

# From Data Strategy to Delivery

## Reconciling scientists' data needs with business constraints

A product from the LAD project

Felipe Albrecht, PhD

Roche Pharma Research and Early Development (pRED), Data & Analytics, Strategy & Governance Chapter

---

## Felipe Albrecht – Senior Scientist in the Strategy & Governance Chapter in Pharma Research at Roche

- Education
  - Bachelor and Master in Computer Science
  - PhD in Computer Science focusing on Gene Regulation at the Max Planck Institute for Informatics
- Experience
  - 20+ years Software Development & Data Engineering
  - 10+ years of Scientific Data Management & Analysis
  - 5+ years of Enterprise Architect, Lab Automation, and Data management solutions in Pharma Research
  - 1+ year Data & Analytics Strategy and Governance

## Roche is a global leader in pharmaceuticals and diagnostics

**CHF 58.7 bn**

Annual sales  
(2023)



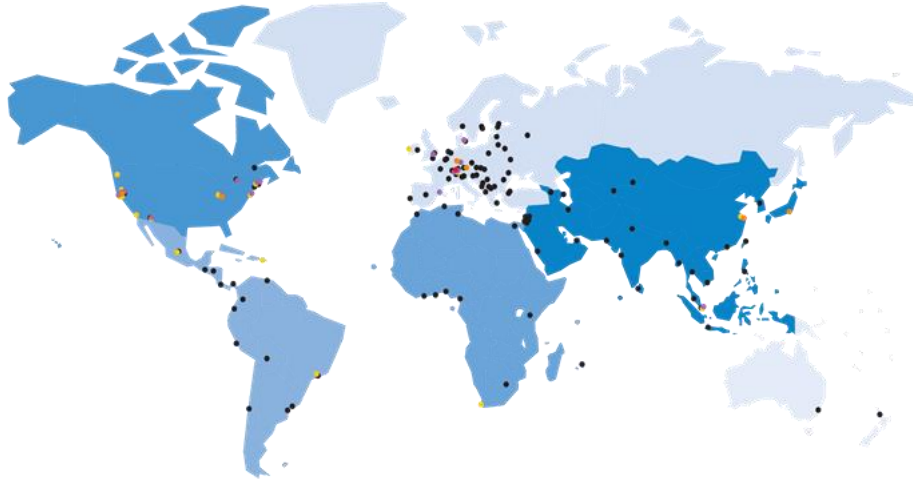
**103,605**

Employees  
(worldwide)



**1896**

Founded in Basel  
(Switzerland)



**17** R&D locations

**26** Manufacturing sites

**99** Affiliates

**150** External partner  
organizations

(worldwide)



