An Autonomous self-optimizing bioprocess reactor, cyberpunk style – Dall.E-3

sanofi

Digital Transformation in Process Development

the drivers for structured and contextualized experimental data

Christian Airiau, PhD
Global Head Data Sciences, CMC, R&D, Sanofi
iCMC-Digital Transformation lead

27JAN2025

Objectives of today’s session

- Pharmaceutical Industry progresses towards its Digital Transformation goal
- Investment, commitment, initial results
- Momentum is clear, but is it enough to deliver on the industry expectations

Today’s objectives:

- Share progress of Digital Transformation program for Process Development
- Focus on the efforts needed for Data structuration / contextualization
- Understand Business drivers and Business Value
- Reflect on challenges – focus on Talent attraction: Tech Vs BioTech

Digital Transformation: a business decision

- Pharma Industry is making progress in its Digital Transformation journey
- However, we are still lagging many other industries (Automotive, aeronautics, Supply chain)
- Major issue remains on structuring/ accessing our data at scale to enable advanced analytics
  - ***Running out of excuses…***

- Automotive industry, self driving cars manage much more complex data & infrastructure than our industry
  - Occupancy NN uses 1.4B video frame
  - 30PB of data
  - Deploy NN to 800,000 cars
  - 75,000 NN models trained (2022)
  - Planner decision every 50ms for Autopilot mode

The slow pace of Digital Transformation in our industry is not due to technical limitations: this is a Business Decision

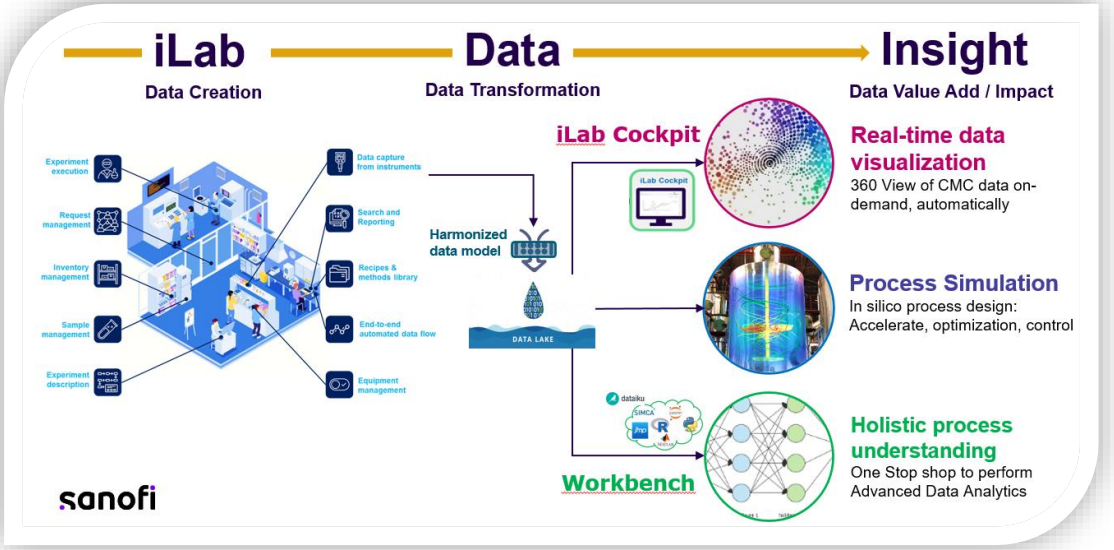
01 A Digital Transformation journey

Sanofi is engaged in a multiyear Digital Transformation program for Process Development activities

*More productive CMC Labs & **faster dev processes** through **data-driven products**, feeding **AI/ML/Modeling** agents, supporting Sanofi launches*

