



*The Clinical Data Lake:*

# Empowering the Data Caterer Vision with Veeva, Databricks, SAS, and Open Source

Sascha Ahrweiler (Bayer AG)  
Holger Dach (Bayer AG)

ET23  
PHUSE US Connect  
Austin, Texas  
March 2026





# Agenda

- 1 Why a new Clinical Data Technical Landscape?
- 2 The Data Caterer Persona
- 3 The Kitchen Architecture
- 4 Veeva – The Food Market
- 5 Clinical Data Lake – The Mis-en-place
- 6 SCE on SAS Viya – Cooking & Plating
- 7 Digital Protocol – The Menu
- 8 Early Lessons Learned

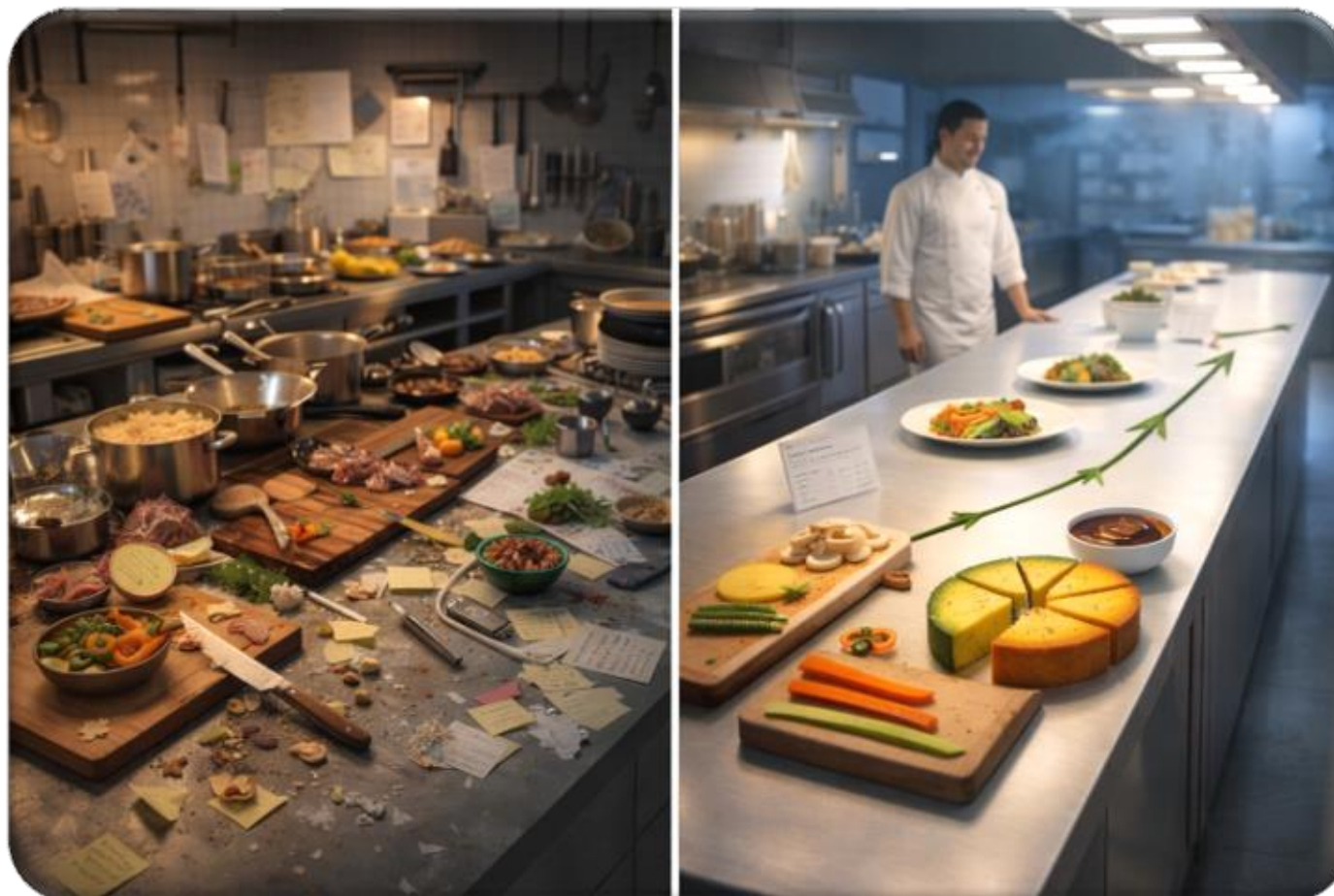




