". The bottom status bar indicates "2,606 of 2,624 rows", "0 marked", "101 columns", and "ELN\_DATA".

Scientist-built dashboard flags data issues in red

Digital maturity grows through iteration:  
Each step forward reveals new challenges, creating opportunities to learn, adapt, and improve.

# Case 2: Building an Analytical Data Foundation

## Business Need

CMC decisions need early prediction of long-term stability



**Decisions can't wait for real-time data**

Short-term accelerated and stressed data must reliably predict long-term stability to inform

- Shelf-life and out-of-specification risk
- Formulation, process and comparability choices
- Earlier, model-informed filing readiness\*

## Proof of Concept

Verify model prediction using historical data



**Test the published science on our molecules**

- Applied published predictive-stability approaches to historical BMS molecules
- Compared short-term predictions with available long-term observations

## Gap Analysis

Identify major challenges

**The bottleneck**

## Data cleaning

Records needed consistent naming and context; cleaning took considerable time.

**Next priority: a reusable data foundation**



\*Consistent with published biologics models and the ICH Q1 (2025) draft; draft guidance does not guarantee regulatory acceptance.

# Build Reusable Data Foundation

Prior data and digital leadership prepared us to lead a broader initiative

## Two shared data problems



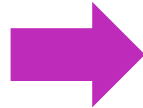
### Historical reports

Unstructured content and inconsistent formats made reuse difficult.



### LIMS data

Large, inconsistent datasets needed cleaning and context at scale.



## How we led beyond modeling

1

### Lead the cross-functional team

Unite scientists, IT, data engineers and modelers around one goal.

2

### Define the data requirements

Agree source data, standards and subject-matter-expert review.

3

### Set a common direction

Build reusable data products for shared needs, not one-off cleanups.

**OUR ROLE** Own the data that makes modeling possible, not just the model

# Data Products Now Drive Prediction and Decisions

## DELIVERED

20+

BMS molecules with standardized data, used to develop and validate our novel predictive stability model



### Historical data products

Standardized CQA, release and stability data



### Dashboards

Forced degradation and stability views across molecules and conditions



### Novel predictive stability model

Internal validation supports earlier shelf-life decisions and risk assessment

## WHAT'S NEXT



### NEXT

### Automate future data capture

Build a pipeline so consistent, model-ready data becomes part of the routine workflow.



### LONGER-TERM VISION

### Connect analytical knowledge

Ontology, data fabric and knowledge graph linking analytical and process data from discovery to commercial, for holistic CQA assessment and control strategy.

**OUR ROLE** Deliver for portfolio needs while building data strategy and long-term capability

# With Each Journey, Our Role Expanded



## Expert analyst and modeler

Define and execute sound analysis and modeling.

## Program leader

Lead cross-functional team to deliver solutions to meet portfolio needs.

## Connector

Translate business needs into data requirements. Connect business and IT.

