CERTAINTY DISCOVERY

CERTARA | Chemaxon

Integrating Artificial Intelligence into the Drug Discovery Pipeline

Gian Marco Ghiandoni

AstraZeneca

Associate Principal AI Engineer

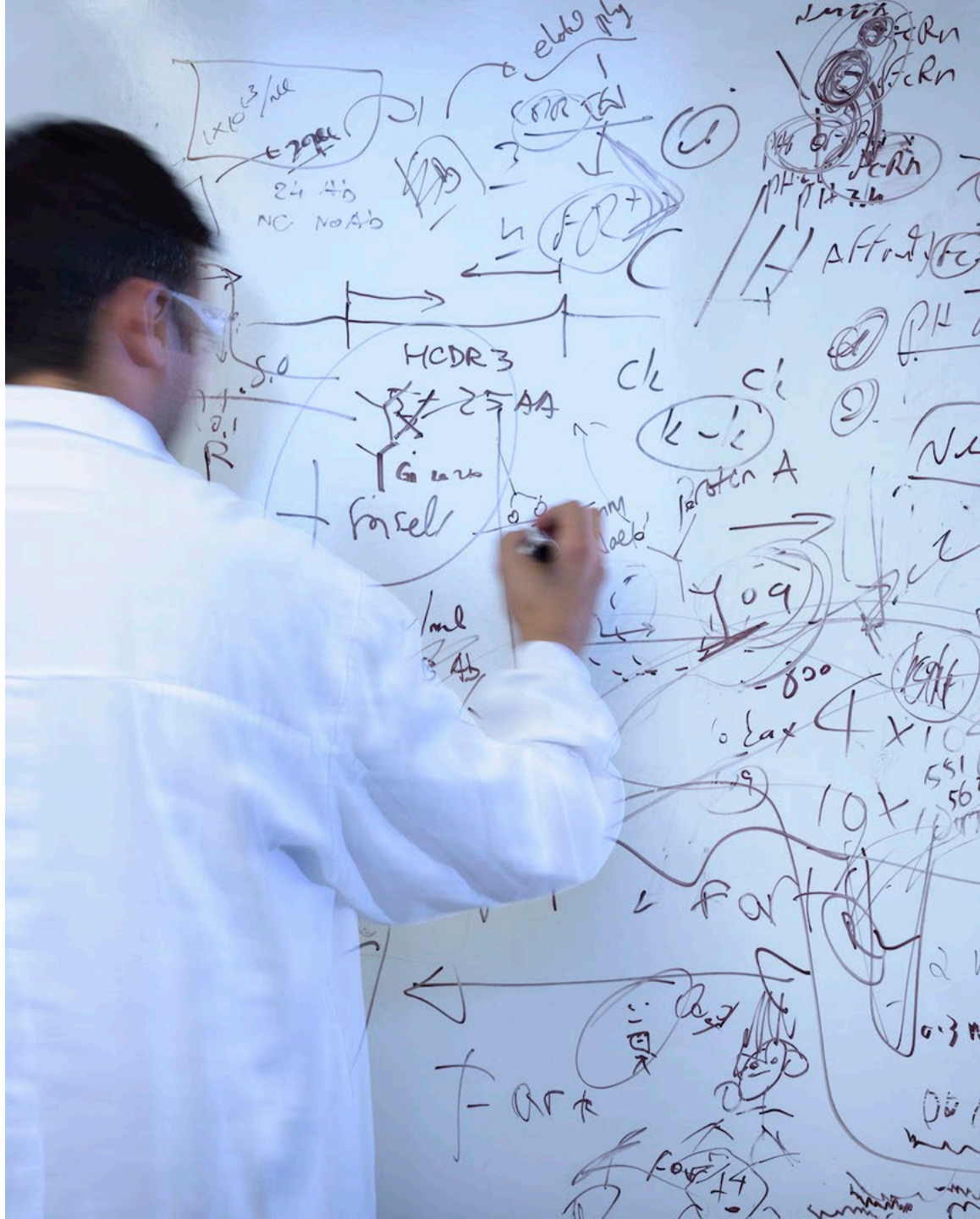
Scaled AI and Chemistry (SAIC) Tech Lead

AstraZeneca

Integrating AI into the Drug Discovery Pipeline

Gian Marco Ghiandoni

05/11/2025
Certainty Discovery
Design Offices Frankfurt Wiesenhüttenplatz, Frankfurt, Germany



Myself

Gian Marco Ghiandoni (Giammy)

CheMLOps Lead | Associate Principal Engineer

- 6 years at AstraZeneca R&D IT
- Currently within the Augmented DMTA platform (ADMTA)
- Leading a team of scientific developers and chemistry/ML engineers
- Side-by-side with the business – we create capabilities for the scientists
- Disclaimer - ask Bill if you want to know why we run so many calculations in D360!

Contents:

- Theory vs. Practice, DMTA
- Predictive Insight Platform (PIP)
- PIP-D360 interaction
- Extended PIP functionalities in D360
- Vision