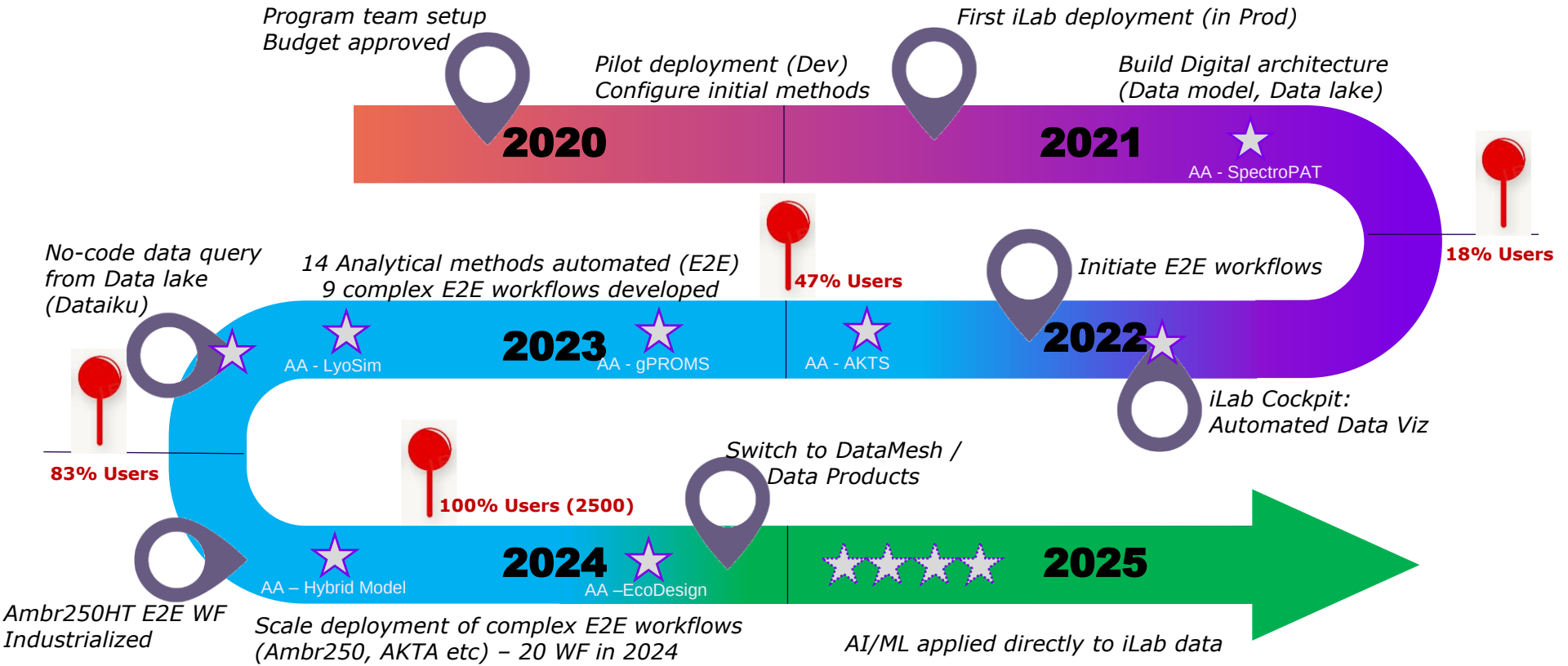
Scope:

- *+2,500 users, ~100 labs, ~20 sites*
- *Process Development in R&D (CMC) and Commercial (MSAT)*
- *All modalities – Biologics/Synthetics*
- *Pharma & Vaccines*
- *GMP and non-GMP*