## Capability builder

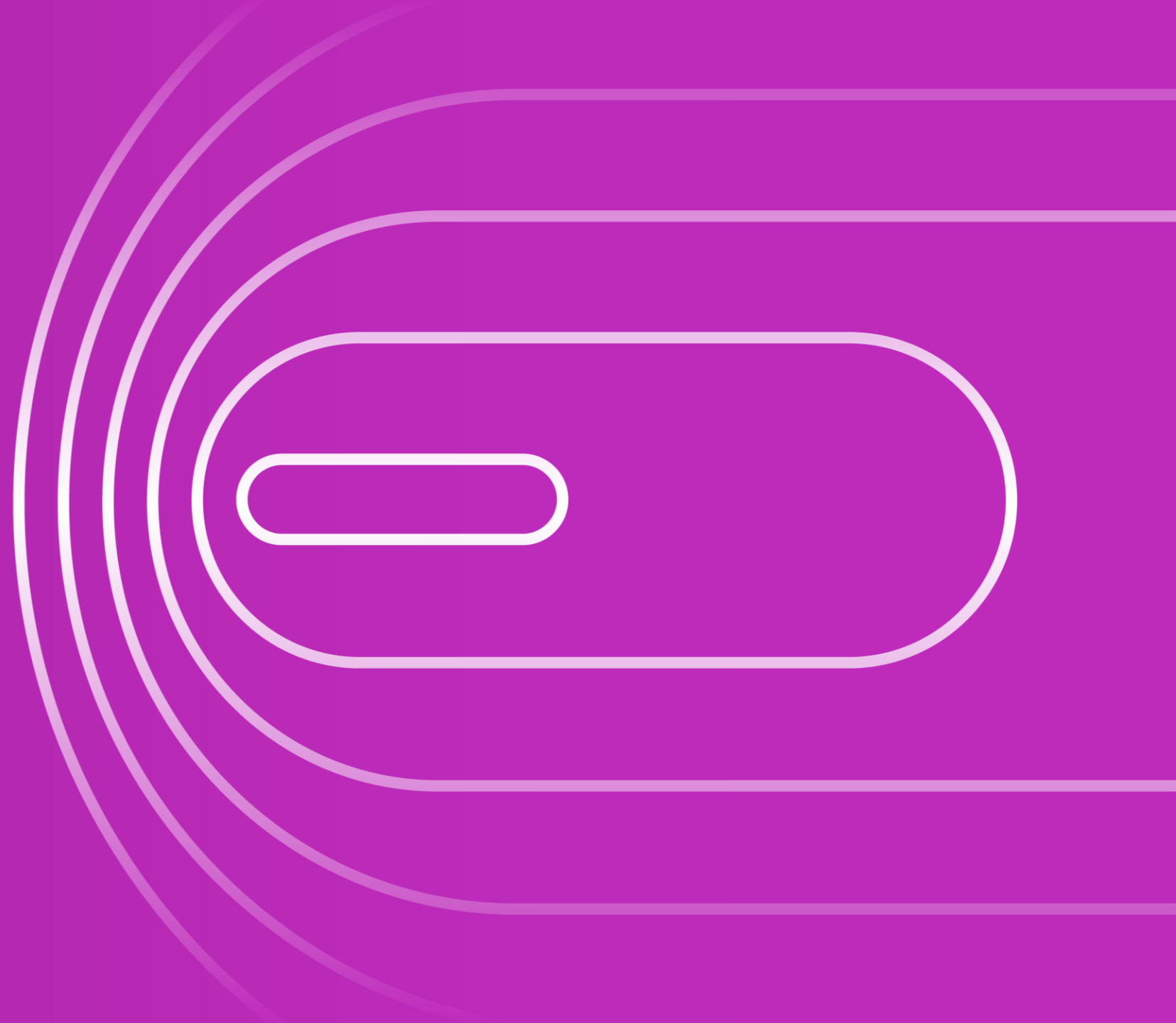
Develop scientists to own the program. Build data products and pipeline for shared needs.

## Strategy leader

Shape business strategy and guide portfolio priorities through data-driven insights.

**We keep learning by owning the next barrier to better decisions**

# The Path Forward



# Data Strategy Succeeds through Shared Ownership

## TODAY

Capture for compliance

Analysts join downstream

Local fixes and handoffs

Budget is the main lens

## THE SHIFT

### Capture for reuse

Scientists record structured data and scientific context at the source.

### Statisticians engage early

Define the intended decision and data requirements before collection.

### Named owners and stewards

Scientists, statisticians and IT share standards and quality responsibility.

### Leaders sponsor decision value

Prioritize reusable data products; tie investment to CMC needs.



## CULTURE

### Hands-on experience builds ownership

- Hackathon events: teams built a model or AI chatbot on real, imperfect data.
- Data gaps became visible; almost every team valued data quality more.
- Next: make it routine with shared standards and stewardship.

# The Statistician's Role in an AI-Enabled Workflow

**Execution  
accelerates.**

**Judgment remains  
essential.**

**Review AI outputs, test their  
limits and guide reliable  
implementation.**



## Scientific context

Turn a business need into the right question and analytical intent.



## Variability and uncertainty

Challenge assumptions, evaluate performance and probe edge cases.



## Patterns across programs

Recognize recurring data issues and shape common standards.