Bayer's aged technology landscape slowed down modern needs like data reuse, automation, and AI. We had to **rethink our clinical data landscape!**

Need faster flow from protocol → insights

Systems built for "decades" create rigidity



Digital transformation changes **work**, not only tools



Our **Digital vision** is to improve the end-to-end work, accuracy, traceability, and decision-making efficiency.



**study intent → governed data → analytics → decision**



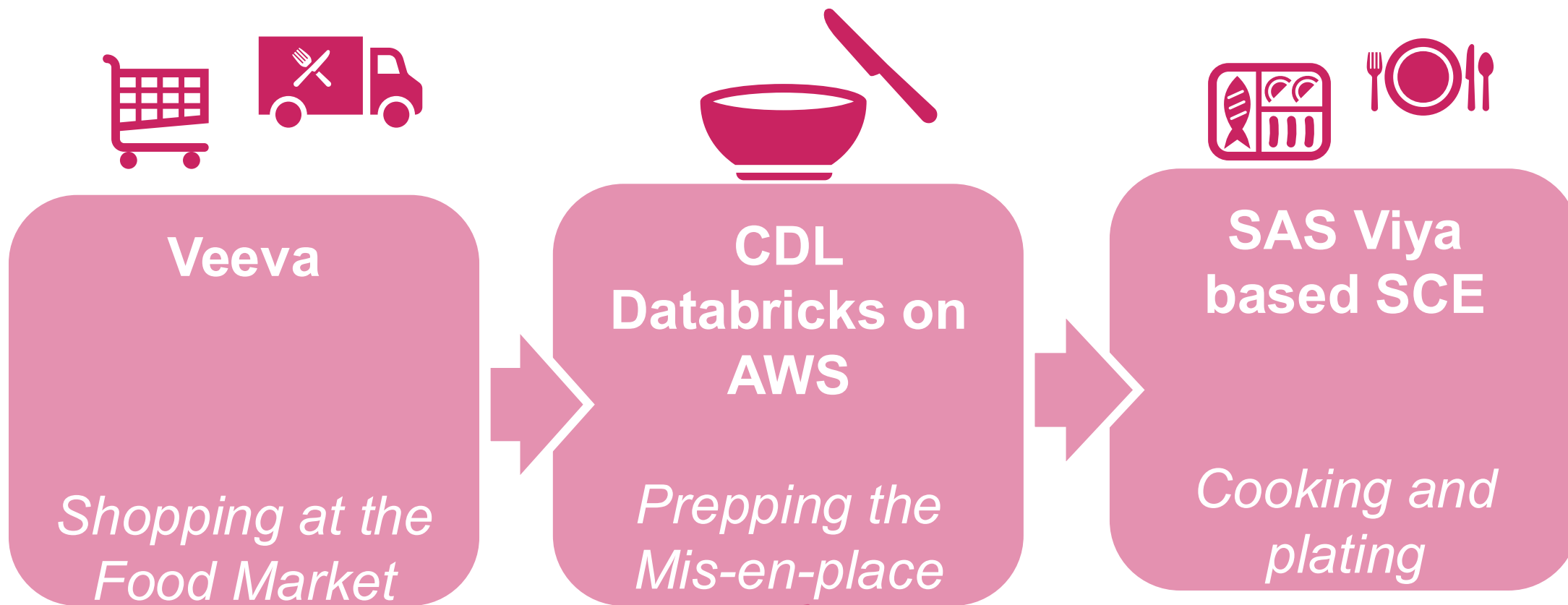
# With the rise of AI, statistical programmers evolve to become **Data Caterer** and move from code execution to service oriented insight delivery

