Unifying Data Science: A Single SCE for Innovation and Compliance

James Black, Executive Director
SCE Product Owner
Advanced Quantitative Sciences

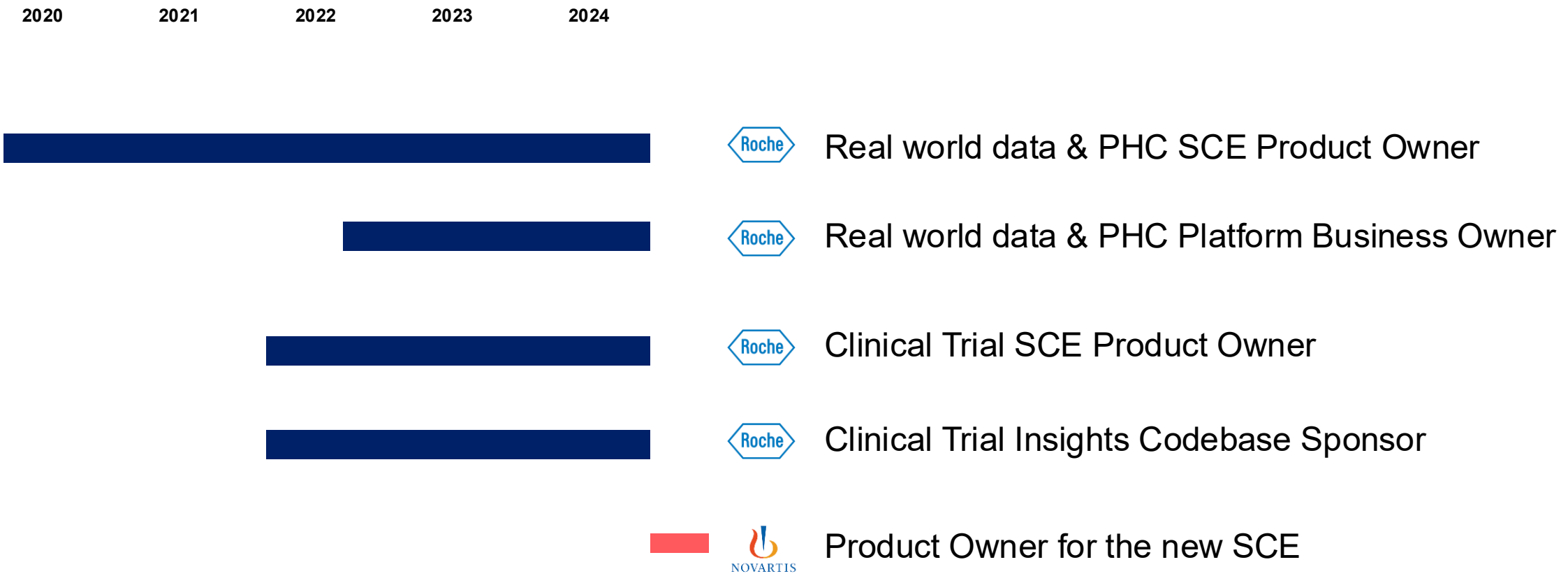
RevX Life Sciences | Philadelphia
May 20th | 2025

NOVARTIS | Reimagining Medicine

Agenda

1. Introduction
2. The evidence generation dichotomy
3. The shift to a single SCE
4. Where we're focused next
5. Q&A

My journey solving SCE challenges



Why GxP and exploratory workflows remain siloed

GxP SCE

The ‘stats programming’ platform

Allows a single propriety language

Sources codebase 10-20 years old

a common metadata store

Exploratory SCE

The ‘rest of the department’ platform

Often created due to needing R

Multi-lingual and open source focused

Modern tooling like git

Deep layer of abstractions present

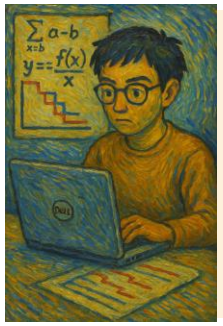
More of the ‘data science’ process exposed to users