AI in Drug Discovery: Theory vs. Practice

Theory: AI revolution

Practice: There is a lot of work to do…

AI

Data Science

Chemistry

Scaling

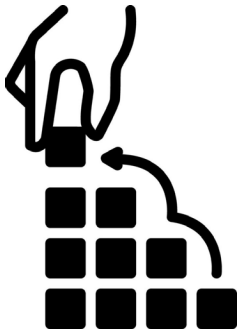
Integrations

Frameworks

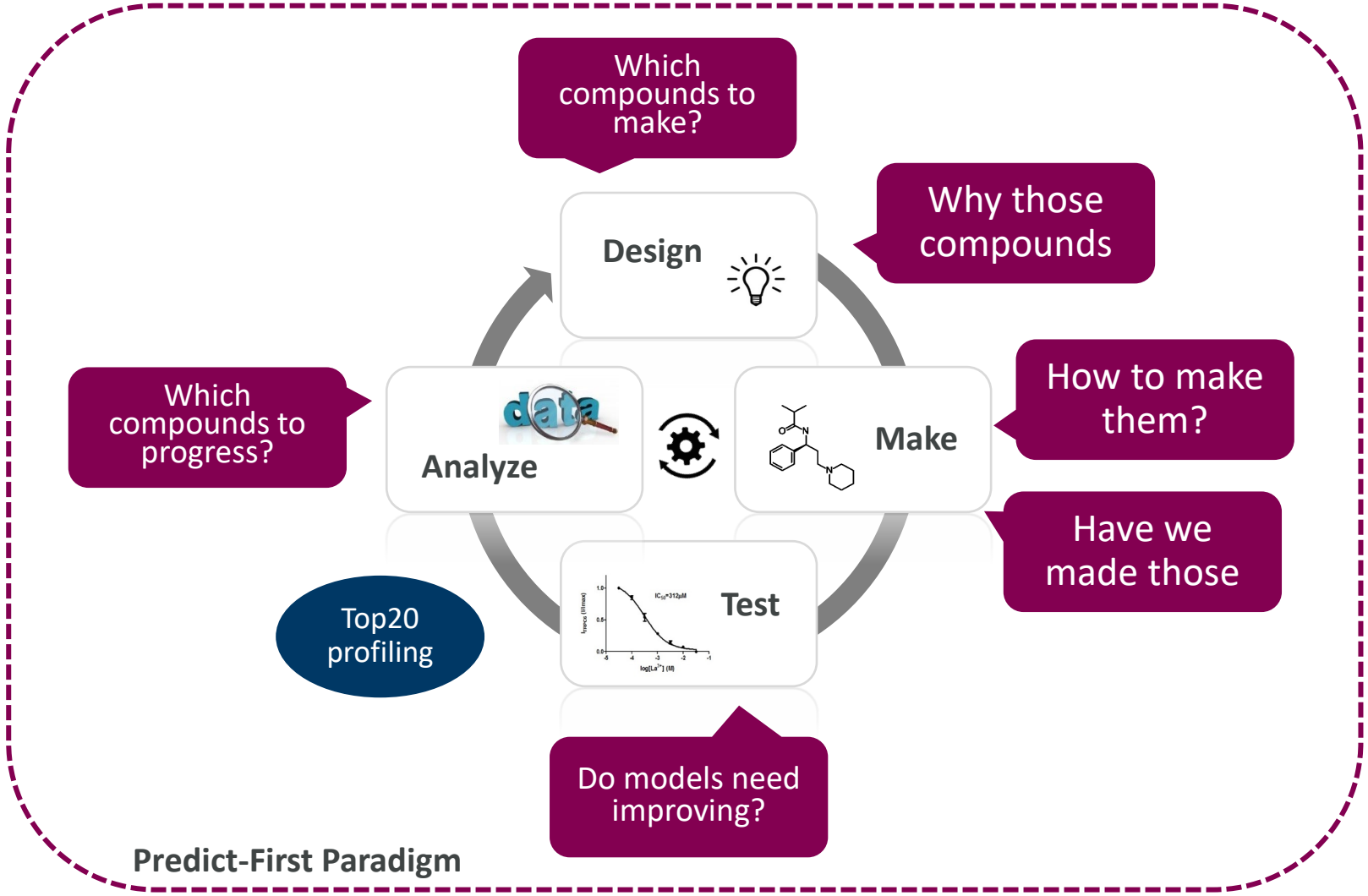
Pipelines

Data

Infrastructure

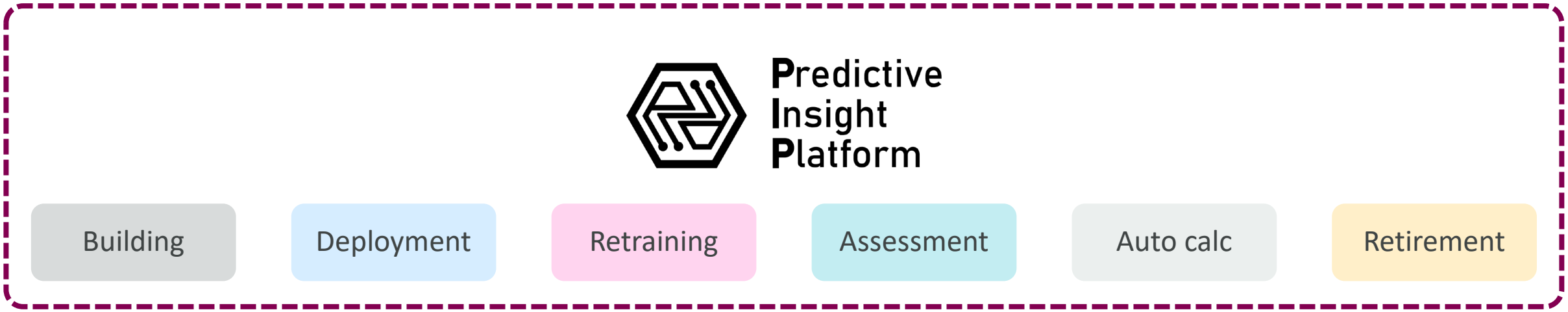


Predictive AI in the DMTA cycle



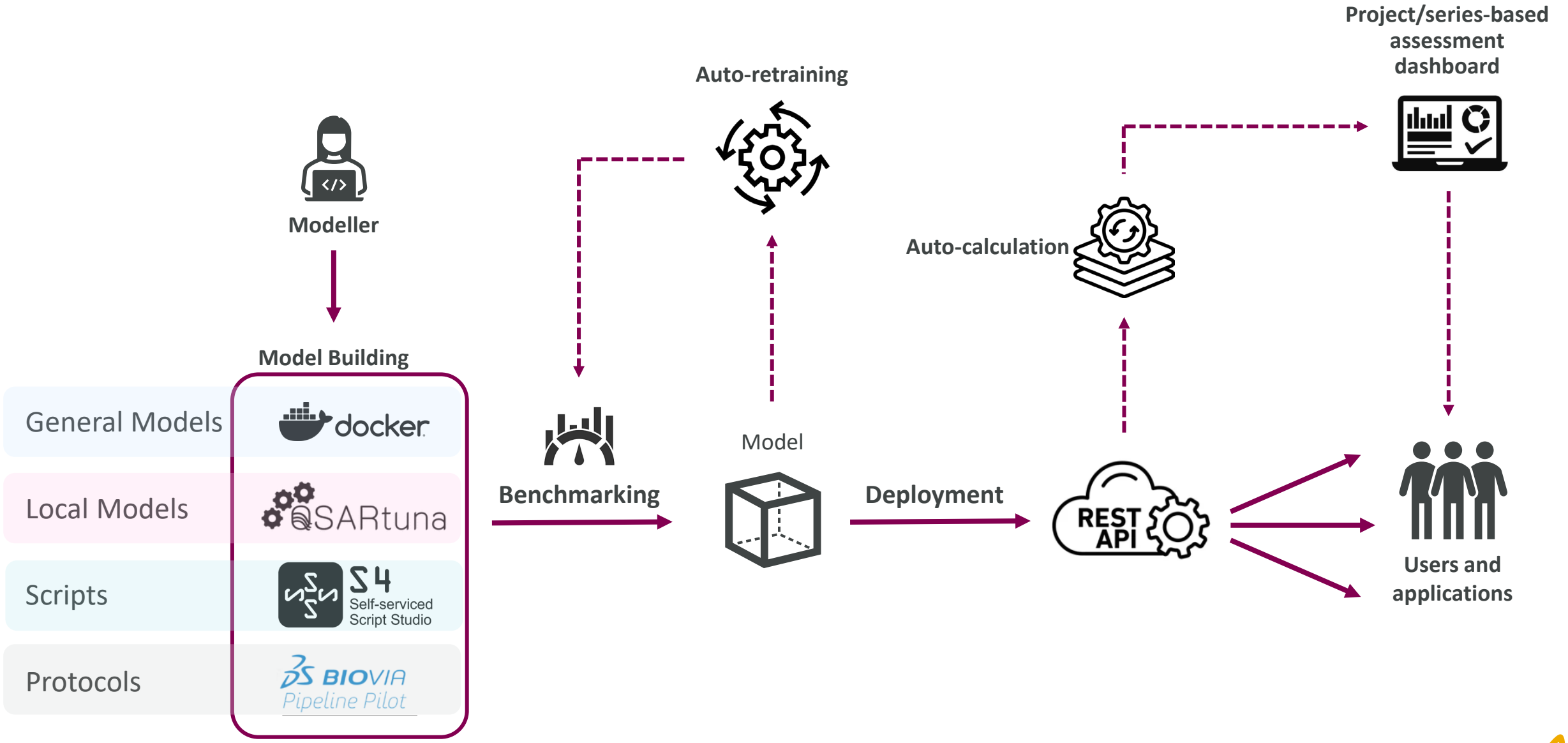
The Predictive Insight Platform (PIP)