sanofi

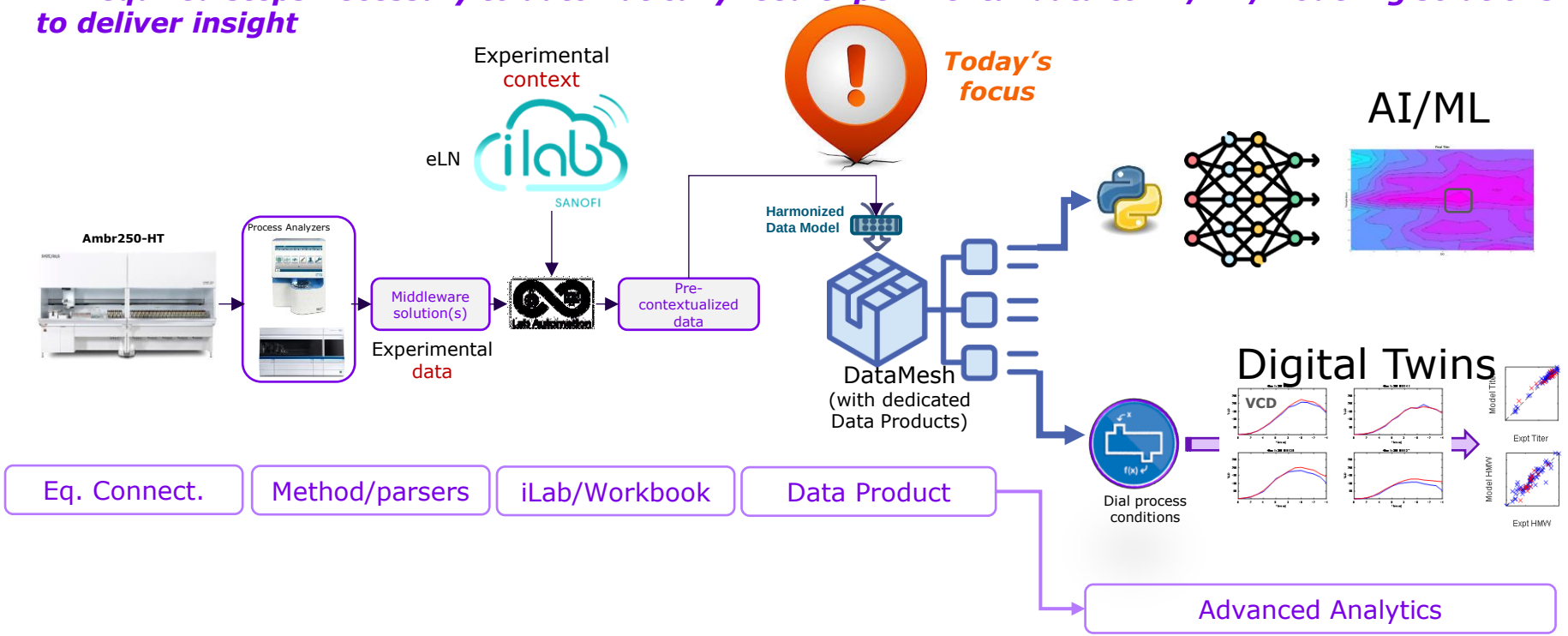
This is a Journey…



sanofi ★ ***Industrialization of an Advanced Analytics Capability***

Focus is now on implementation of E2E Workflow

All required steps necessary to automatically feed experimental data to AI/ML/Modeling solutions to deliver insight



All necessary and critical steps implemented by program Factories to deliver Business value

02 Data Structure and contextualization

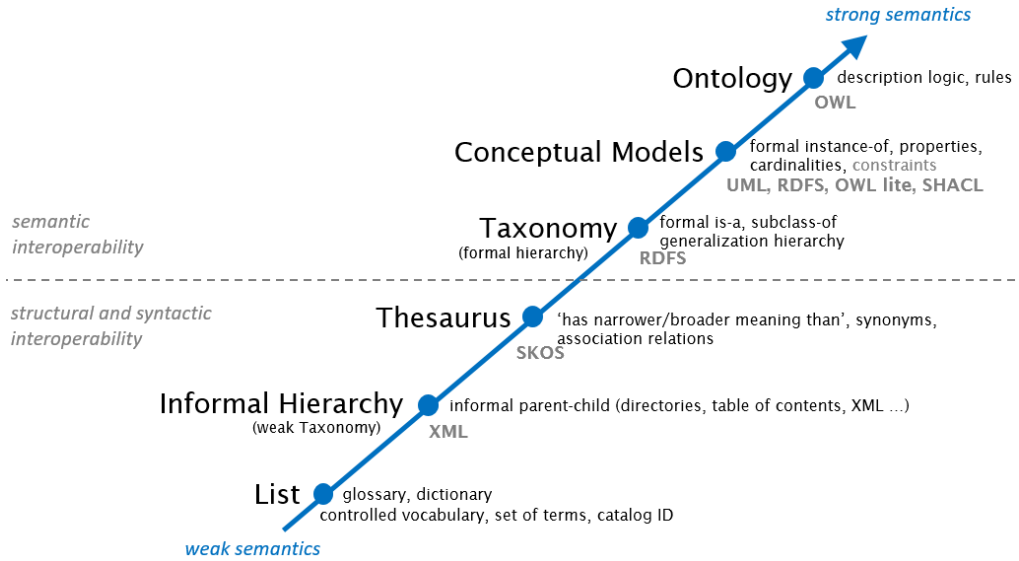
A NON-NEGOTIABLE STEP

Let’s align on some terminology

Ontology

A formal description of knowledge as a set of concepts within a domain, and the relationships that hold between them.

It ensures a common understanding of information and makes explicit domain assumptions thus allowing organizations to make better sense of their data. (NIIMBL consortium)



Sources

- Deborah L. McGuinness. "Ontologies Come of Age". In Dieter Fensel, Jim Hendler, Henry Lieberman, and Wolfgang Wahlster, editors. Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press, 2003.
- Michael Uschold and Michael Gruninger “Ontologies and semantics for seamless connectivity” *SIGMOD Rec.* 33, 4 (December 2004), 58-64. DOI=http://dx.doi.org/10.1145/1041410.1041420
- Leo Obrst “The Ontology Spectrum”. Book section in of Roberto Poli, Michael Healy, Achilles Kameas “Theory and Applications of Ontology: Computer Applications”. Springer Netherlands, 17 Sep 2010.
- Leo Obrst and Mills Davis "Semantic Wave 2008 Report: Industry Roadmap to Web 3.0 & Multibillion Dollar Market Opportunities”. 2008.