- Understand stakeholder craving
- Deliver insights in the right form
- Less manual wrangling, more interpretation





# The **architecture of our kitchen** is based on three connected pillars to enable the Data Caterer









We shop our ingredients from from the **Veeva** food market. Veeva ensures provence, compliance, and structured data entry that feeds the ecosystem



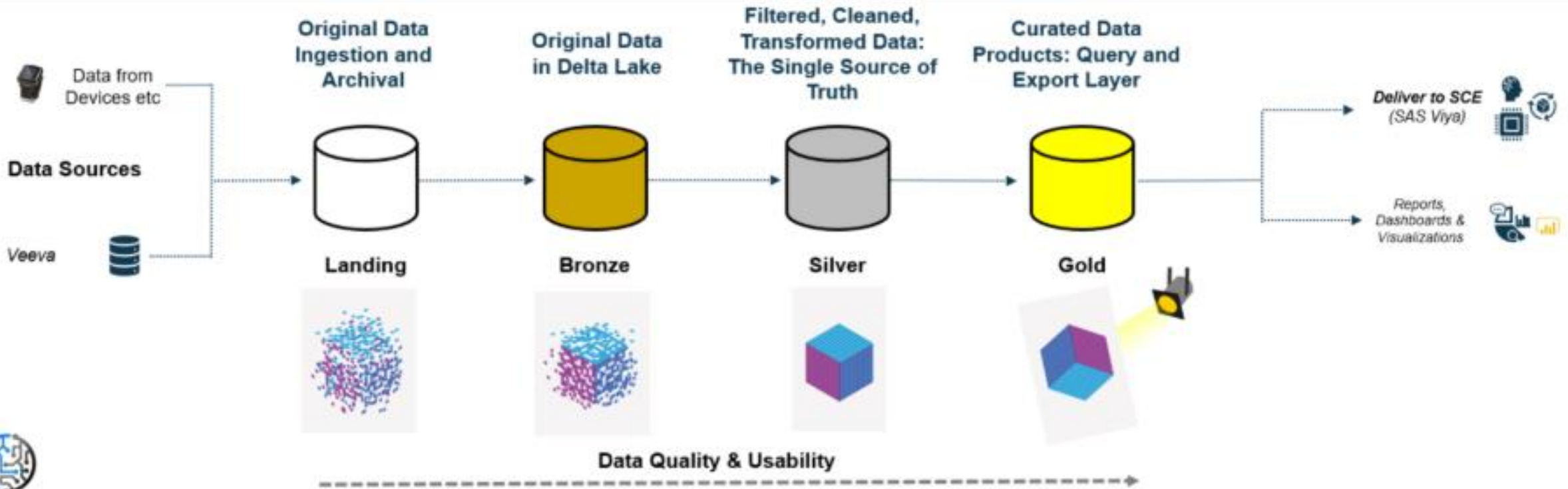
Veeva

*Shopping at the Food Market*

-  Operational platform across protocol development, clin ops, data entry
-  Audit trails & lifecycle control
-  REST/API interfaces & event-driven integration
-  Decision on **Veeva across multiple functions** to create synergies on different vaults



# We process and prepare our data using a medallion model in our Mis-en-place: The **Clinical Data Lake based on Databricks on AWS**





# Cooking our meals and plating them from customer consumption in reproducible quality in our **SAS Viya based SCE** kitchen