- The Predictive Insight Platform (PIP) is a cloud-native platform for AI/ML drug discovery at AZ
- Hosting more than 400 models for small-molecule and new modality property prediction
- Integrated with 15+ chemistry applications
- Critical for operational systems - sample management and compound registration
- Enables “Predict-First” at AZ with almost 500M calculations being run every month



Ghiandoni GM, Rathi PC, et al. Augmenting DMTA using predictive AI modelling at AstraZeneca. Drug Discovery Today. 2024;29(4):103945.

How PIP works



Ghiandoni GM, Rathi PC, et al. Augmenting DMTA using predictive AI modelling at AstraZeneca. Drug Discovery Today. 2024;29(4):103945.

PIP augments all aspects of the discovery pipeline

De novo generation

PhysChem, in vitro ADME, synthetic accessibility

Hit-to-lead

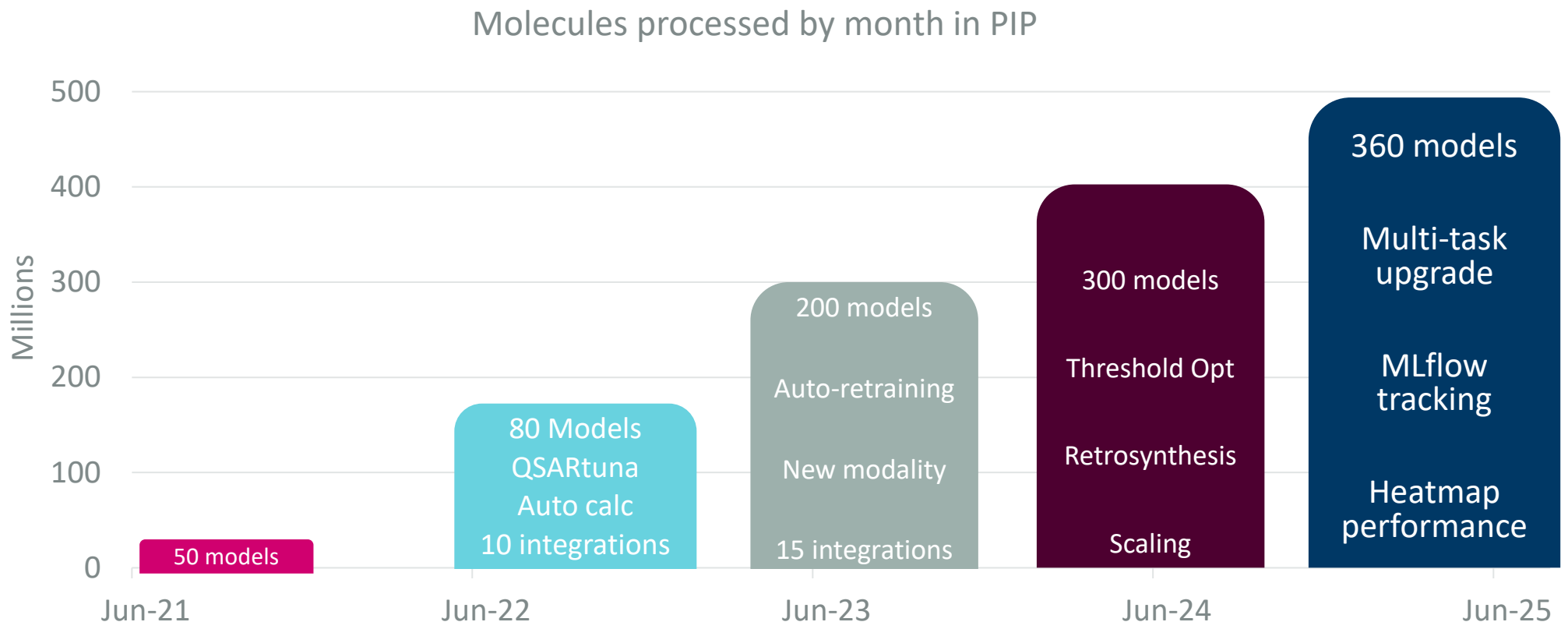
Loose MPO, in vivo ADMET predictions

Lead optimisation

Strict MPO, safety, translation into human

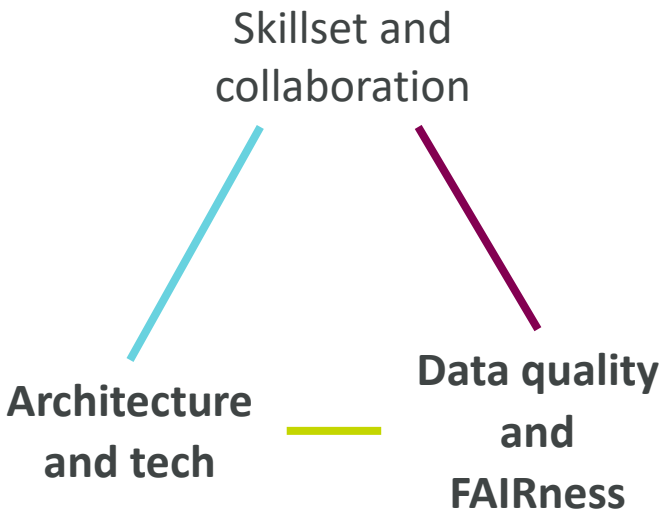
Ghiandoni GM, Rathi PC, et al. Augmenting DMTA using predictive AI modelling at AstraZeneca. Drug Discovery Today. 2024;29(4):103945.

Evolution of PIP



Ghiandoni GM, Rathi PC, et al. Augmenting DMTA using predictive AI modelling at AstraZeneca. Drug Discovery Today. 2024;29(4):103945.