From Arne Balzer, Mathias Wolf, Olivier Gigonzac

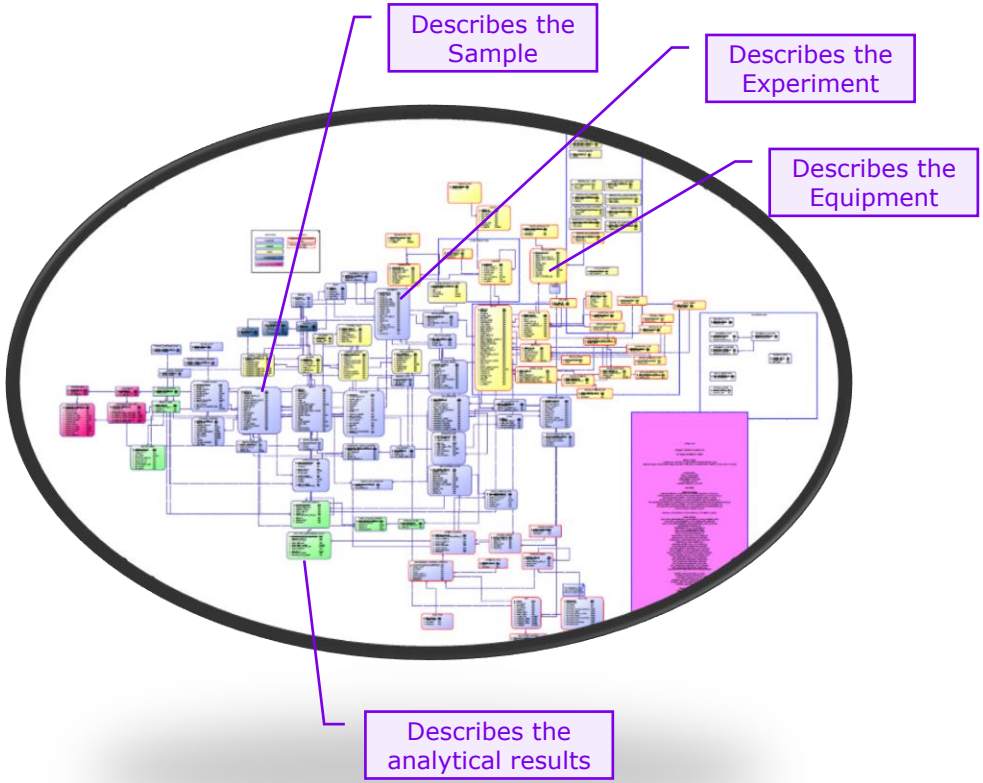
Cornerstone of the Digital Transformation: Data Model

A non-negotiable step

Data model organizes data elements and standardizes how the elements relate to one another. It explicitly determines the structure of data

Ontology: sets the general framework
Logical Data Model provides a visual / conceptual relationship of data entities
Physical Data Model is the translation of the Logical Data Model into the structure of a database

Each data point generated in the lab is mapped in the Data model



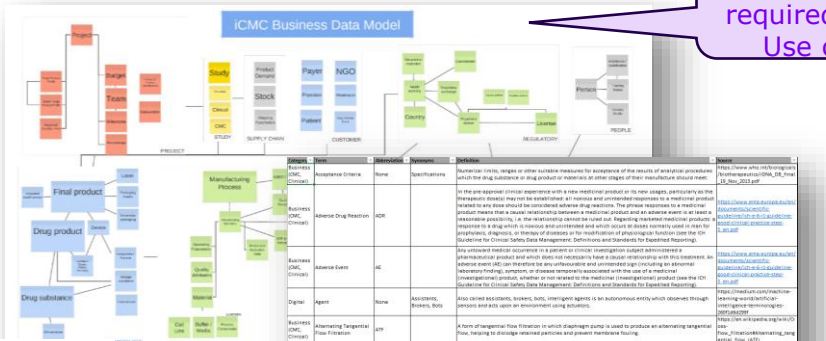
Define Data Model, through Use Cases

*Starts with multiple **Use Cases** defining the Logical data model*

❶ End-users explain what they need to do

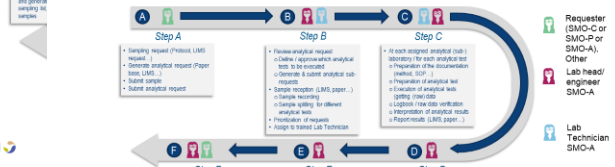
❷ Data mapping / connectivity required by the Use cases

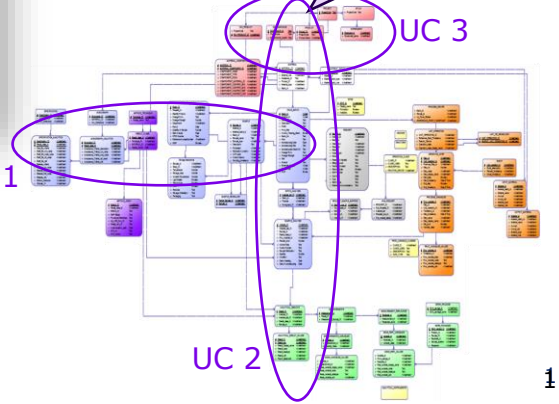
❸ Combination of all use cases builds the full picture



