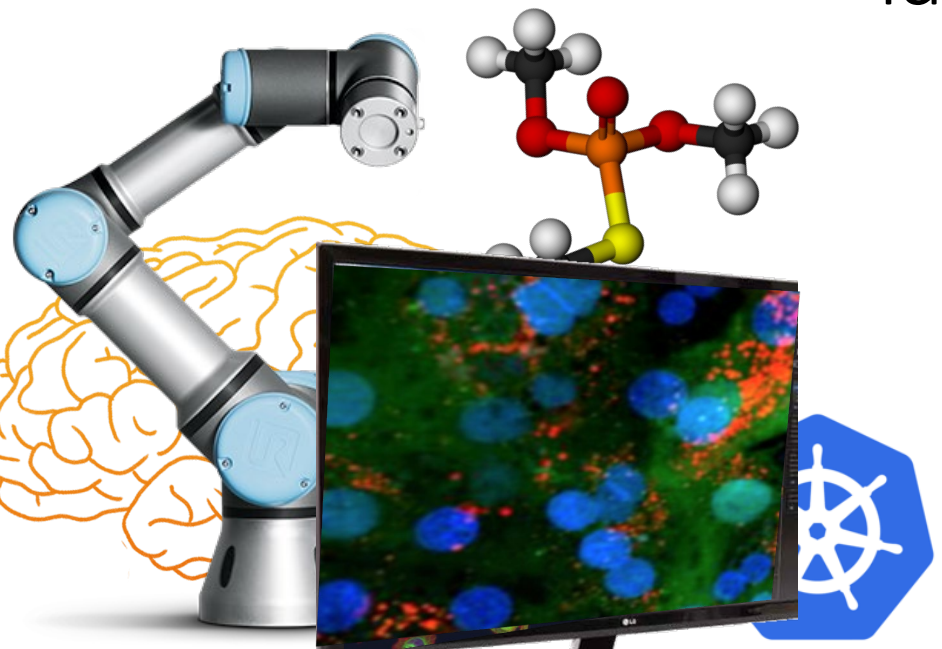
UPPSALA UNIVERSITET

pharmb.io

Automating the process of continuously prioritising data, updating and deploying AI models in a robotised lab for drug discovery



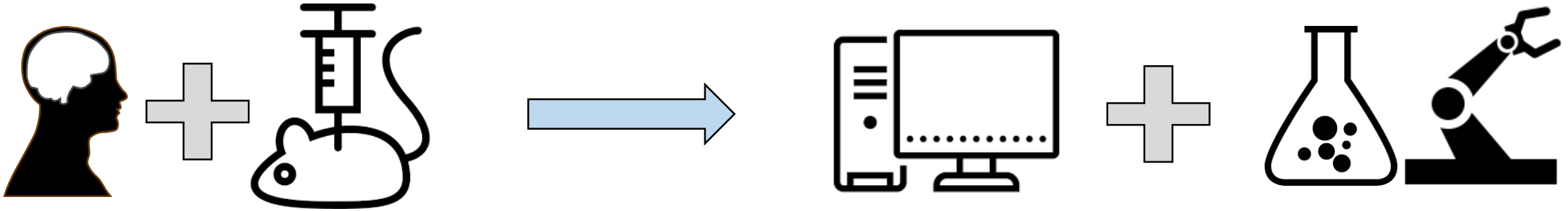
Ola Spjuth <ola.spjuth@farmbio.uu.se>

Department of Pharmaceutical Biosciences, Uppsala University

www.pharmb.io

UPPSALA UNIVERSITET

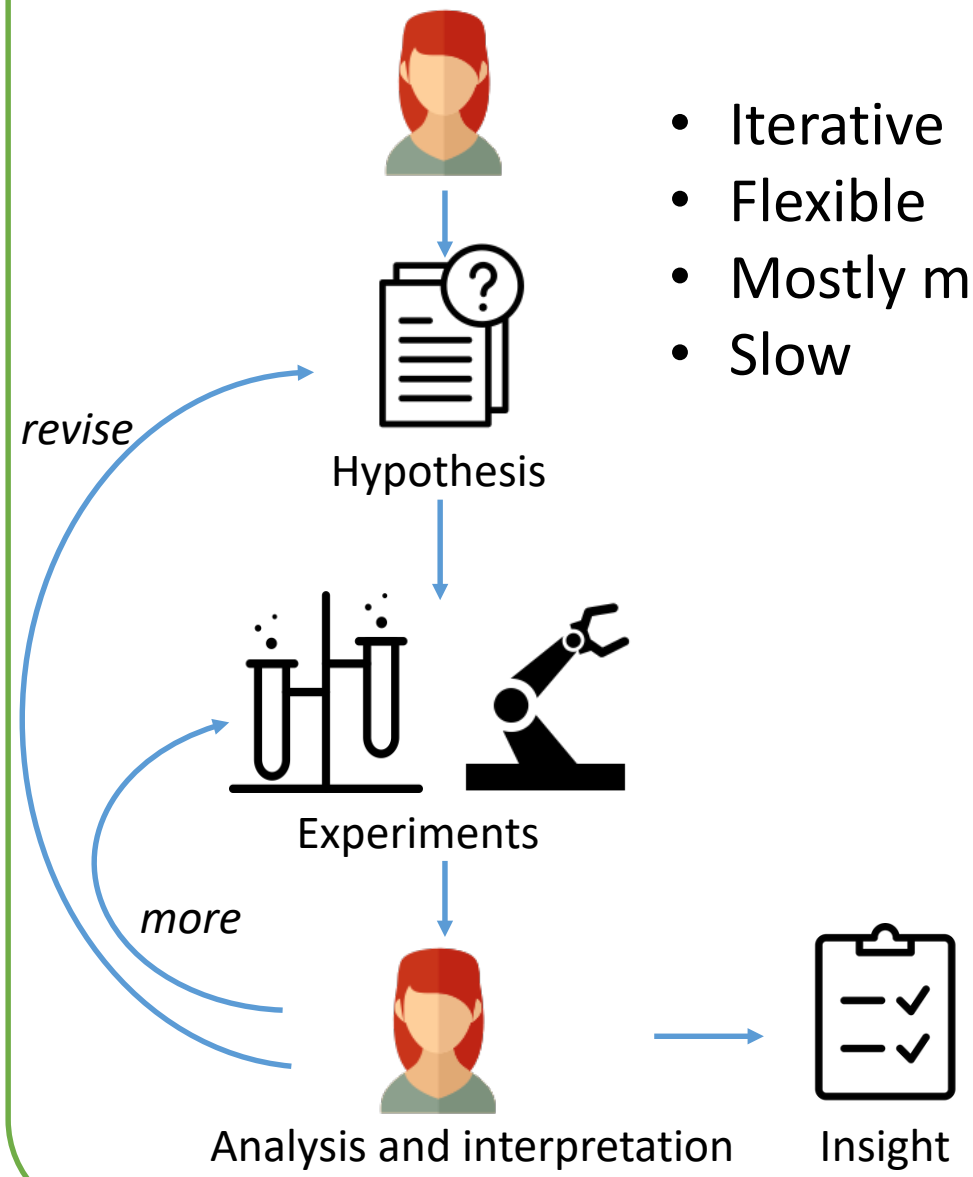
pharmb.io



Objective: Accelerate drug discovery using AI and intelligent design of experiments.

- Predict safety concerns (fail early)
- Explain drug mechanisms
- Screen for new drugs