Recipe for success



Architecture and tech


Cloud


Automation


Reusable Code Templates

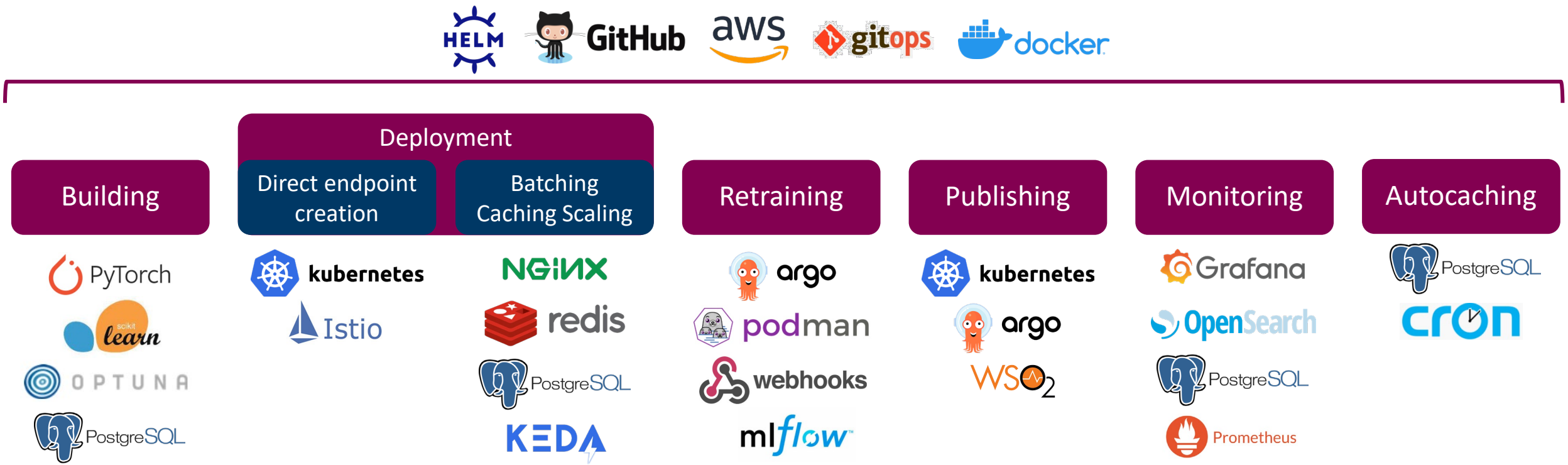
Source Control (GitOps, IaC)

Hybrid technologies


Consistent, augmented APIs

✅ Sustainable development

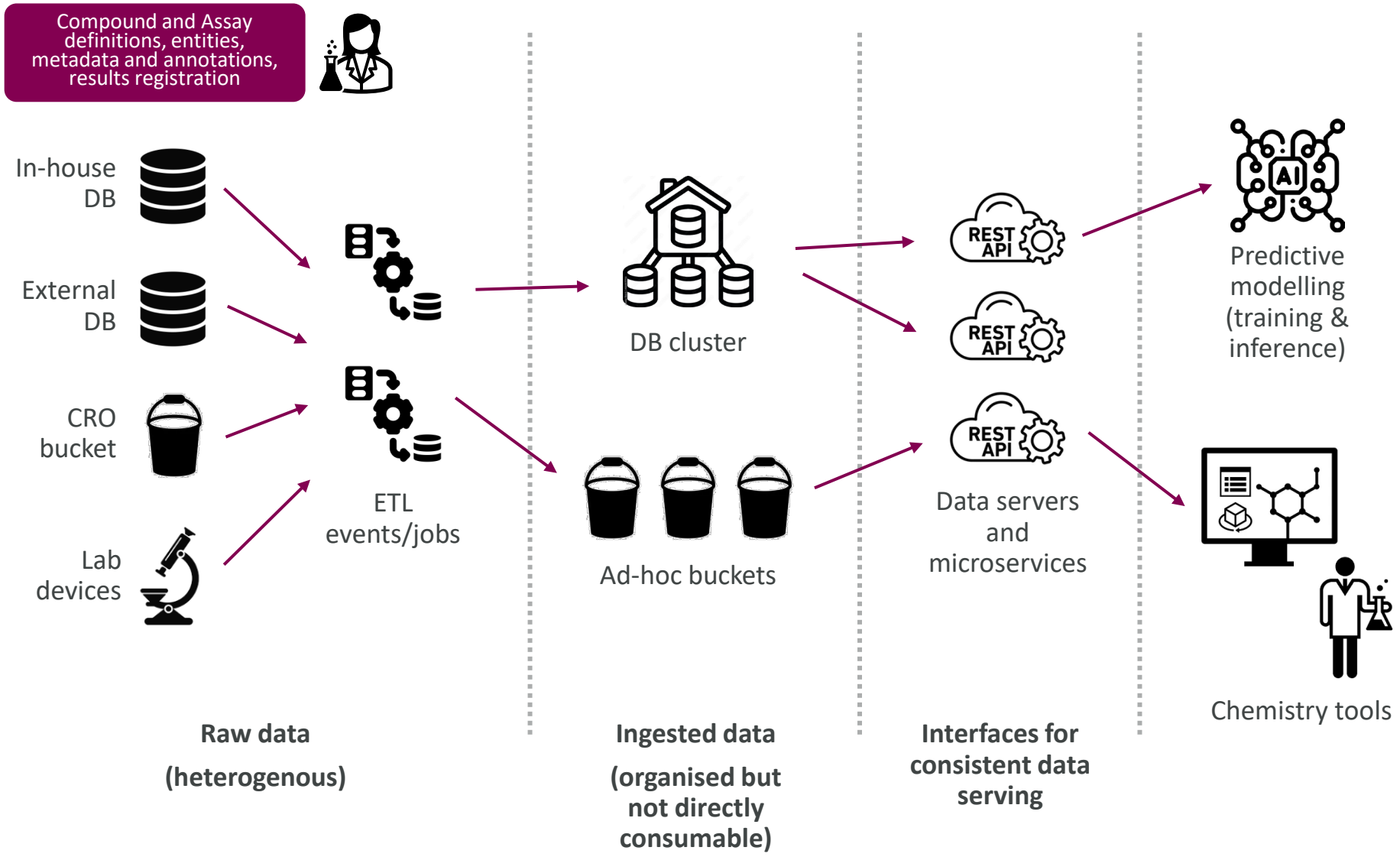
Architecture and tech: Hybrid technologies



The platform is implemented using a blend of proprietary and open-source technologies

Ghiandoni GM, Rathi PC, et al. Augmenting DMTA using predictive AI modelling at AstraZeneca. Drug Discovery Today. 2024;29(4):103945.