Use Case 2: Biotechnology: Technical Batch Documentation

Uses Case 1: Analytics: Batch Comparision



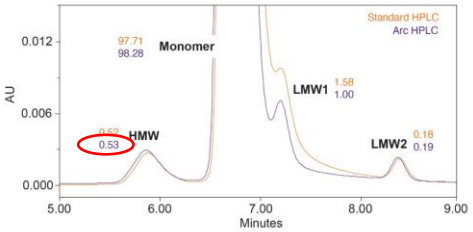


Drivers to build strong data model

The context around "0.53"

Method

- Version
- Analyst
- Date
- Site
- Mobile phase
- Etc…



0.53 %

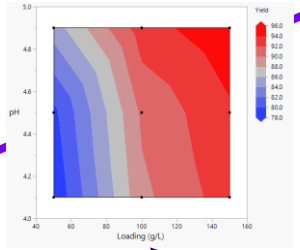
Unit

Downstream process Step

- Project
- Purification step
- Equipment #
- Sample submitted
- Process conditions
- Etc…

Bioprocess step

- Project
- Development Phase
- Ambr250 run
- Vessel #
- Equipment #
- Process conditions
- Etc…



Study associated with experiment (DoE)

- Workpackage
- Objective
- Conclusions
- Model
- Etc…

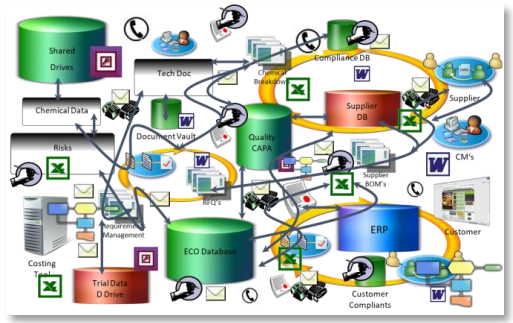
sanofi

Drivers to build a strong data model

Providing Insight

"Our PPQ run shows an Out of Trend result in HMW: 0.50%. Have we generated such results in the past?"

When your data is not structured:

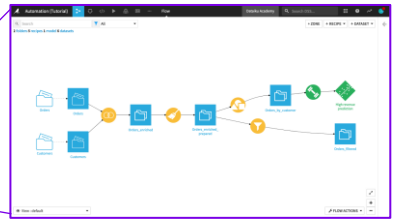


5 scientists, over 2 weeks

- "We could not find an example of such result"
- "What do we do with our PPQ run?"

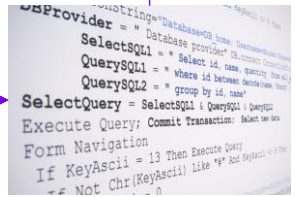
When your data is structured:

Lab Scientist



Self-served data prep.

No-code SQL



- "This project generated 0.53% of HMW during Stretch DoE study in 2022"
- "Can this help justify PPQ Out of Trend results?"

Bioprocess Manufacturing – common Ontology

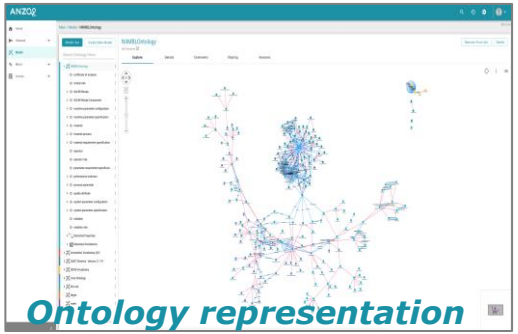
Implemented via NIIMBL – Big Data program

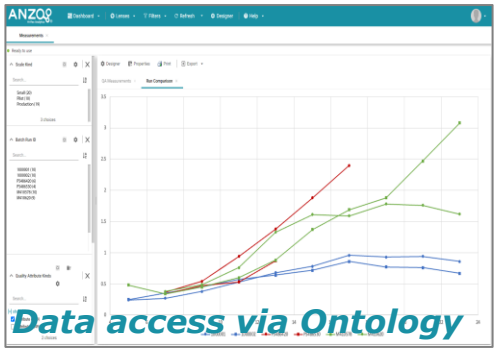
NIIMBL developed a *Bioprocess manufacturing Ontology* using NIIMBL members input.

Ontology ismanaged under Industrial Ontology Foundry (IOF) and "Basic Formal Ontology" (BFO) model

The NIIMBL Ontology for Bioprocess Manufacturing is now Open Source.