## Cross-functional translation

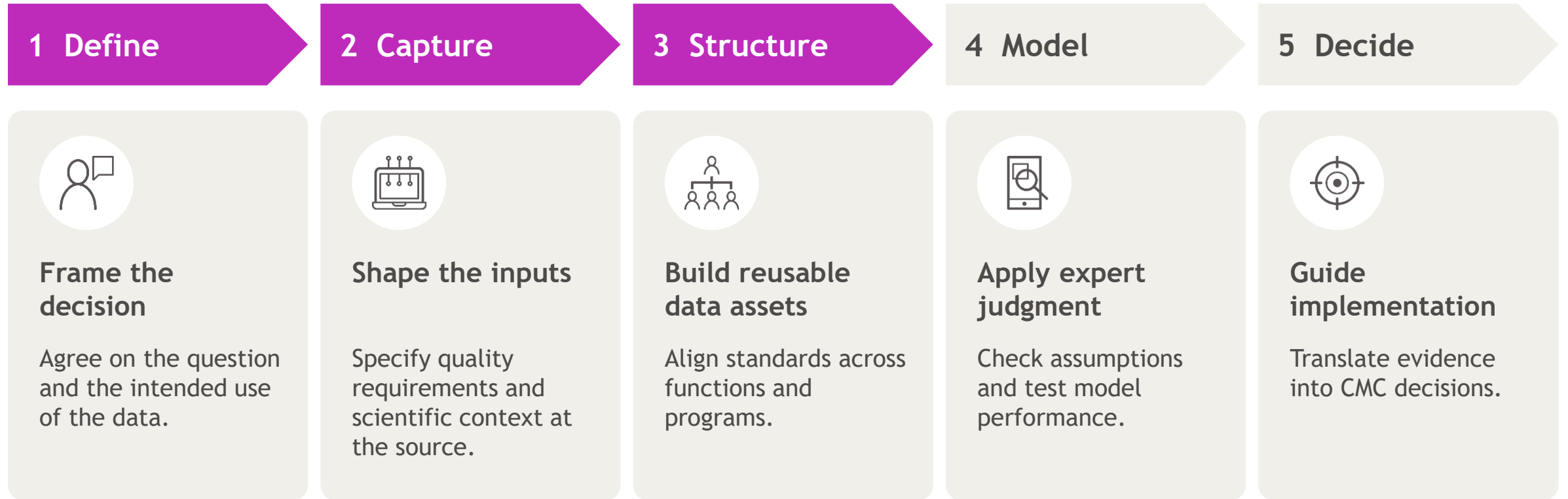
Connect scientists, IT and leadership around fit-for-purpose data.

*Complementary to, not exclusive of, the strengths of other disciplines.*

# Influence the Full Data-to-Decision Chain

Shift left: where we can add the most leverage

Where we most often engage today



**THE IDEA** Influence grows when we shape the workflow beyond the analysis itself

# A Practical Path from Analyst to Strategic Partner

|                    | Contributor                                               | Initiative leader                                                               | Strategic partner                                                                |
|--------------------|-----------------------------------------------------------|---------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| SCOPE              | Own the decision question                                 | Own a shared workflow                                                           | Shape enterprise priorities                                                      |
| WHAT YOU DO        | Join earlier, clarify intended use and surface data gaps. | Convene scientists and IT, pilot a reusable data product and improve usability. | Connect program needs and shape priorities for data capabilities and governance. |
| SIGNAL OF PROGRESS | A clearer problem and fit-for-purpose evidence.           | An adopted pilot and shared standards.                                          | A roadmap tied to CMC decision needs.                                            |

**Your next step: own one data barrier or decision.**

# Take-Home: Shape the Data, Shape the Decisions

1



**Data foundations decide whether model or AI delivers**

Data quality, structure and context turn models into value.

2



**Statisticians are well placed to build them**

Context, uncertainty, cross-program patterns and translation across functions.

3



**Influence grows with ownership**

Move from analysis to initiatives to strategy by owning the next barrier.

---

Think big,  
stay agile,  
and embrace  
change



# Thank you

## Questions & Discussion

**With appreciation to our partner functions:**

**Thank you for your expertise and collaboration throughout these journeys.**