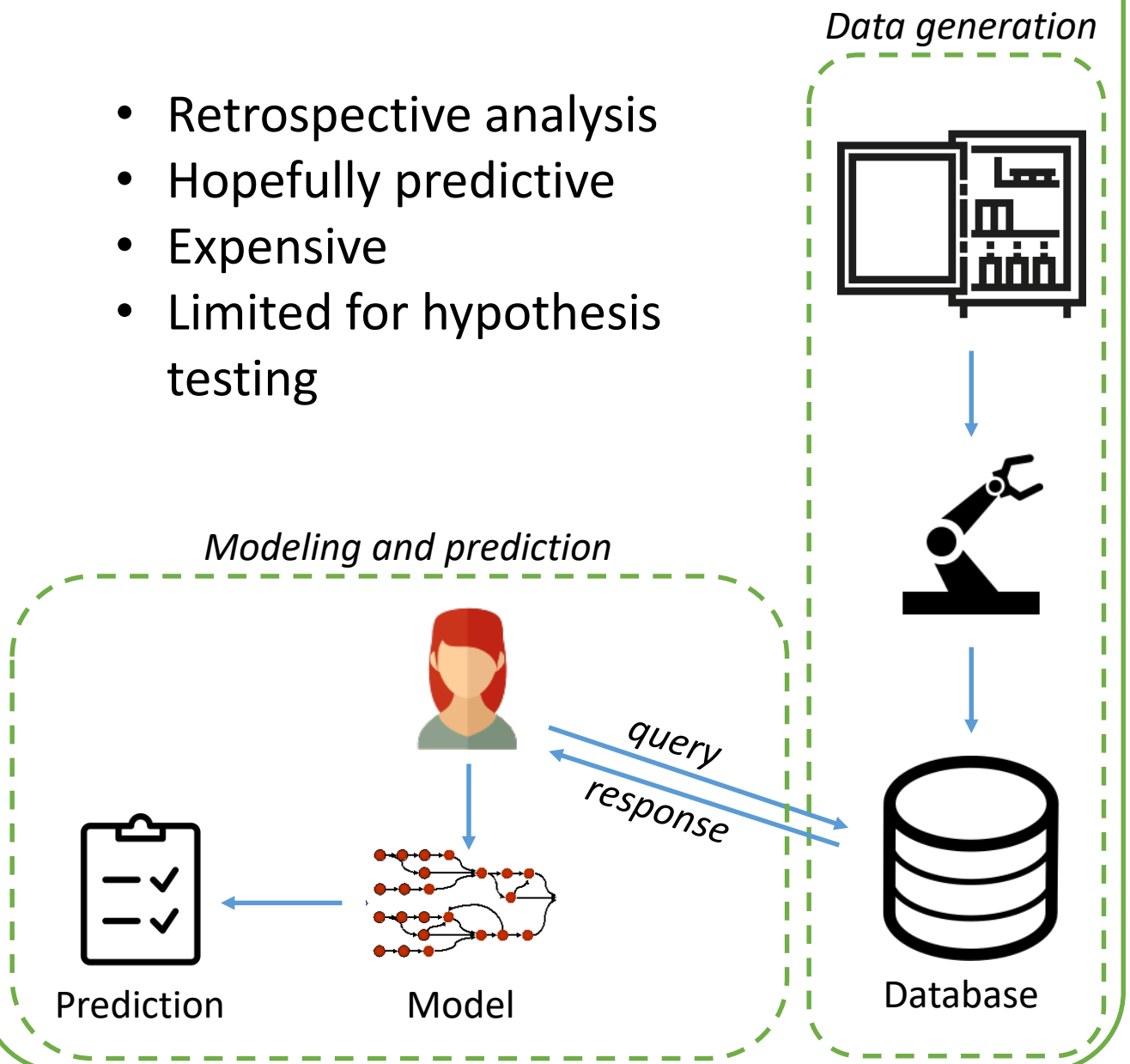
Traditional hypothesis testing



- Iterative
- Flexible
- Mostly manual
- Slow

Predictive modeling

- Retrospective analysis
- Hopefully predictive
- Expensive
- Limited for hypothesis testing



Data-driven science

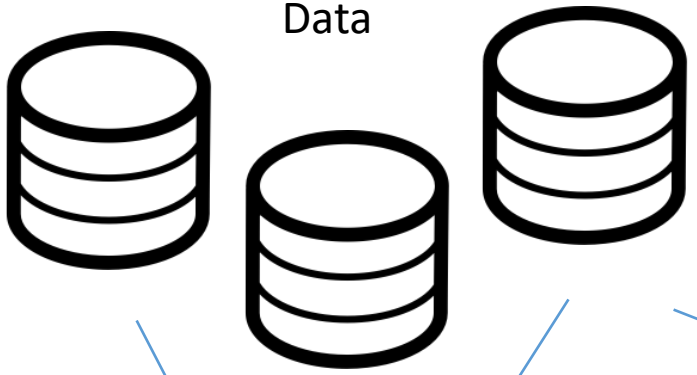
Hypothesis

test

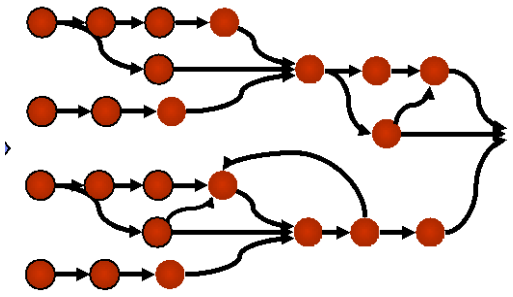
generate

Hypothesis

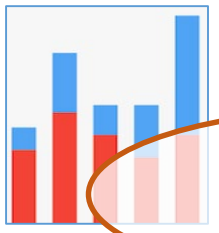
Data



Hard problem!



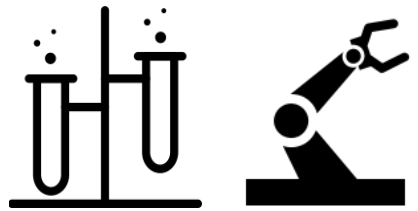
Models



Evaluation

Poor accuracy?