- Available to all in the Biopharmaceutical industry
- Lower the access of Ontology/Data model across the industry
- Provide standardized framework to manage data



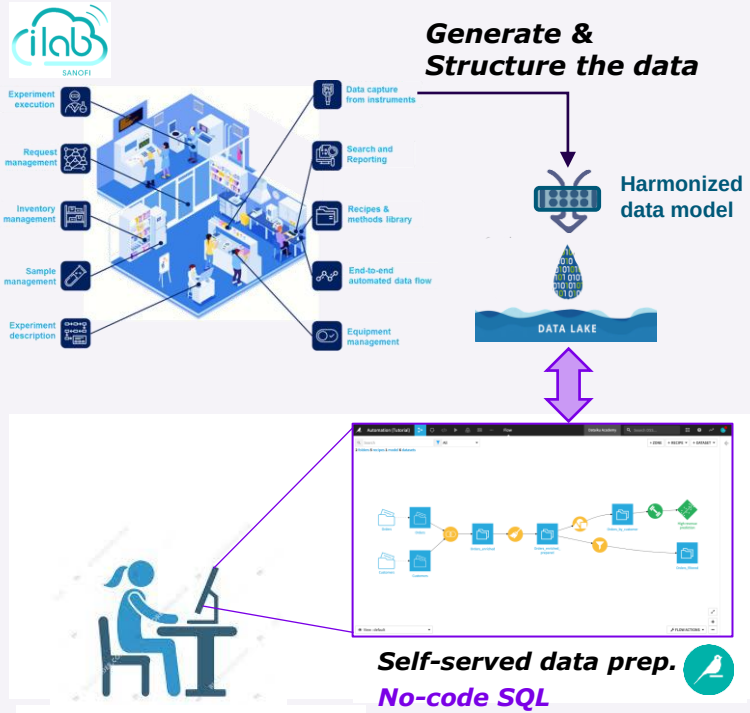


sanofi

From Boonserm Kulvatunyou, Ph.D.
National Institute of Standards and Technology part of NIIMBL Big Data Team

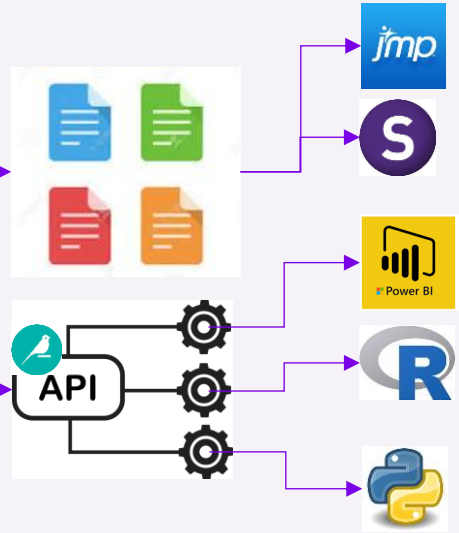
03 Where are we headed

Joining the dots: Generate data / query / analyze



Experimental data automatically structured, contextualized, centralized in Data lake

Direct query by Scientists / Data Scientists to feed into Stats / MVA / Modeling applications



Stats / DoE

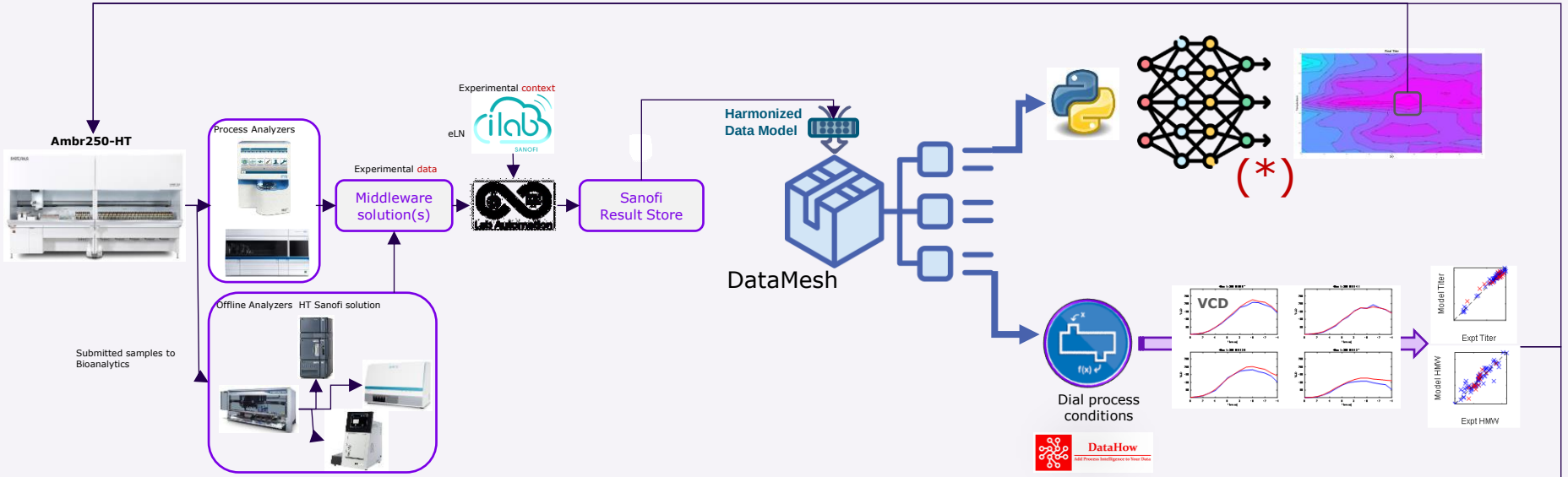
Multivariate analysis

Advanced Analytics
ML

Join the Dots

All cell culture experiments to feed into In-Silico design (Hybrid) and in ML models