**#1** R&D investor  
in healthcare

- CHF 12.2bn R&D investment in 2020
- Top-10 R&D investors across all industries

---

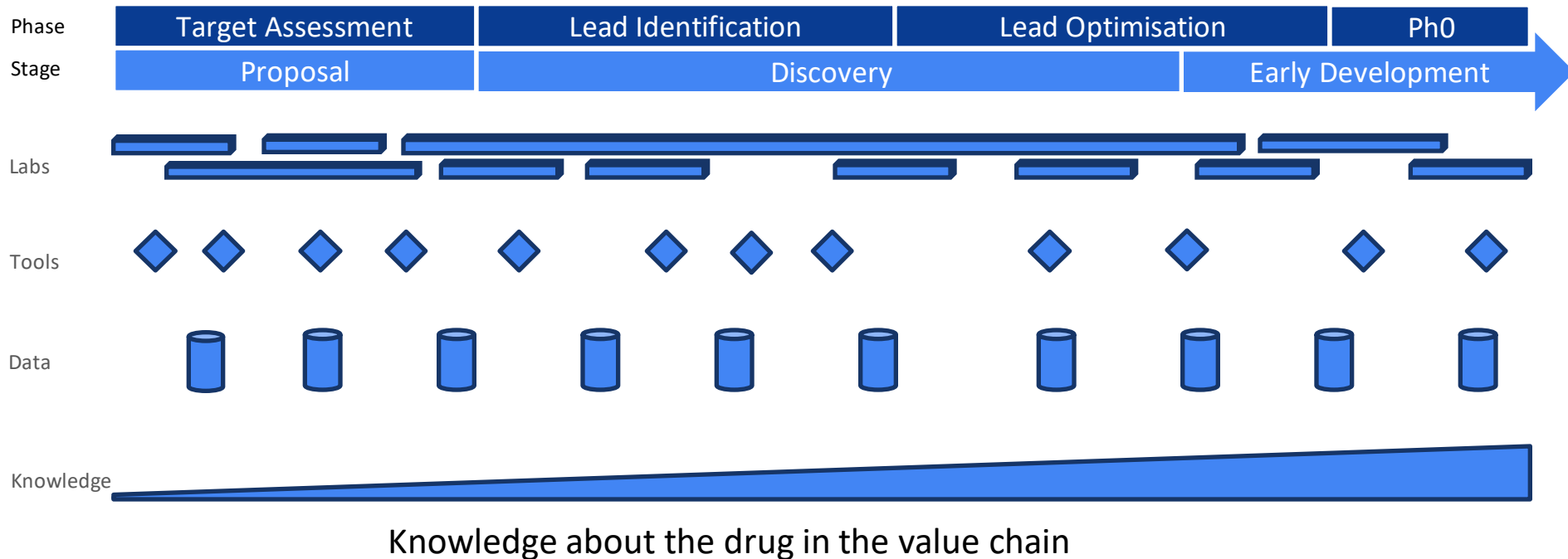
## What we cover today

- 1. The pharma research value chain generates and consumes data that must be actively managed.*
- 2. Actively Managing Data requires a data strategy.*
- 3. Without a Data Strategy, there can not be systematic use of Artificial Intelligence (AI)*
- 4. Value realization comes from the better use of the data by AI and scientists, and it is a long-term process.*
- 5. Data Strategy development follows a clear process that combines strategy, technology, and architectures.*

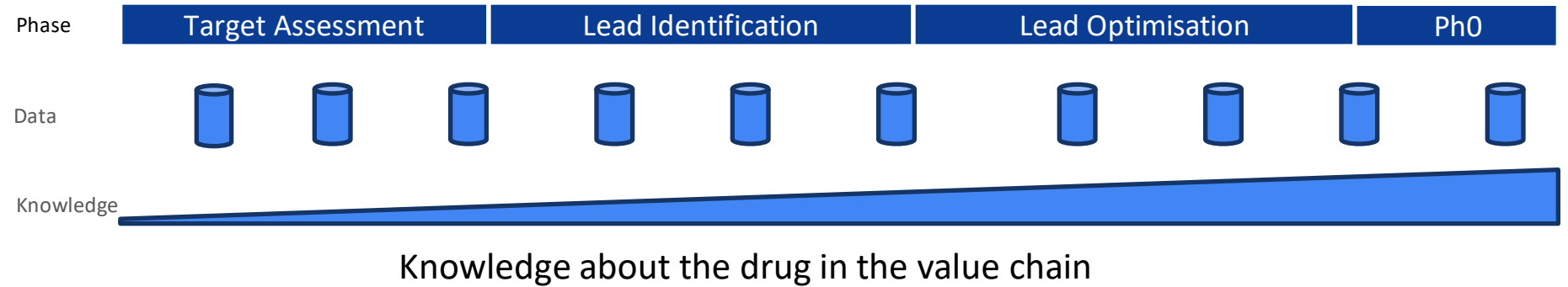
I present what we need to build the foundation of an AI from the perspective of data engineering

## Value chain in pharma research

The goal is to bring the best possible molecule to development in the shorter amount of time