Data quality and FAIRness



Interaction between PIP and D360

Usage and history

D360 accounts for 90% of the calculations in PIP and the first downstream app integrating PIP

Inter-dependence

D360 uses PIP for predictions – PIP uses D360 for data

Functionalities

Users can access extended PIP functionalities (e.g., images, global/data parameters) via D360

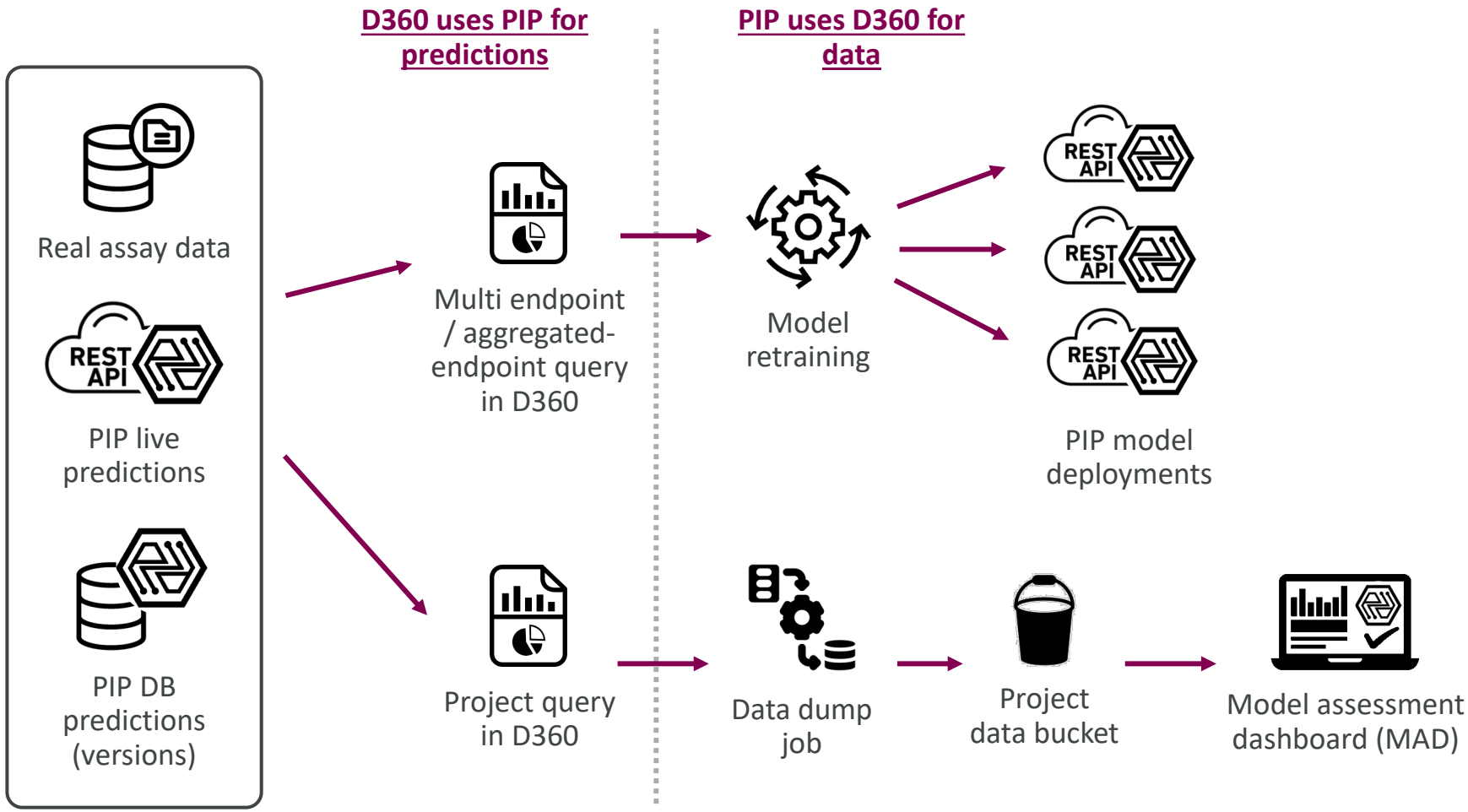
Versioning

D360 can query historical/versioned PIP predictions

Federation

D360 exposes real and virtual assay (PIP) data

Inter-dependence, Federation, and Versioning



Model assessment dashboard:
Built outside D360 but uses it as a data source

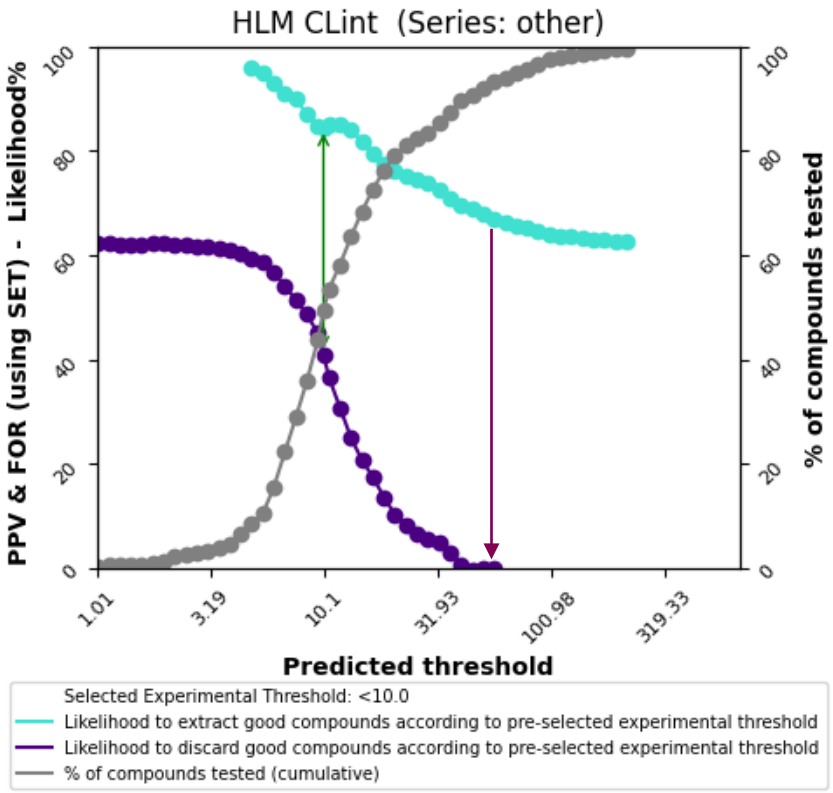
Model assessment dashboard (MAD)

Precision modelling suite

- Metrics monitoring
- **Threshold optimisation**
- Heatmap
- MPO optimisation (in progress)

What predicted threshold gives best enrichment for Series: other

	Predicted Threshold	PPV %	FOR %
Selected Experiemental Threshold	10.0	84	41
Recommended Threshold	45	68	0

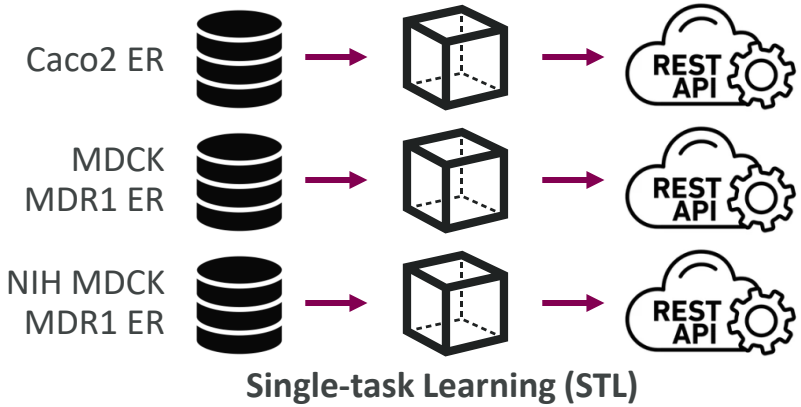


Scannell et al. Predictive validity in drug discovery: what it is, why it matters and how to improve it. Nat Rev Drug Discov. 2022;21, 915–931.