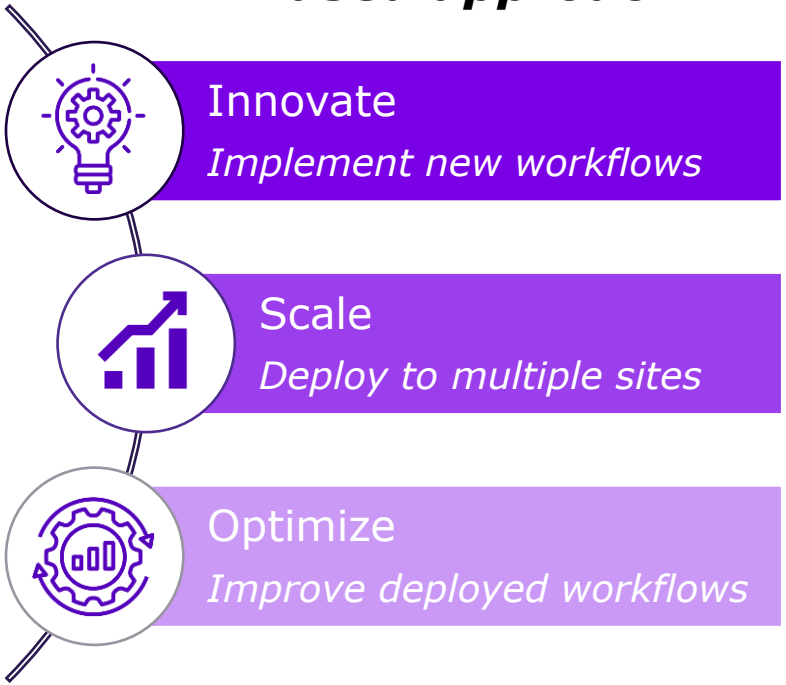
Use proposed new set of experimental conditions to drive to **Process Optimality**



(*) with automated data flows, large numbers of run are becoming available, suitable for ML approaches, specifically intending for Transfer learning across products
- 6 first months of deployment at 1 site: 420 batches / 6700 samples / 175,000 measurements

E2E Data Flow – Moving to full scale

Phased approach



10 new complex workflows per year
- Ambr / AKTA / Chemspeed / Tecan etc

About 20 new analytical methods per year

20 Replicates of workflows / analytical methods
- Copy/paste by first intent (accelerate deployment)
- But heterogenous site-specific equipment connectivity

Continuous improvement - established workflows
- Industrialization, support (Long term Op. model)

sanofi

4 areas of focus - Ambition

Become a CMC organization powered by AI at scale

Process Modeling

Physics based Digital Twins

- Reduce wet-lab experiments by dry-lab Digital twins
- Design / Optimize / Control all main Unit Operations in CMC
- Define optimum process conditions, faster and cheaper

AI / GenAI

Optimize processes and increase productivity

- Expert AI agents:
  - *Identify complex process/quality relationships in large datasets*
  - *Drive process productivity*
- GenAI agents for CMC:
  - *Technical reports / dossier generation*
  - *AI-driven knowledge management system - "Digital SME"*

Automation / Robotics

Hyper productive labs

- Automation of low value-added tasks:
  - *Increase project team productivity*
  - *Enable scientists to focus on science*
- Lead the Robotics race:
  - *Integrate complex laboratory tasks*
  - *Assess Humanoid robotics in lab environment*

Digital Transformation

Underpinned by complete access to structured data

sanofi

04 Skills and roles

A NON-NEGOTIABLE STEP

Key challenge
Workforce evolution matching Digital Maturity Ambition

- **Strategic investment in our employees:**
  - Structured training in Data Science / Digital
    - *Specific topics: Statistics / MVA / ML / Modeling*
    - *Different levels: Introduction / practitioner / expert*
    - *Community of practice – post training, reinforcement*

- **New skills in Digital / Business organizations**
  - Investment in new skills and new roles

- **Skills:**
  - Data engineers
  - IoT experts (EqCx)
  - Solution Architects
  - Semantic Engineers
  - Data Governance
  - ML
  - Systems modeling

- **Roles:**
  - Data Owners
  - Product Owners
  - Scrum Masters
  - Business System Owners
  - Business Analyst
  - Modelers
  - Data Scientists
  - Digital Training coordinator

- Upskilling
- Reskilling

- ***Increasing Digital Maturity does not happen organically.***
- *It requires time, efforts, investment*
- *This needs to be a key part of the Digital Transformation Strategy*