SAS Viya based  
SCE

*Cooking and  
plating*

## Legacy

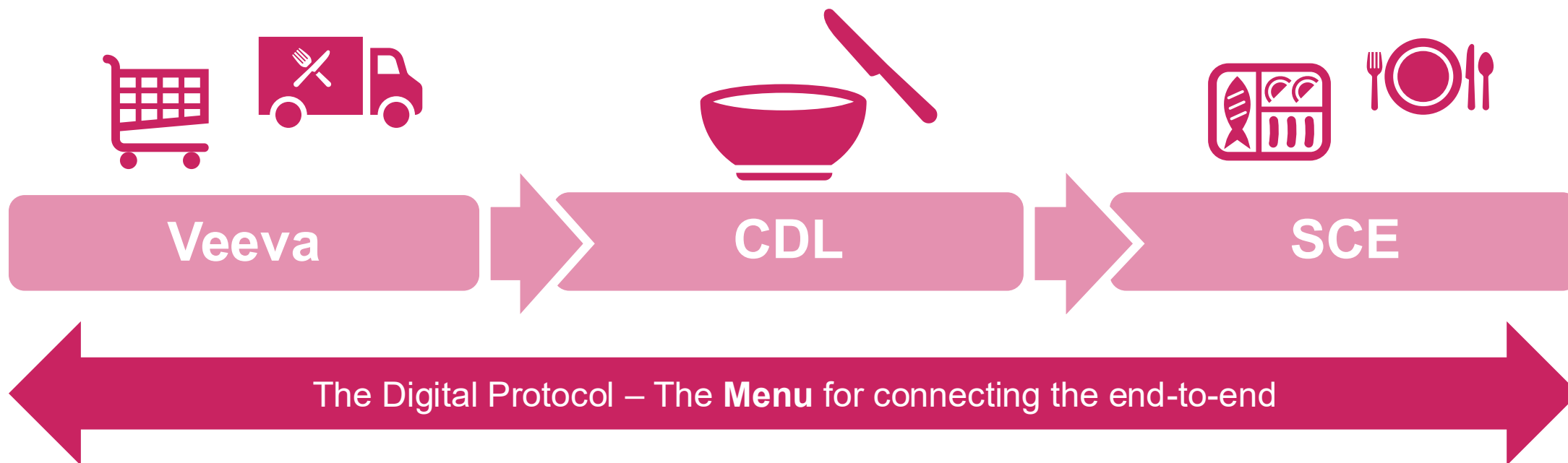
home grown SAS 9.4 Platform (>10 years) combined with validated R environment




## Future

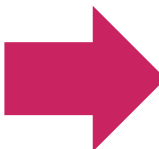
- 🍴 Scalable integrated SCE with SAS, R, Python under validation
- 🍴 Containerized compute
- 🍴 Role-based access & workflows for unblinding and snapshots
- 🍴 Traceability, promotion control



# The **Digital Protocol** as the overarching recipe completes the flow from study intent through execution to analytics



-  Machine readable connection
-  Eliminates translation steps
-  Enables automated downstream processes



**Multiplies automation & AI readiness**



# Our early implementation reinforced a set of **pragmatic lessons**



**Flow matters more than tools.** If you cannot describe the end-to-end flow, you will optimize local steps and lose overall speed.



**Governance must be designed in from day one.** Lineage, versioning, and controlled sharing are product features, not afterthoughts.



**Preparation and execution should be separated.** The lake is for harmonization and reuse; regulated outputs require controlled execution.



**Transformation libraries should be standardized early** to avoid rebuilding complexity study by study.



**The digital protocol is a multiplier.** When study intent becomes machine-readable, the stations connect and automation becomes realistic. AI readiness then becomes less about model choice and more about controls: access, traceability, explainability, and change management.



# Conclusion

Re-considering a clinical data technical landscape is **not about choosing one tool**. It is about building an ecosystem that aligns platforms, processes, and people around decision enablement.

Just as in a restaurant, **all components have to fit together** to enable the Data Caterer to serve what our customers are craving for: Data Insights for data driven decisions

- // Veeva provides a governed operational backbone
- // CDL Prepares and harmonizes data for reuse and analytics
- // The SCE provides controlled execution and analytics delivery

When the kitchen is built well, clinical data is not only stored. It is curated and served in a way that helps our stakeholders **make faster and more confident decisions**.

Health for all, Hunger for none



Thank  
you!