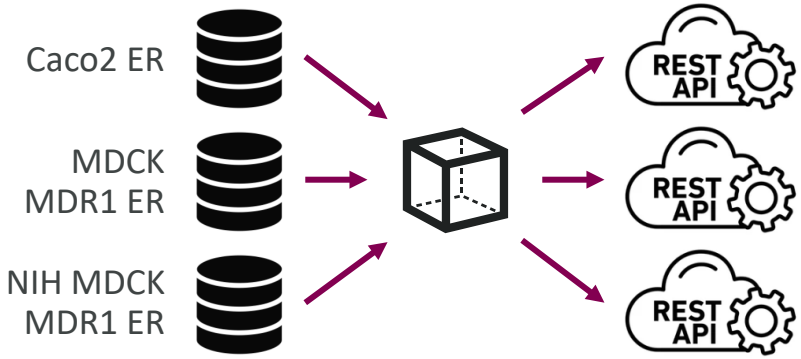
Viklund J, Ghiandoni GM, et al. Practical Guidelines for MPO Score Optimization in Drug Design: A Cross-Company Case Study. (Manuscript in preparation)

Extended functionalities:

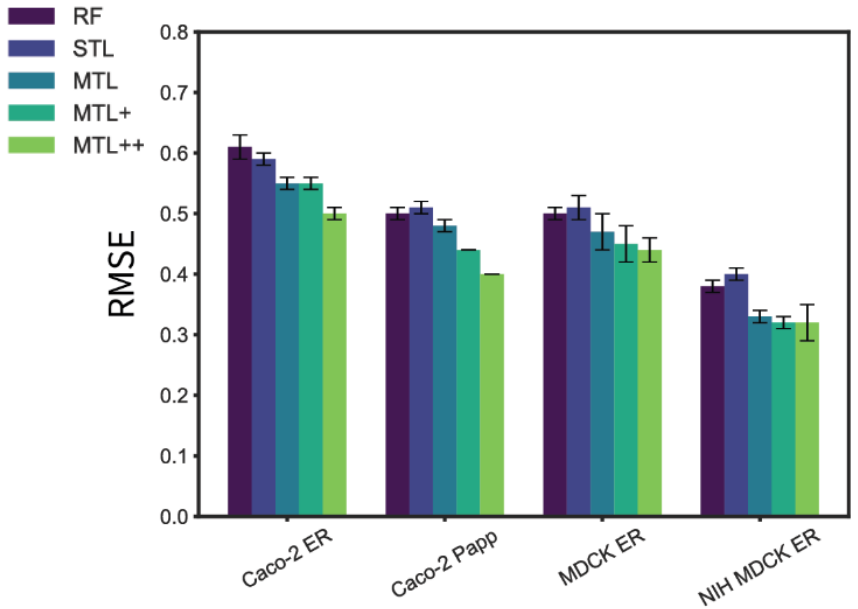
Making better predictions: Multi-task modelling pipeline



Single-task Learning (STL)



Multi-task Learning (MTL)



1. 20% overall increase of accuracy across 15 global ADME-Safety endpoints

2. Enhancement of low-data endpoint accuracy

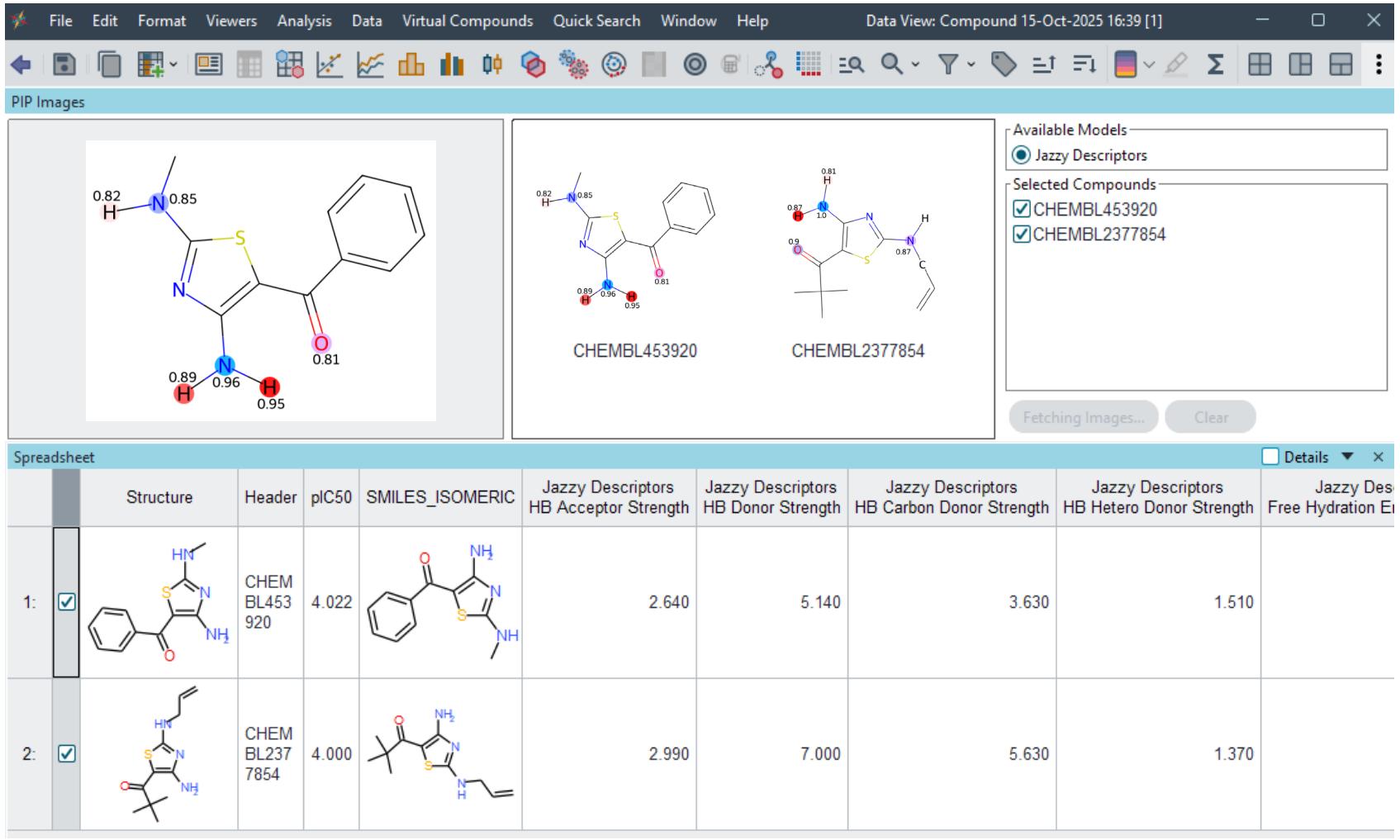
Ghiandoni GM, Subramanian V, et al. Prediction of permeability and efflux using multi-task learning. ACS Omega. 2025 (Manuscript Accepted)