Biotech Industry challenge

Attract the top tier Data Scientists from tech industry

- **Top Data Science graduates are primarily attracted to tech companies:**
  - Tech companies remain the most competitive salary-wise:
    - *$100,000 – 180,000 for Data Scientists in Tech Vs $85,000 - $150,000 in Healthcare*
  - *Digital maturity / access to state of the art Tech Stack*
    - *Big Pharma have a long way to go to provide the exciting Digital environment to attract Data Scientists*

- **What can Biotech do to attract more talents from tech companies**
  - *Be ready to pay more? Probably, but not only*
  - *Leverage the trend to give meaning to their work – Active external communication*
    - *"Tech for good" or "data science for social impact" are key trends in data science community*
  - *Maturity of typical Big Pharma "TechStack" and access to structured data is a minimum threshold to attract key talents*

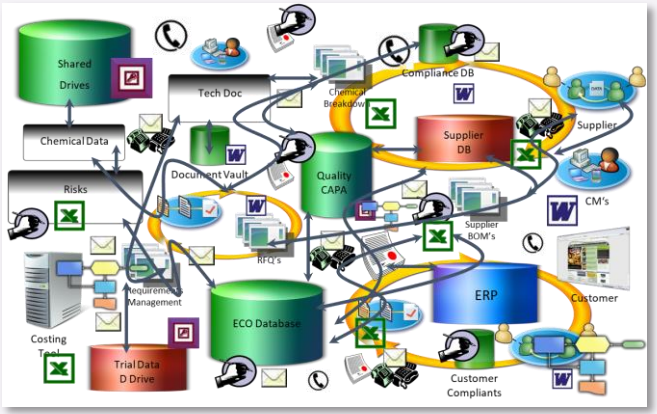
Which Sandbox do you want to play in?

sanofi

Conclusions

Current versus future state

From where we were …

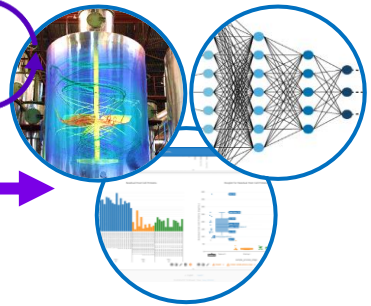


…to where we will be

Hyper efficient labs

FAIR data - structured

GenAI



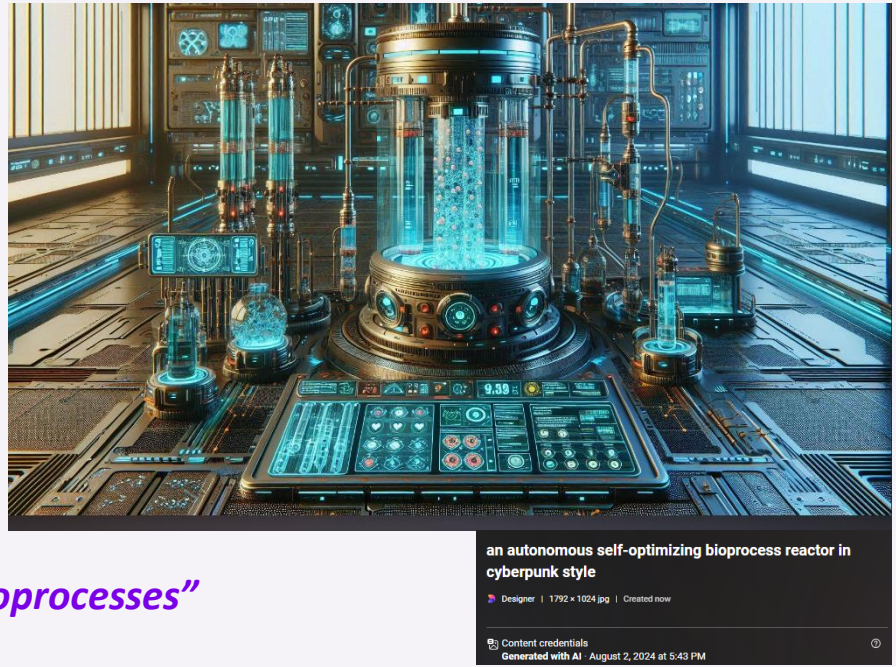
Digital Twins + ML/AI

Launch Excellence + Fast to patient

sanofi

Accelerate the pace

- Industry is moving in the right direction
- We must learn from other industries to catch up on our Data foundation
- Focus and continued investment is required



Moving towards "Autonomous self-driving bioprocesses"

Acknowledgements

- Anne Pomarede-Aschard – Digital lead
- Mohan Boggara
- Christelle Le Beaudour
- Ethan Penner
- Gabriel Lurz
- Louis Joos
- Mathias Ganz
- Olivier Gigonzac
- Sven Kaiser
- Guillaume Lardier