Prediction/
Insight

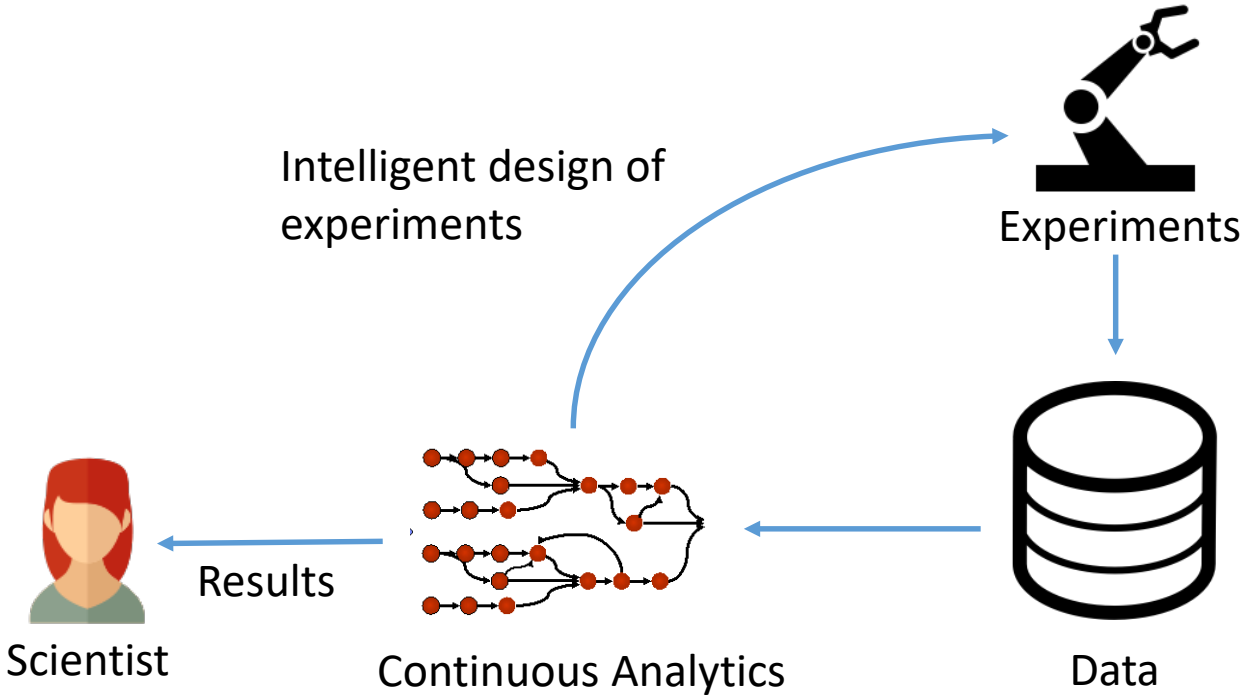


Generate new data

UPPSALA UNIVERSITET

pharmb.io

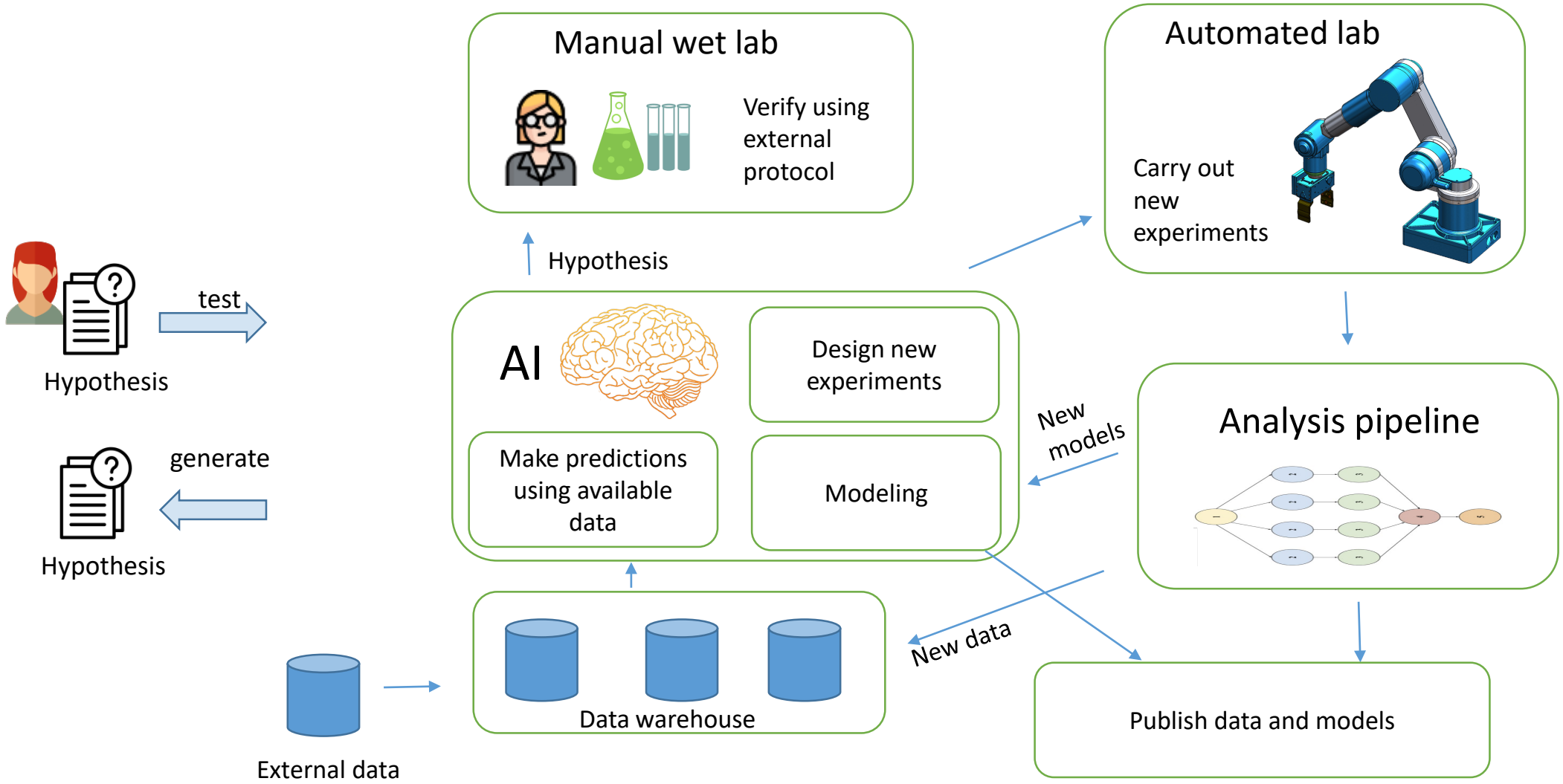
Intelligent experimentation



- What experiments should we do and how?
- Can we reduce search space?
- How store only interesting data?
- Can we replace experiments with predictions?

Informatics system

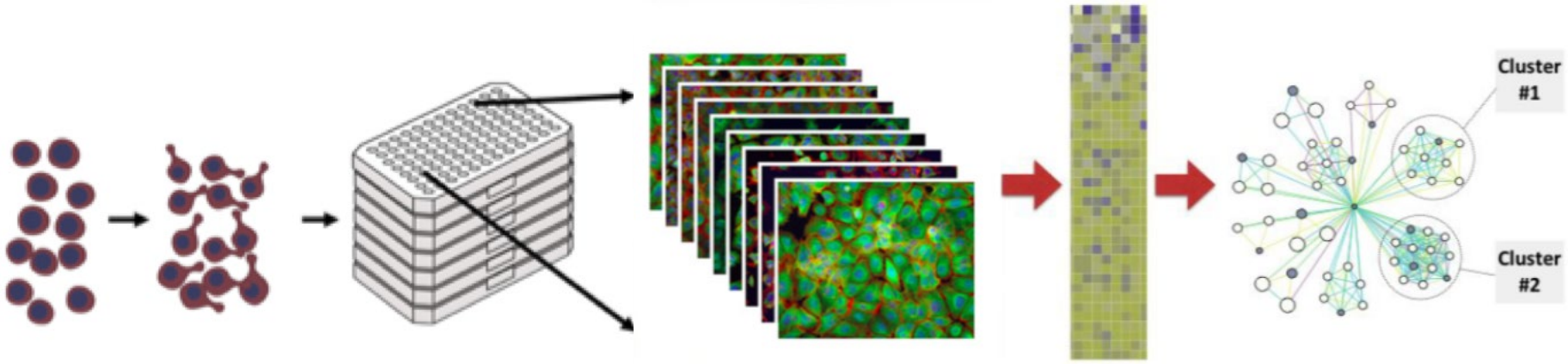
Vision: Intelligent systems for assessing drug leads



UPPSALA UNIVERSITET

pharmb.io

High-content cell profiling



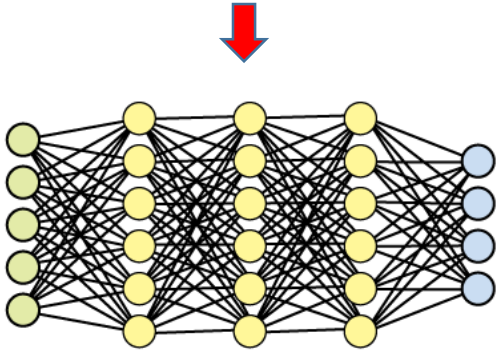
Genetic or chemical perturbations

Experiments in multi-well plates

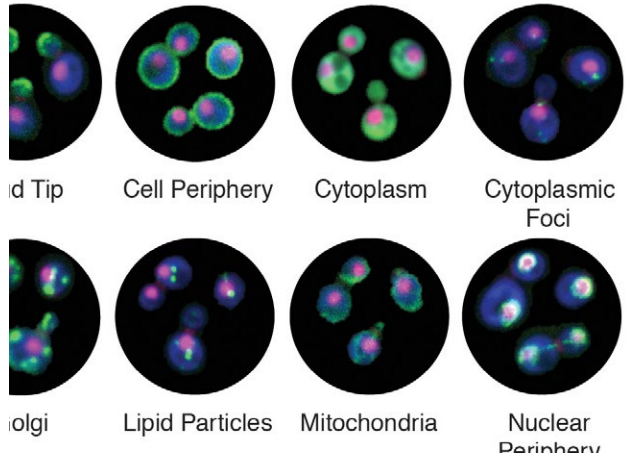
Imaging

Features

Hypotheses



Convolutional Neural Network



Predictions

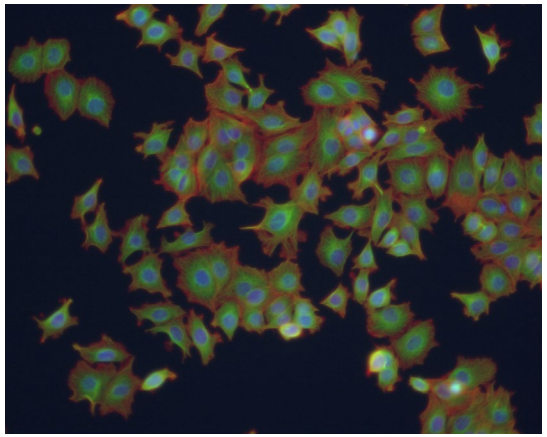
Bray et al. (2016). “Cell Painting, a High-Content Image-Based Assay for Morphological Profiling Using Multiplexed Fluorescent Dyes.” Nature Protocols 11 (9): 1757–74.

UPPSALA UNIVERSITET

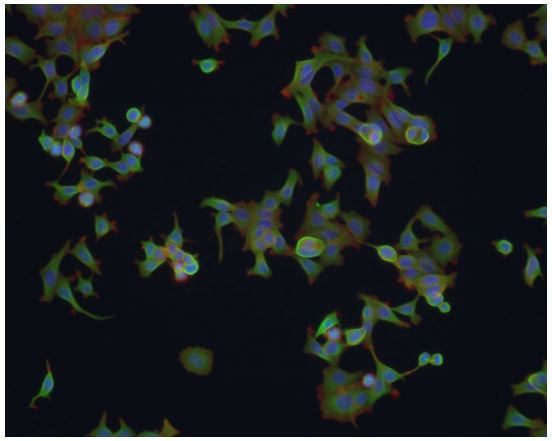
Classify images into biological mechanisms