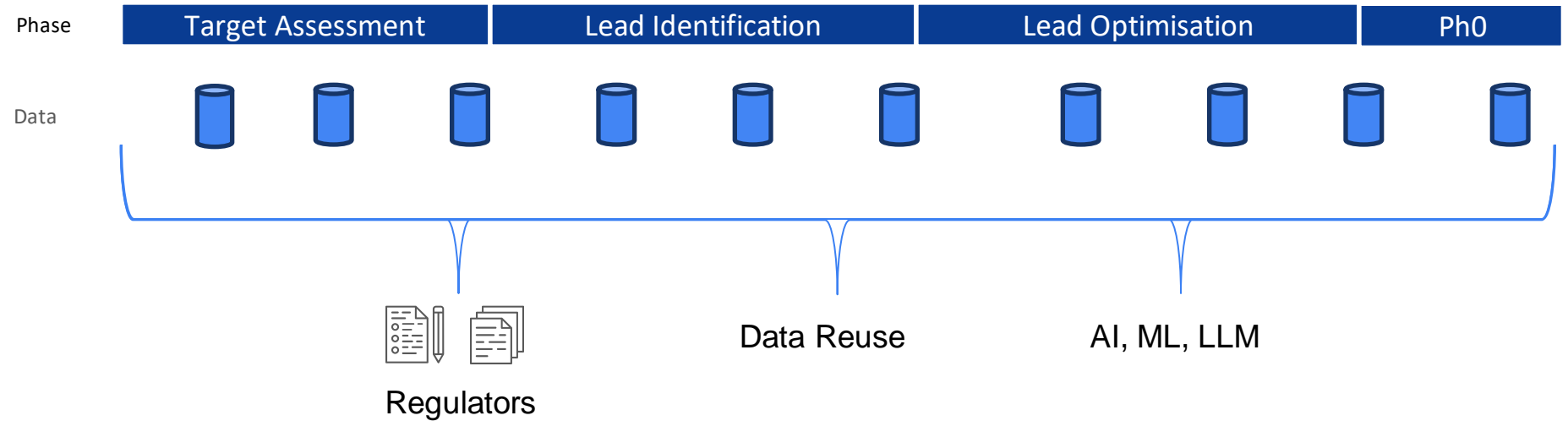


Value chain in pharma research build knowledge based on its generated data.



The generated data has different purposes

The main aspects of today's data is for filing regulatory documents



The value chain is not linear and straightforward,  
but it is an interactive process

Phase

Target Assessment

Lead Identification

Lead Optimisation

Ph0

Stage

Proposal

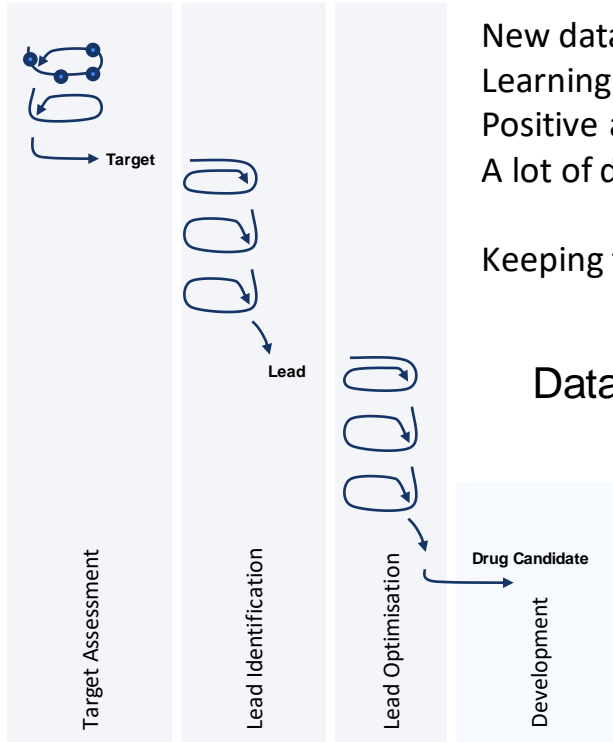
Discovery

Early Development





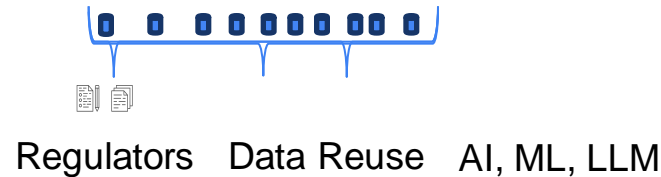
## The data is generated in loops rather than straightforward lines and loops



New data is generated in all steps of drug research  
Learning and optimization cycles are repeated  
Positive and negative results data must be stored and managed  
A lot of data is being generated but being treated as a byproduct

Keeping the data from each step is leveraging its learnings!

Data is an asset but is still being treated as a by-product



---

## Key aspects about data in pharma pre-clinical research

**Data is generated in all steps of pharma research.**

We need to provide this data for secondary re-use by scientists and AI applications

Data products for the data is a powerful idea

**We need to consider that not all the data is equal**

It is too costly and timely to implement data products for data

**We still need data for AI, which requires a large corpus of data in good quality**

**Therefore, we need to govern how we deal with data**

---

We must capture, organize, and manage the data,  
but how can we ensure that we capture the right data in the right quality?

**Better data is better than better algorithms.**

A. Halevy, P. Norvig and F. Pereira, "**The Unreasonable Effectiveness of Data**," in IEEE Intelligent Systems, vol. 24, no. 2, pp. 8-12, March-April 2009, doi: 10.1109/MIS.2009.36.