Extended functionalities:

Understanding predictions: Image explanations

Atomic-level explanation

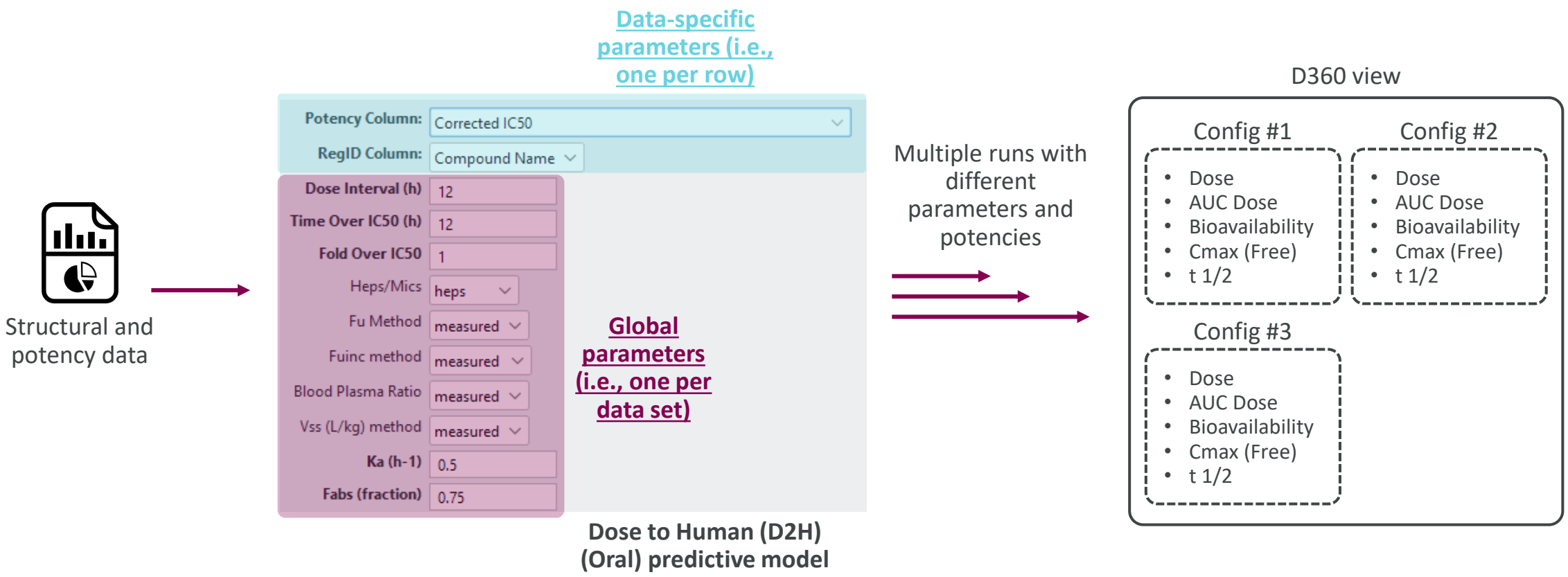
Molecular-level prediction



		Structure	Header	pIC50	SMILES_ISOMERIC	Jazzy Descriptors HB Acceptor Strength	Jazzy Descriptors HB Donor Strength	Jazzy Descriptors HB Carbon Donor Strength	Jazzy Descriptors HB Hetero Donor Strength	Jazzy Des Free Hydration E
1:	☑		CHEMBL453920	4.022		2.640	5.140	3.630	1.510	
2:	☑		CHEMBL2377854	4.000		2.990	7.000	5.630	1.370	

Extended functionalities:

Augmenting predictions: Global and data-specific parameters



Vision – What are we missing?

1. **Efficient interaction with slow & non-cacheable & non-batch-able calculations, e.g.,**
   - Calculations that heavily rely on freshly measured data (D2H)

2. **Improved prediction DB search also allowing multi-type IDs, e.g.,**
   - Optimisation of federated queries between Virtual (remote) and Registered (local) compound DBs

3. **Visualisation of structure-processing calculations, e.g.,**
   - Calculations using sequence inputs
   - Enumerations (one-to-many outputs)

4. **Better management of resources, e.g.,**
   - Identification of queries that run unnecessary calculations
   - Better UX for breaking events in queries (e.g., effect of schema changes to equations)

Acknowledgements

Augmented DMTA Platform

- Nagarajan PS
- David Riley
- Evgenii Maksimov
- Vladimir Chugunov
- Nick Tomkinson
- Susan Leung
- Arthur Garon
- Prakash Rathi
- Rajendra Kumar (ex AZ)

Stakeholders

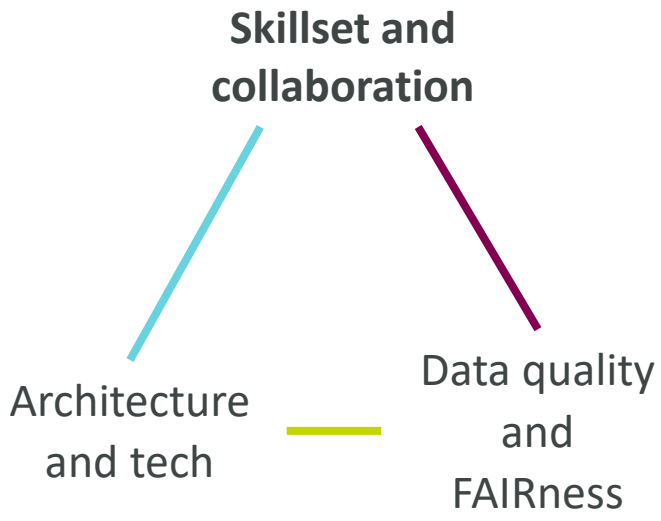
(Oncology & BioPharma R&D)