pharmb.io

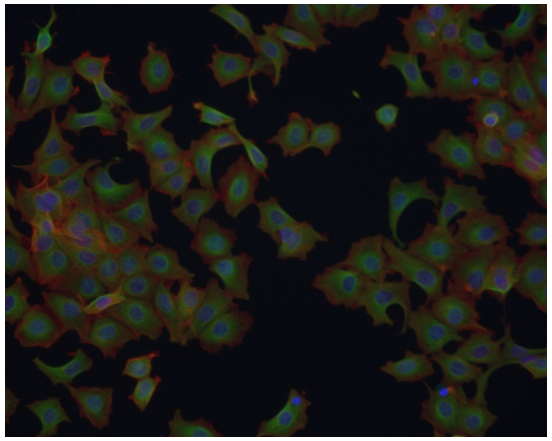
Protein degradation



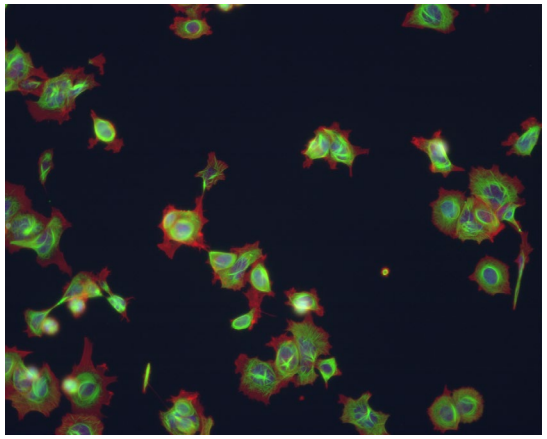
Cholesterol-lowering



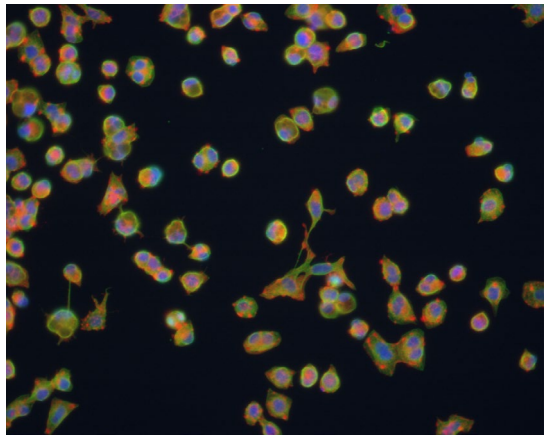
DNA replication



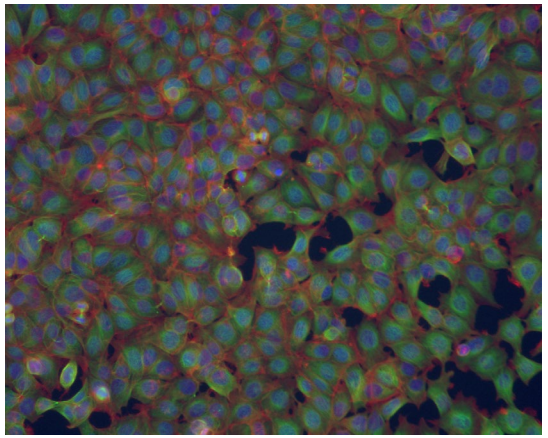
Microtubule stabilizer



Actin disruptor



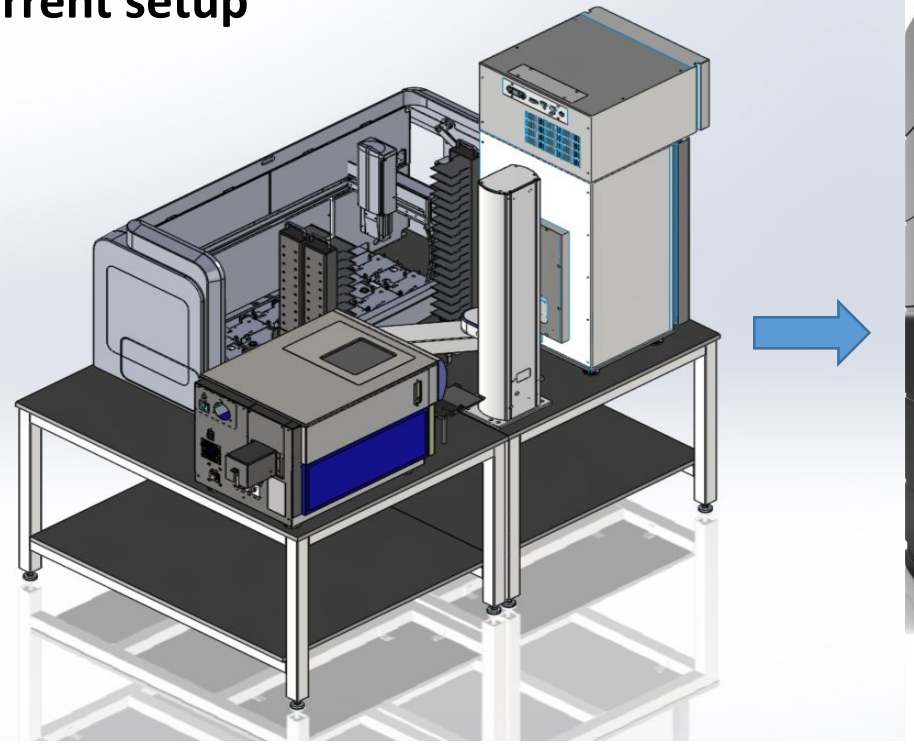
Kinase inhibitor

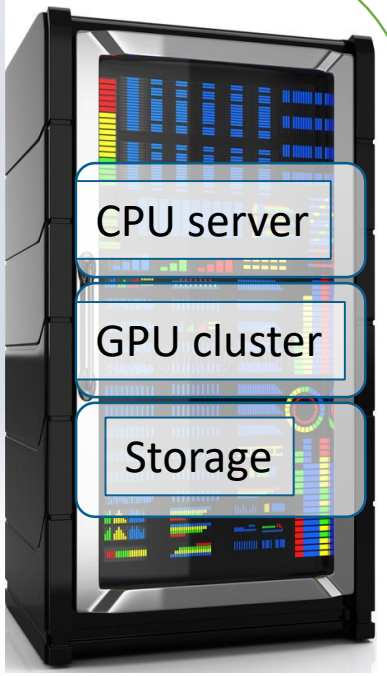


Kensert A, Harrison PJ, Spjuth O. Transfer learning with deep convolutional neural network for classifying cellular morphological changes. *SLAS DISCOVERY: Advancing Life Sciences R&D. (2019)*

Automated cell-based experiments

Current setup





Informatics system

Automated plate handling

Open source robotized cell lab

Automate more protocols

UPPSALA UNIVERSITET

Chemical DB

chemicals

link

Robotized lab

images

cell profiles

OMERO

CPU/GPU/HPC cloud

Models

Notebooks

External user

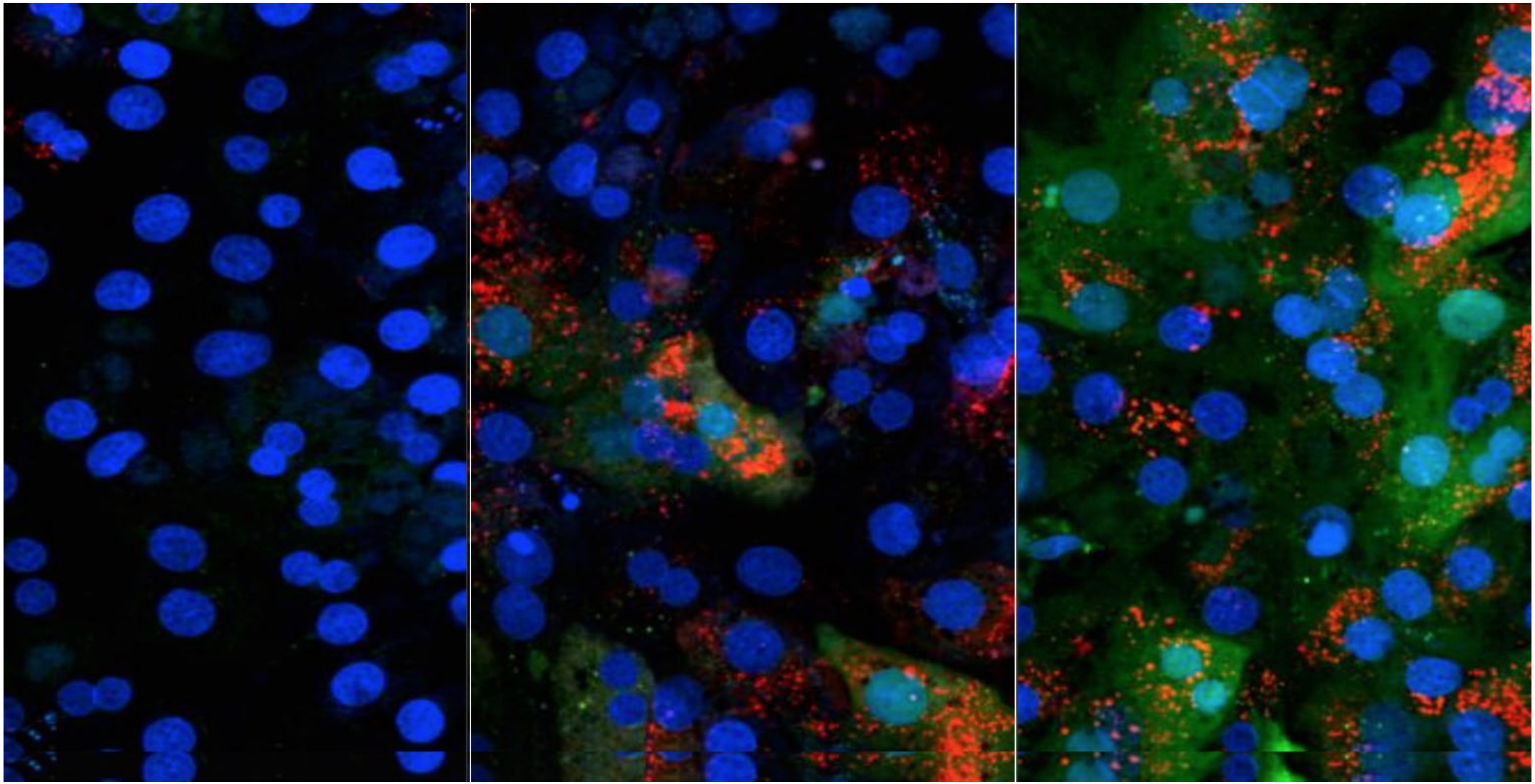
Services

Data

Models

Public services

- Fluorescent LNPs (lipids)
- Fluorescent Cargo (mRNA)
- Fluorescent Product (protein)