Mock, M., Edavettal, S., Langmead, C., & Russell, A. (2023). **AI can help to speed up drug discovery — But only if we give it the right data.** *Nature*, 621(7979), 467-470. <https://doi.org/10.1038/d41586-023-02896-9>

---

## Key questions and steps while bringing Data Governance to an organization

What is the current state of data governance in the organization?

What is critical for business from a performance and compliance perspective?

Engaging with business: understand their needs and our gaps in fulfilling them.

Each lab has its individualities. Understand the different data types, requirements, and usages.

Prioritize data capabilities: focus on the foundation rather than using the data *(more about this later in this presentation)*

Simplify the data capability map.

# Data Strategy Development and Execution

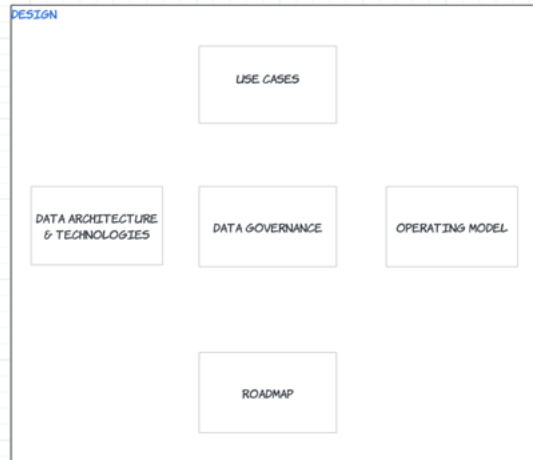
It is a big topic, and I want to bring in these two main points:

The requirements and needs come from the business; work with them:

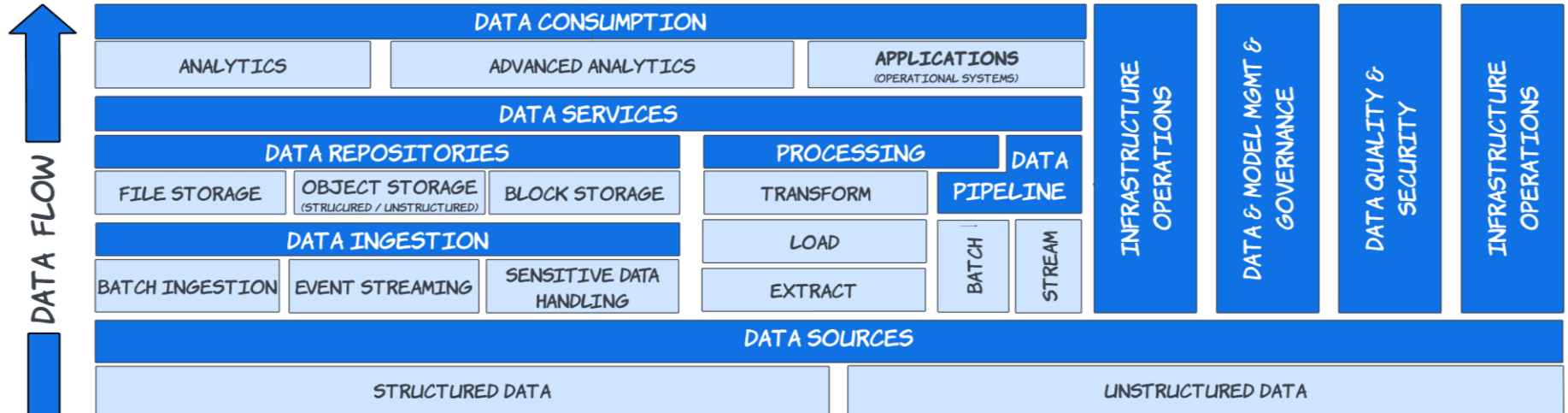
- Listen, give feedback, filter the buzzwords, and focus on the business value.

There are three key steps; do not skip any of them:

Due diligence, Design, and Delivery



## Data governance must follow the data flow



---

## Final remarks

We went through the pharma research value chain,  
Data has to move from the by-product to the main element.

We create knowledge. This knowledge comes from the data.

### **Better data, better knowledge, and higher ROI in the drug research**

There is no one-size-fits-all solution for all different type of data in research

Plan and work with scientists, IT, and Data Specialists to determine what must be done and how.

# Questions?



Other topics that I'd love to engage with you all:

- How to implement the data governance & strategy
- Data Products
- Systems and Architecture
- Dogs and Cheese!

Felipe Albrecht, PhD  
*pRED Data & Analytics, Data & Governance*  
*felipe.albrecht@roche.com*



*Doing now what patients need  
next*