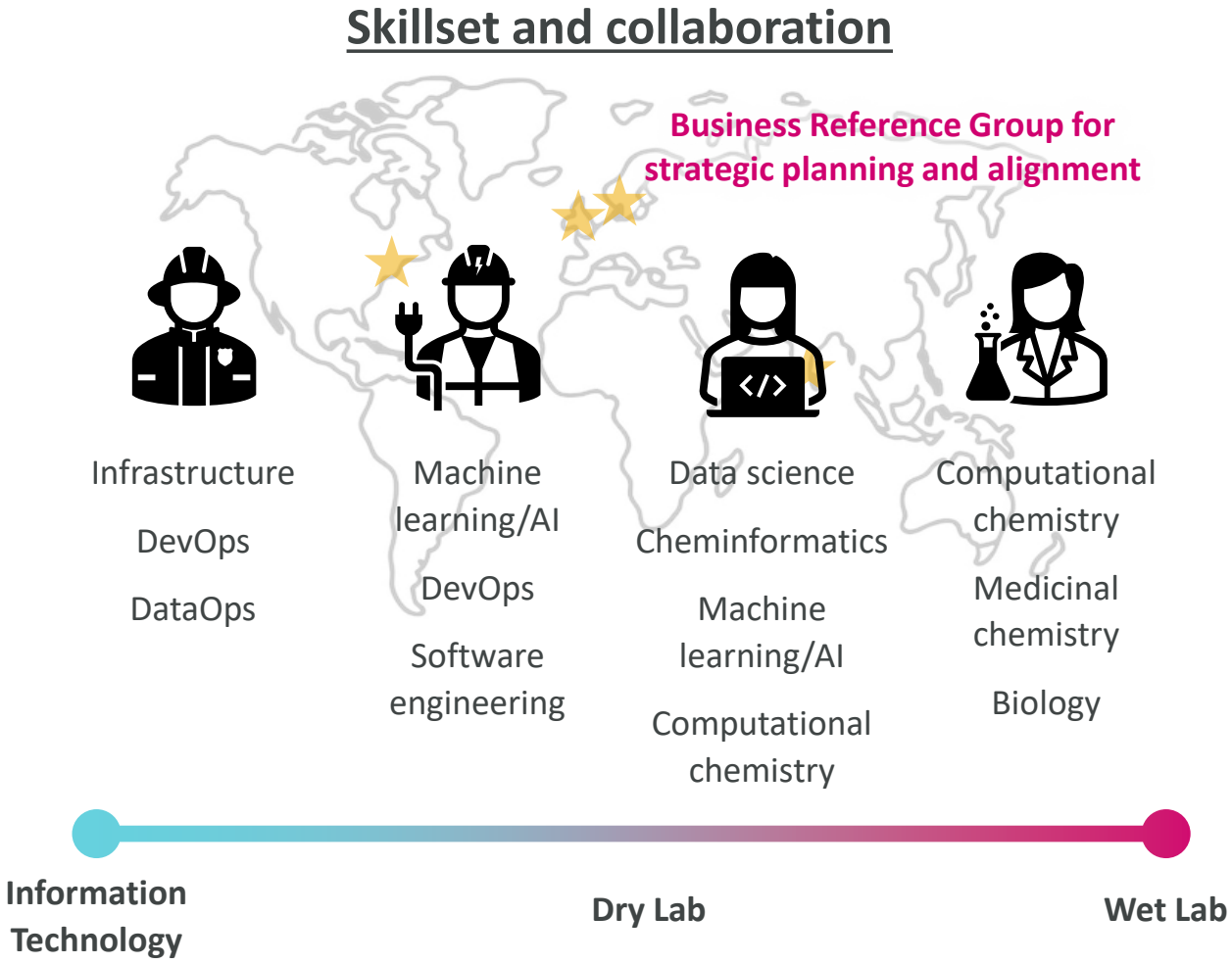
- Emma Evertsson
- William McCoull
- Vigneshwari Subramanian
- Jenny Viklund
- Greg Kauffman
- Susanne Winiwarter
- Graeme Robb
- Graham Smith

AstraZeneca PLC, 1 Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, CB2 0AA, UK
+44(0)203 749 5000
www.astrazeneca.com

Backup

Recipe for success



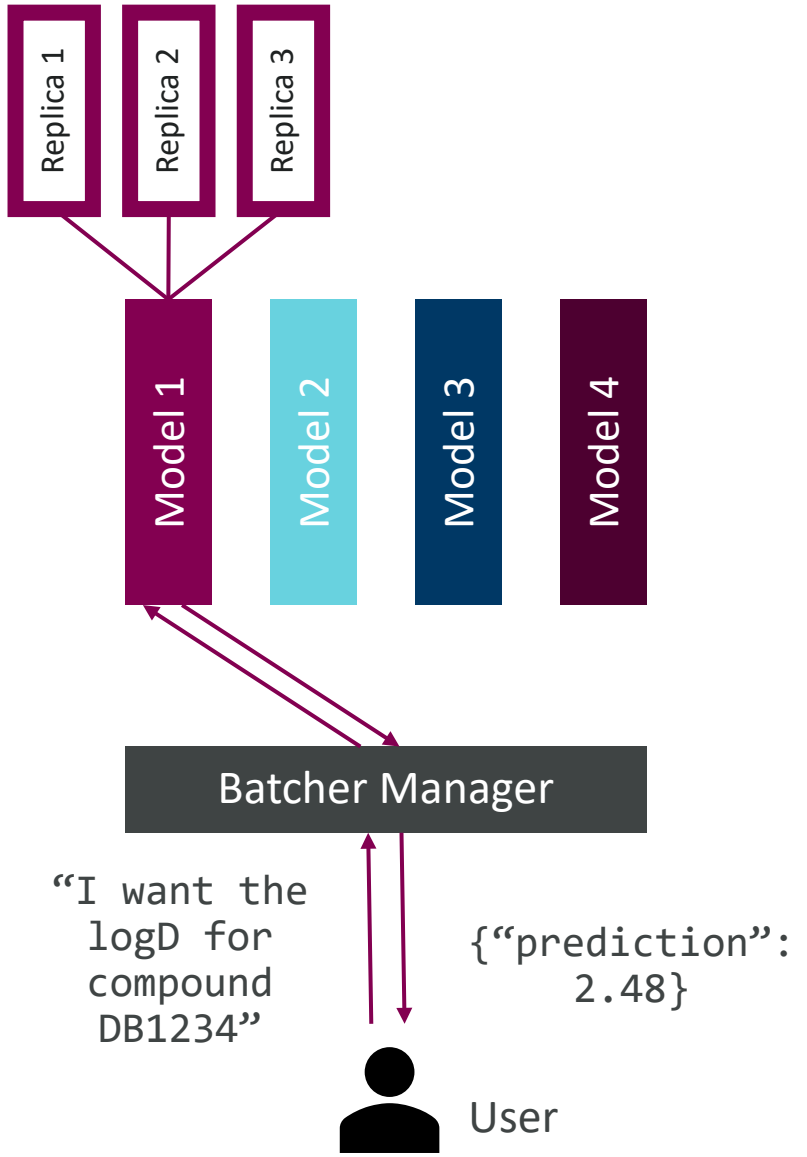


Architecture and tech: Consistent, augmented APIs

Users interact with a gateway called “Batcher Manager” (or Batman, for friends)

Batman provides many functionalities:

- **Consistent interface for all models/calculations**
- **Asynchronous requests**
- **Batching & Queue management**
- **Caching**
- **Database ID lookup**
- **Chemistry (e.g. format conversion, R-group replacement)**



Ghiandoni GM, Rathi PC, et al. Augmenting DMTA using predictive AI modelling at AstraZeneca. Drug Discovery Today. 2024;29(4):103945.

Architecture and tech: Templates and versioning

Reusable templates

~10 templates to (mostly) rule them all

Docker Deployment & Scaling

Multi-serving Deployment (QSARtuna)

Single-task retraining

Chemprop base image

Conformal base image

Multi-task retraining

Source control

One, versioned, source of truth for all environments