No LNPs

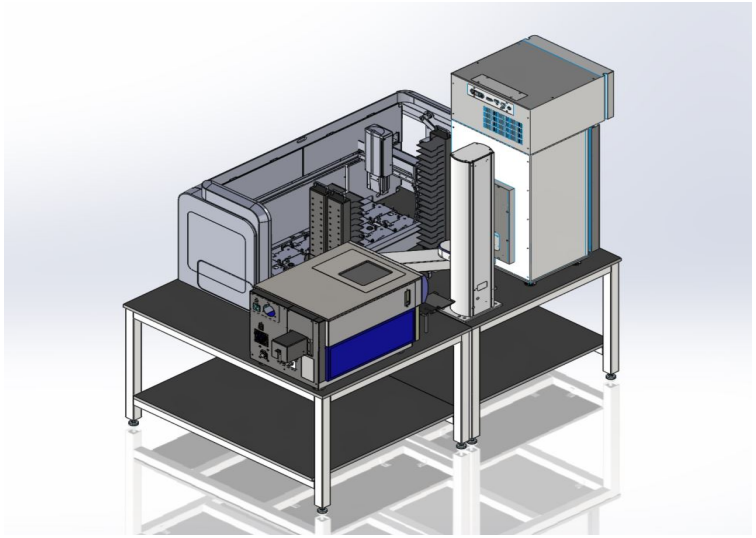
Partial LNP uptake

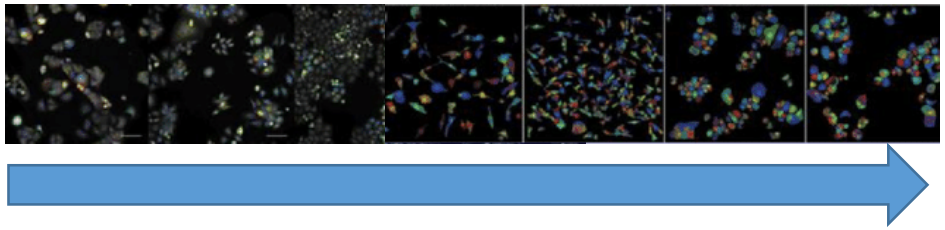
LNP uptake and mRNA decoding

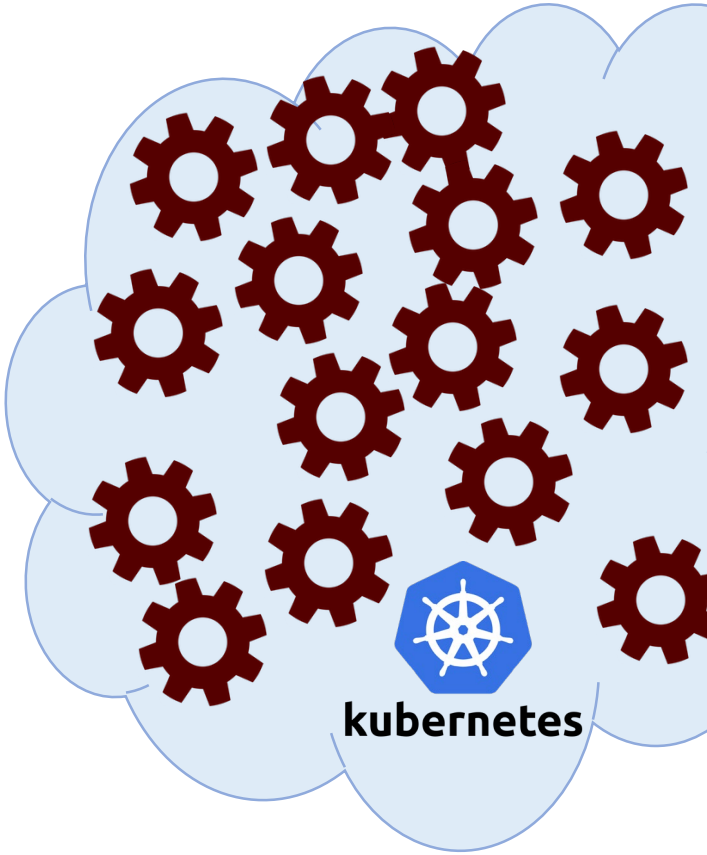
UPPSALA UNIVERSITET

pharmb.io

Dealing with large scale data

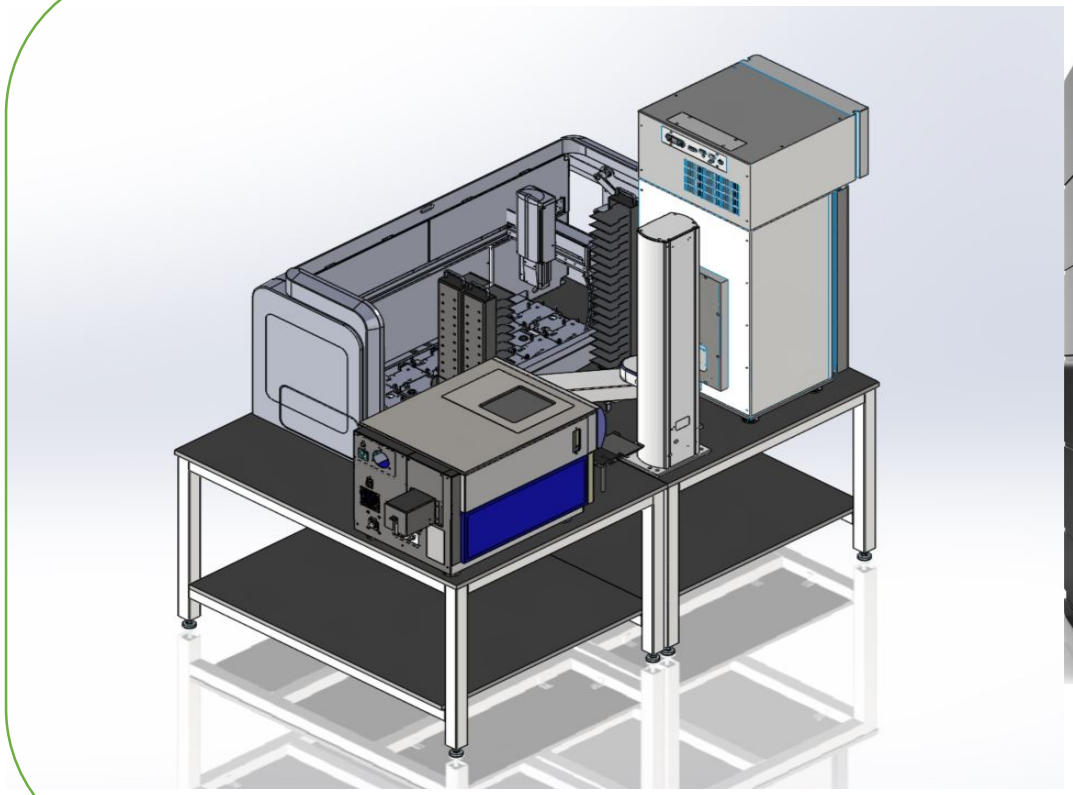


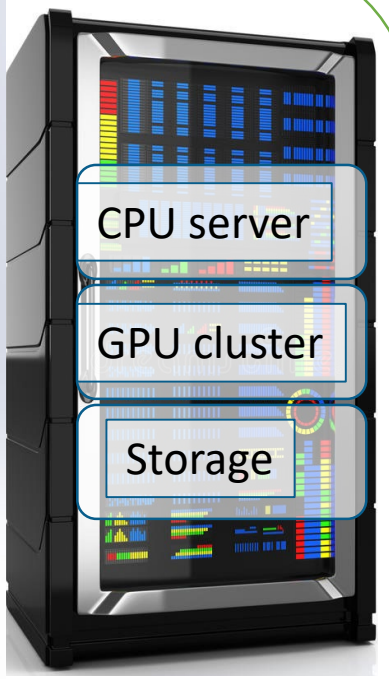




- High volume, high velocity
- Continuously process data, train models, serve models
- Embrace scalable virtual infrastructures (cloud) and microservices (containers)
- Intelligently prioritize what to store

Designing a flexible and scalable informatics system





Spark

kubernetes

TensorFlow

HPC

Cloud

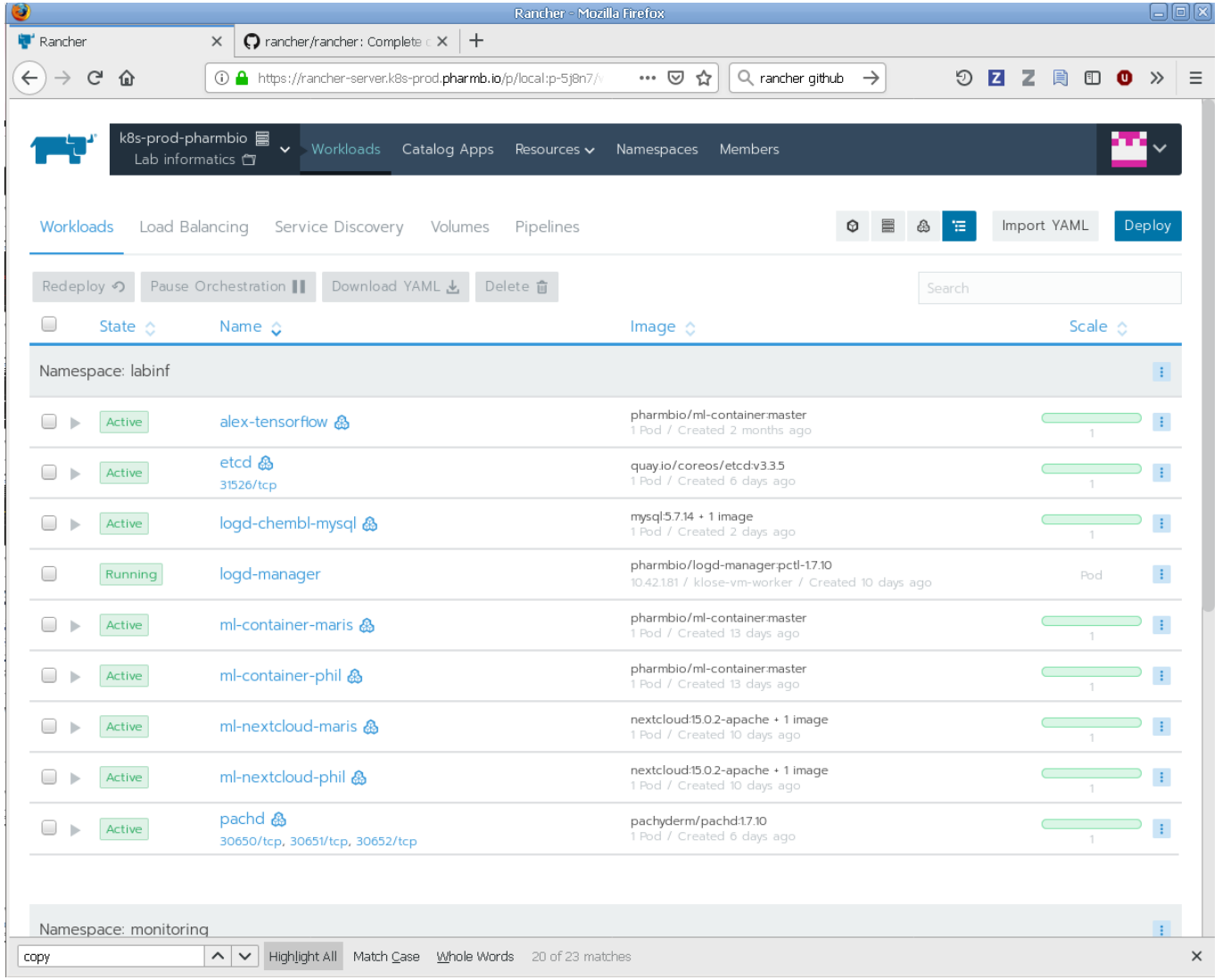
UPPSALA UNIVERSITET

pharmb.io

Multi-cluster Kubernetes management

RANCHER

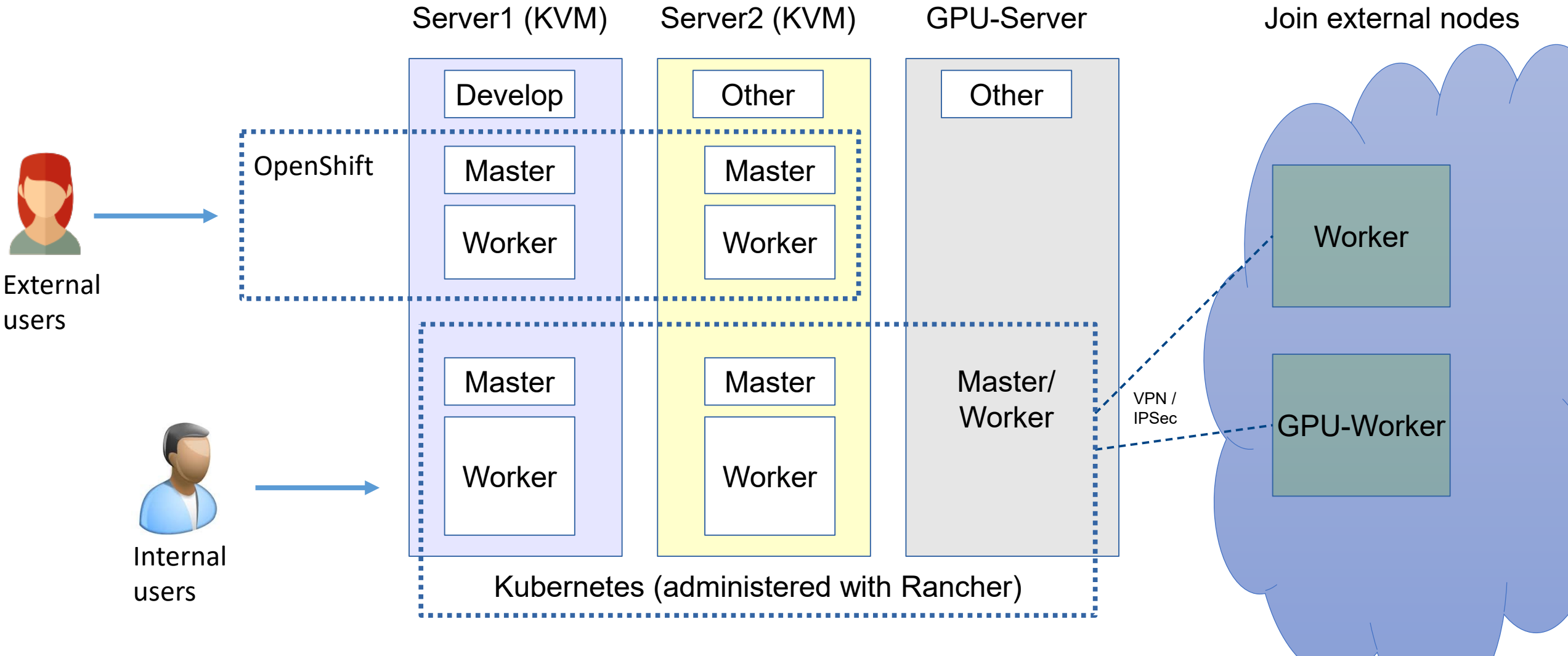
- Kubernetes Setup and admin in the cluster
- High availability. Kubernetes extras such as Networking, Helm, Nginx-load-balancer with tested and matched versions.



UPPSALA UNIVERSITET

pharmb.io

Hybrid infrastructure

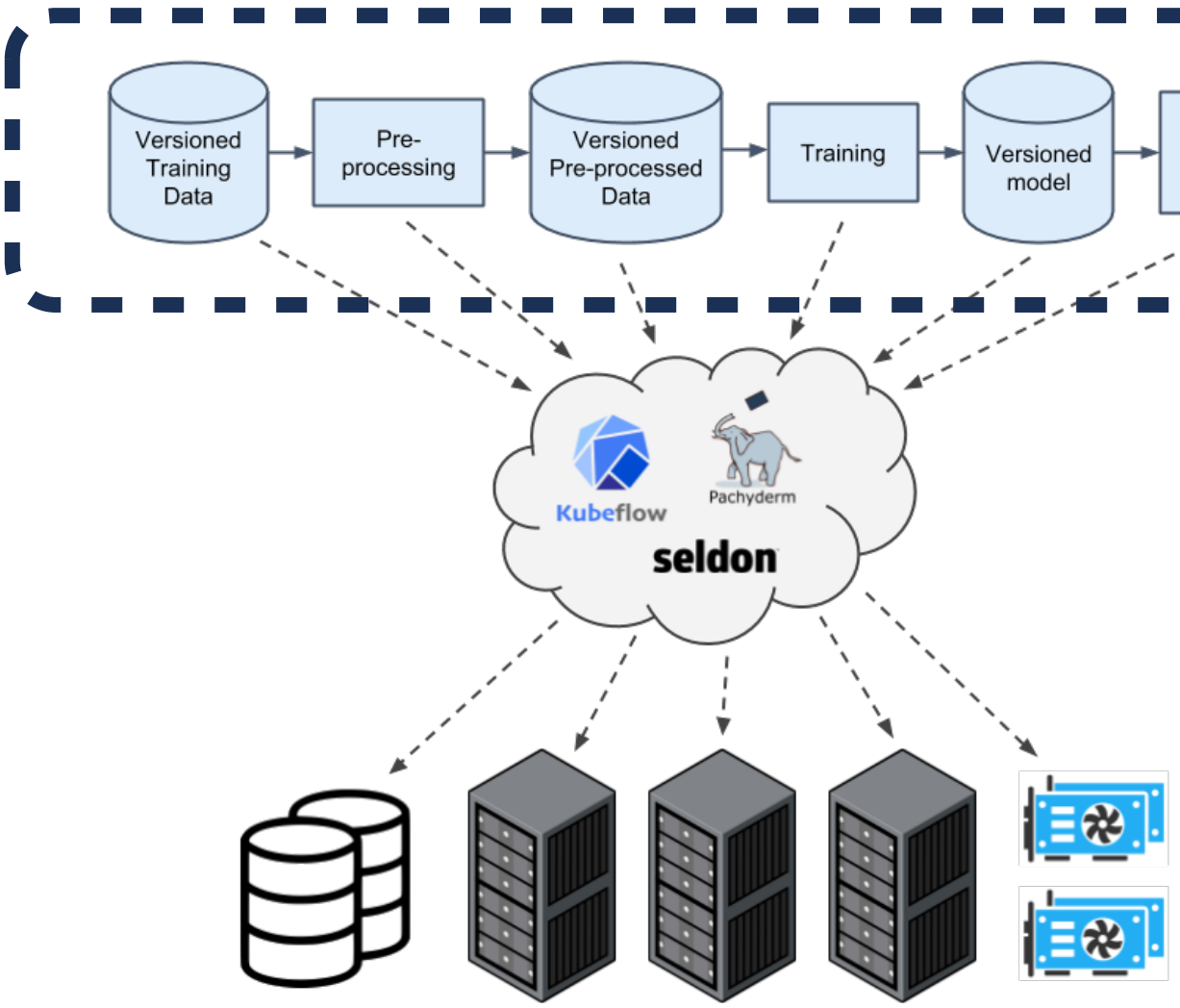


UPPSALA UNIVERSITET

Logging and monitoring

pharmb.io

AI/Machine Learning life cycle



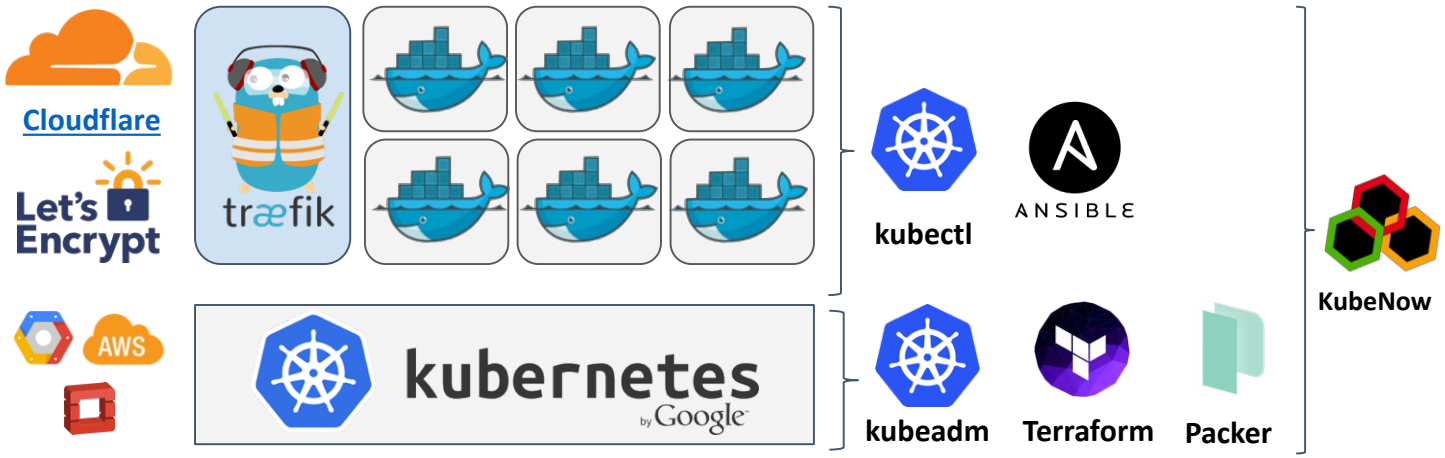
- Structured processing steps
- Governance
- Versioning
- Streamline analysis on high-performance e-infrastructures
- Support reproducible data analysis
- Enable large-scale data analysis

Tools we develop and use

https://github.com/kubenow

KubeNow

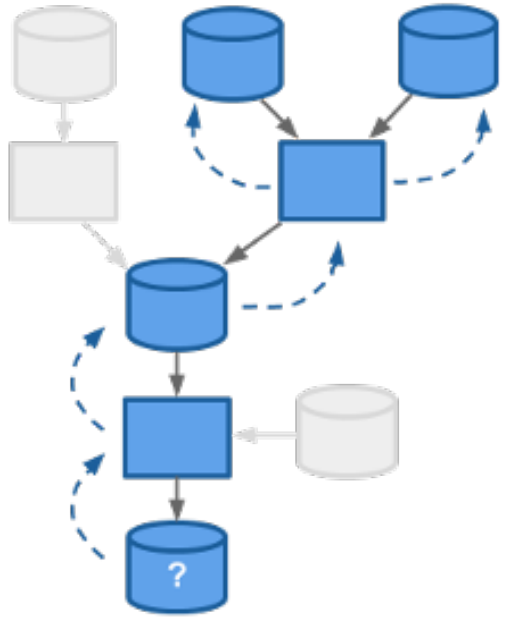
- Easy deployment of virtual infrastructures on IaaS
- Containerize tools, orchestrate microservices with workflow systems on top of Kubernetes and OpenShift



Capuccini et al. **On-Demand Virtual Research Environments using Microservices.** https://arxiv.org/abs/1805.06180

https://www.pachyderm.io

- Data pipelining and data versioning layer for Kubernetes



Novella et al. **Container-based bioinformatics with Pachyderm**. *Bioinformatics*. 35, 5, 839-846. (2018). DOI: http://dx.doi.org/10.1093/bioinformatics/bty699

Stiftelsen för Strategisk Forskning

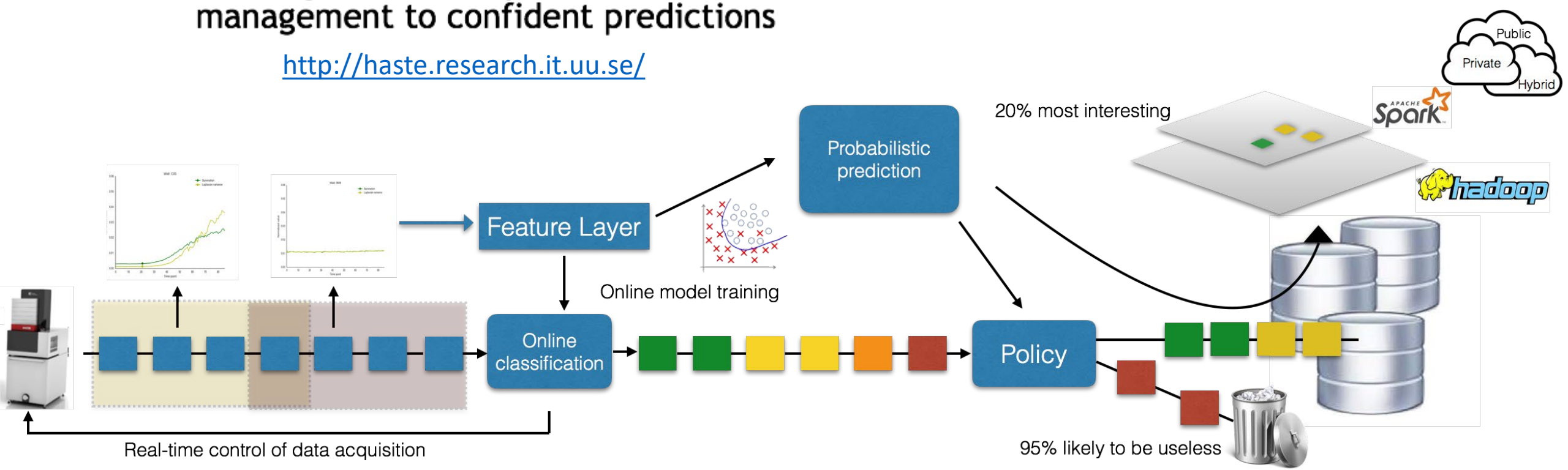
HASTE

AstraZeneca Vironova SciLifeLab eSENCE

HASTE: Hierachical Analysis of Spatial and Temporal data

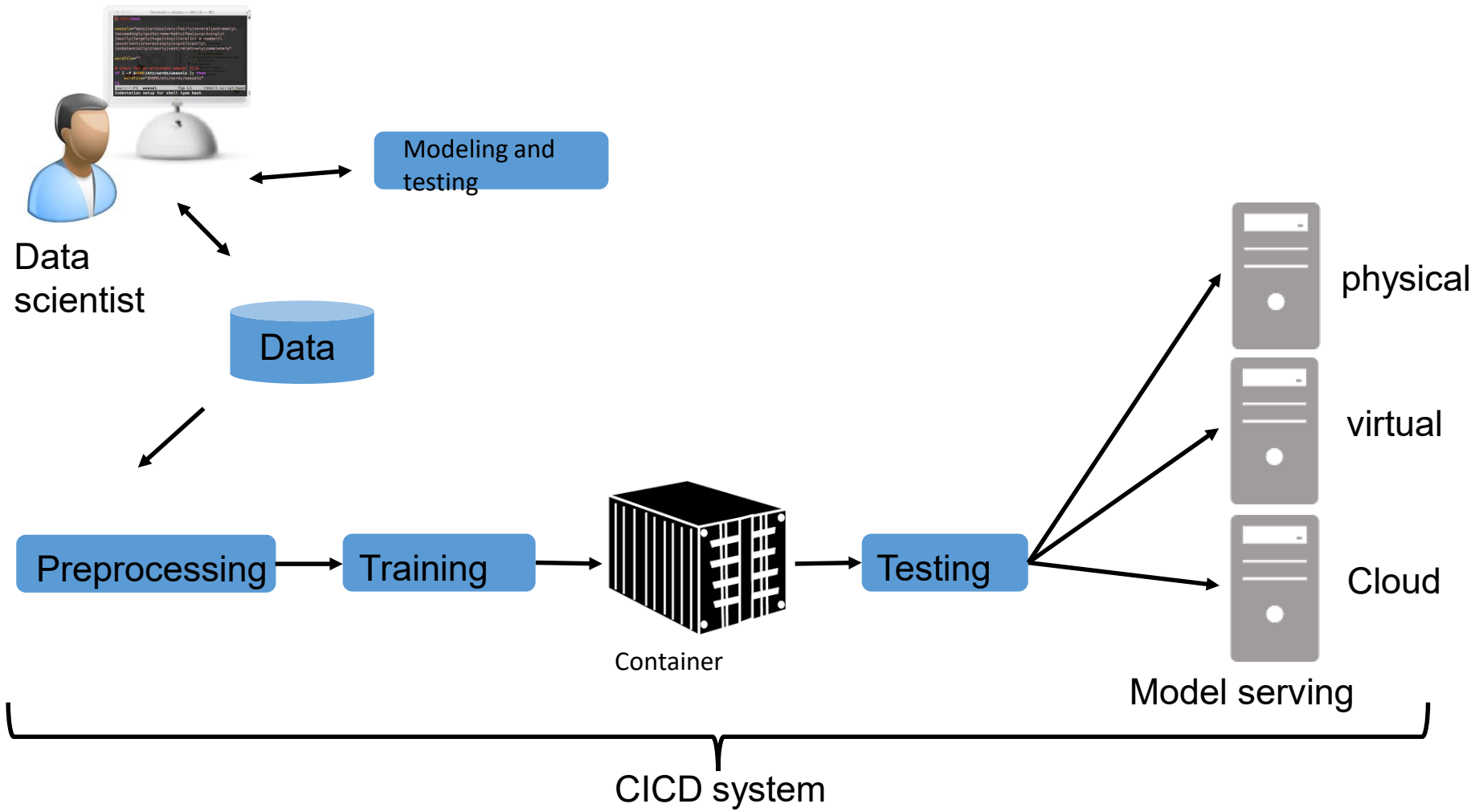
- from intelligent data acquisition via smart data-management to confident predictions

http://haste.research.it.uu.se/



Continuous analytics

- Empower data scientists and automate from data to deployed models



UPPSALA UNIVERSITET

pharmb.io

FAIR data and services

FAIR[1]

- **F**indable
  - Semantic service discovery (JSON-LD)
- **A**ccessible
  - Web UI
  - Programmatic API (OpenAPI)
- **I**nteroperable
  - Open API
  - Semantic annotations (JSON-LD)
- **R**eproducible
  - Published scientific workflow (Pachyderm)

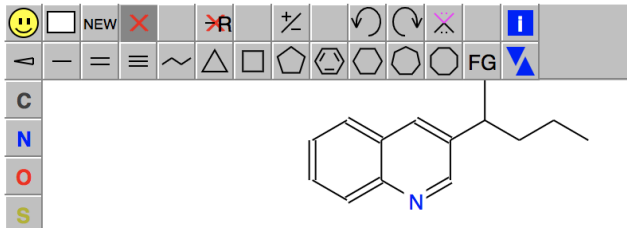
Services

- Portable (Containers)
- Resilient (Kubernetes)
- Scalable (Kubernetes and cloud computing)

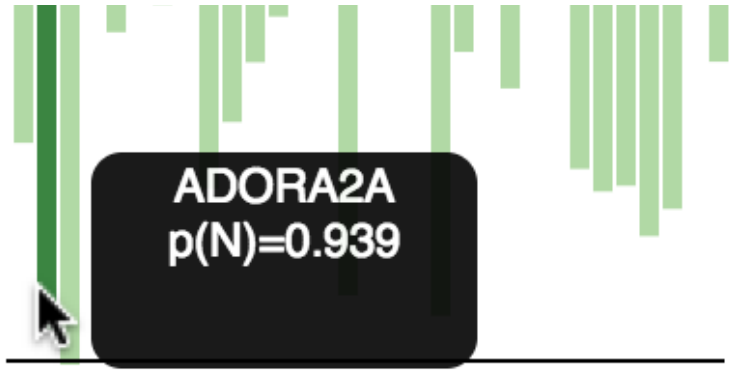
[1] Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3 (2016).

Examples of what we serve

Target (safety) profiles

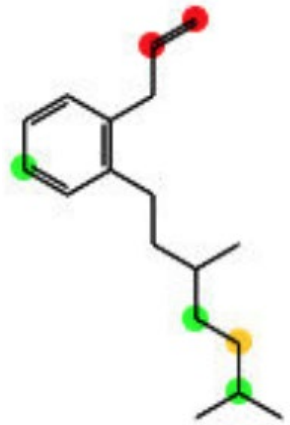


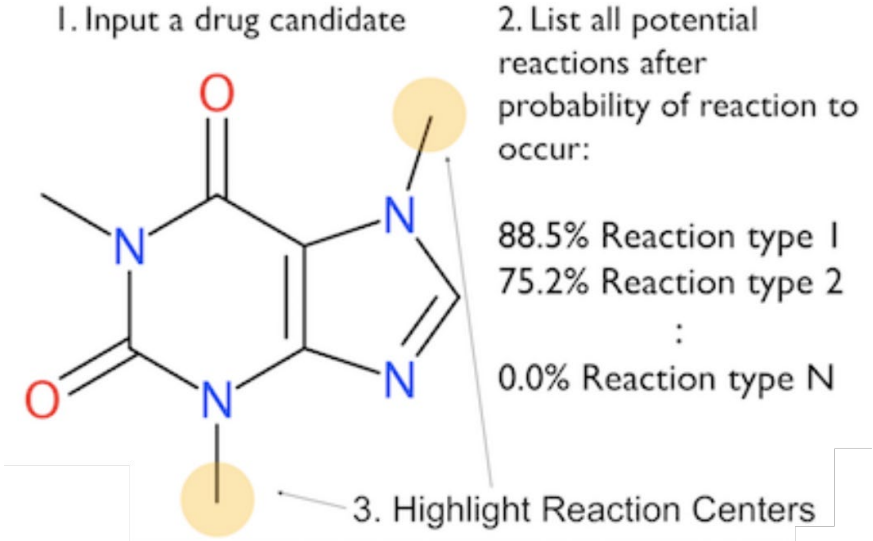
ACHE	ADORA2A	ADRB1	ADRB2	AR	AVPR1A	CCKAR	CHRM1	CHRM2
CHRM3	CNR1	CNR2	DRD1	DRD2	EDNRA	HTR1A	HTR2A	KCNH2
LCK	MAOA	NR3C1	OPRD1	OPRK1	OPRM1	PDE3A	PTGS1	PTGS2
SCN5A	SLC6A2	SLC6A3	SLC6A4					



http://ptp.service.pharmb.io/

Site-of-metabolism and reaction types





https://metpred.service.pharmb.io/draw/

UPPSALA UNIVERSITET

Chemical DB

chemicals

link

Robotized lab

images

cell profiles

OMERO

CPU/GPU/HPC cloud

Models

Notebooks

External user

Services

Data

Models

Public services

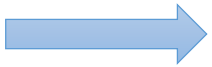
UPPSALA UNIVERSITET

pharmb.io

DevOps and components

- **IaaC**: Infrastructure is portable, testable and simplifies maintenance
- **Containers**: Components become portable and scalable
- **Kubernetes**: Resilient container orchestration over multiple nodes
  - logging and monitoring - profiling
  - Hybrid cloud - elasticity
- **Pachyderm**: Workflows of containers in Kubernetes with data versioning
- **CICD**: Streamline development and testing

→ **DevOps**: Software Developers, Data Engineers and Data Scientists working together in same infrastructure



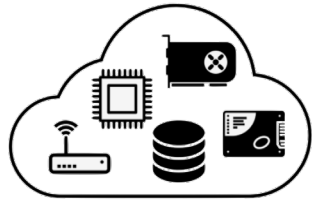
scaleout SYSTEMS

UPPSALA UNIVERSITET

pharmb.io

Implications: Continuous Analytics

- We can handle the continuous data processing from instruments with robust, resilient data pipelines
- We can continuously re-train models as data is updated
- We can continuously publish data and models

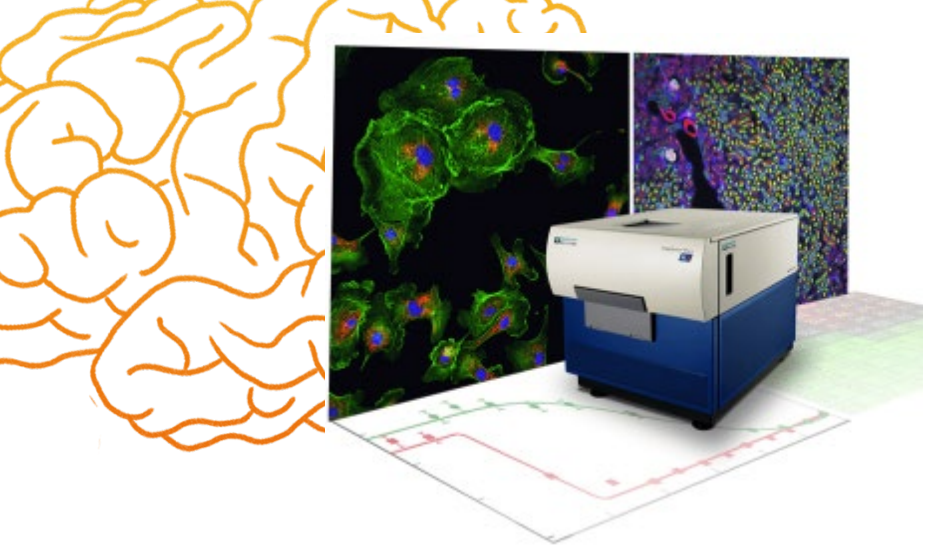


- Agile research group of different competencies
  - Scientists get access to necessary infrastructure
  - DevOps roles, no dedicated sysadmin / tool devel / scientist roles

Spark

kubernetes

TensorFlow



Funding:

Horizon 2020
European Union Funding
for Research & Innovation

FORMAS
ETT FORSKNINGSRÅD FÖR HÅLLBAR UTVECKLING
A SWEDISH RESEARCH COUNCIL FOR SUSTAINABLE DEVELOPMENT

Stiftelsen för Strategisk Forskning

SeRC

eSSENCE
THE e-SCIENCE COLLABORATION

VINNOVA

elixir

eurostars™

NeIC
NORDIC E-INFRASTRUCTURE COLLABORATION

- Thank you -

pharmb.io

Research group website: http://pharmb.io