Moving between GxP and non-GxP frustrates users and divides the department

Who they are: Key roles in biometrics

Statistical Programmer

Emphasis on efficiency, quantity of outputs, auditability



Statistician

Emphasis on statistical design, data flow end-to-end

Bioinformatician or Data Scientist

Emphasis on agility, large data and high compute

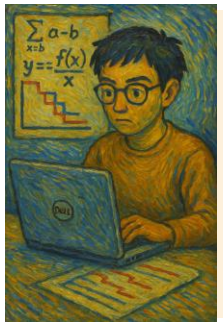
More GxP pre-specified analyses

Less GxP pre-specified analyses

How they work: Tools and environments

Statistical Programmer

RStudio SAS



Statistician

RStudio Shiny jupyter

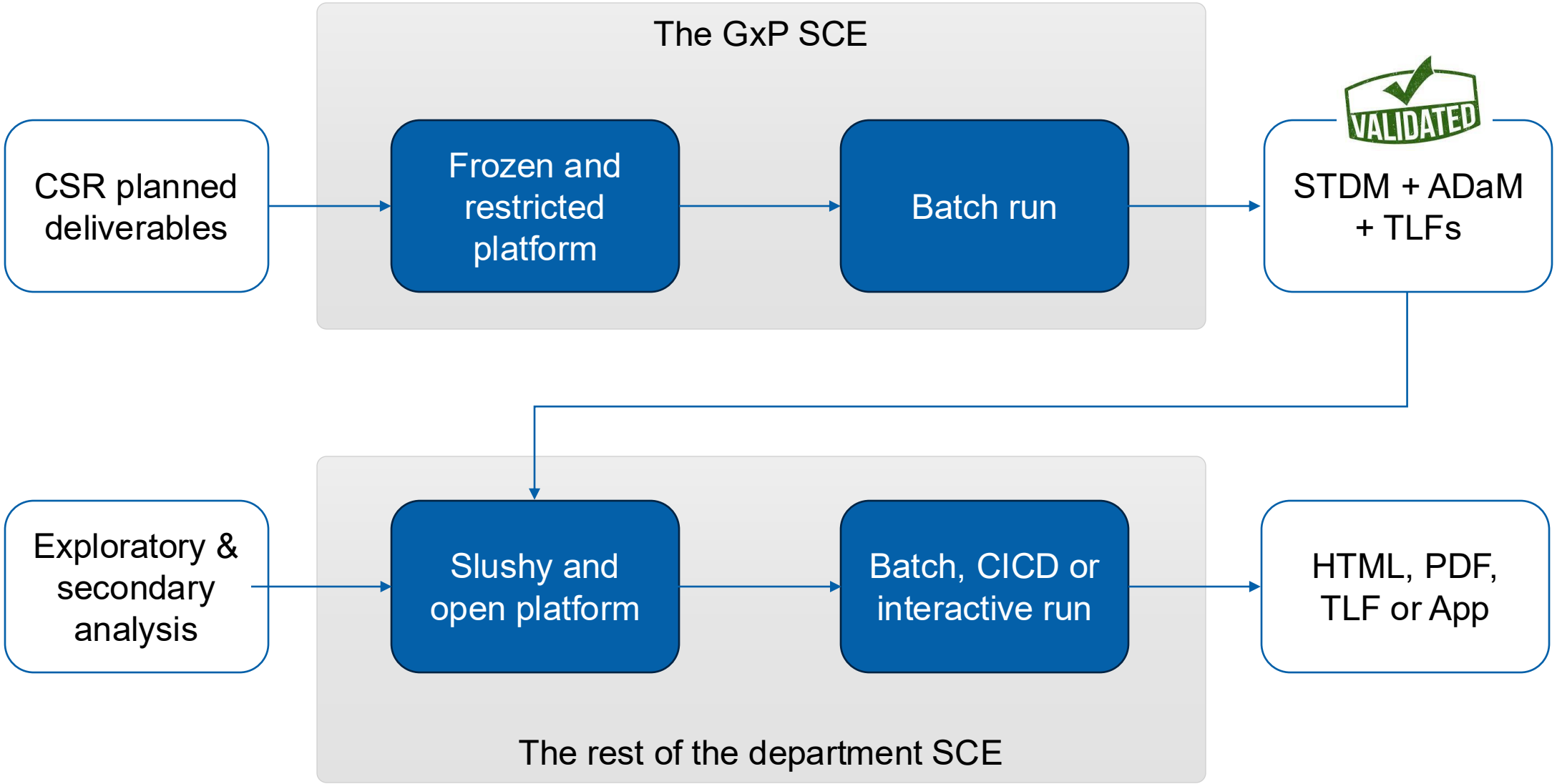
Bioinformatician or Data Scientist

RStudio jupyter nextflow

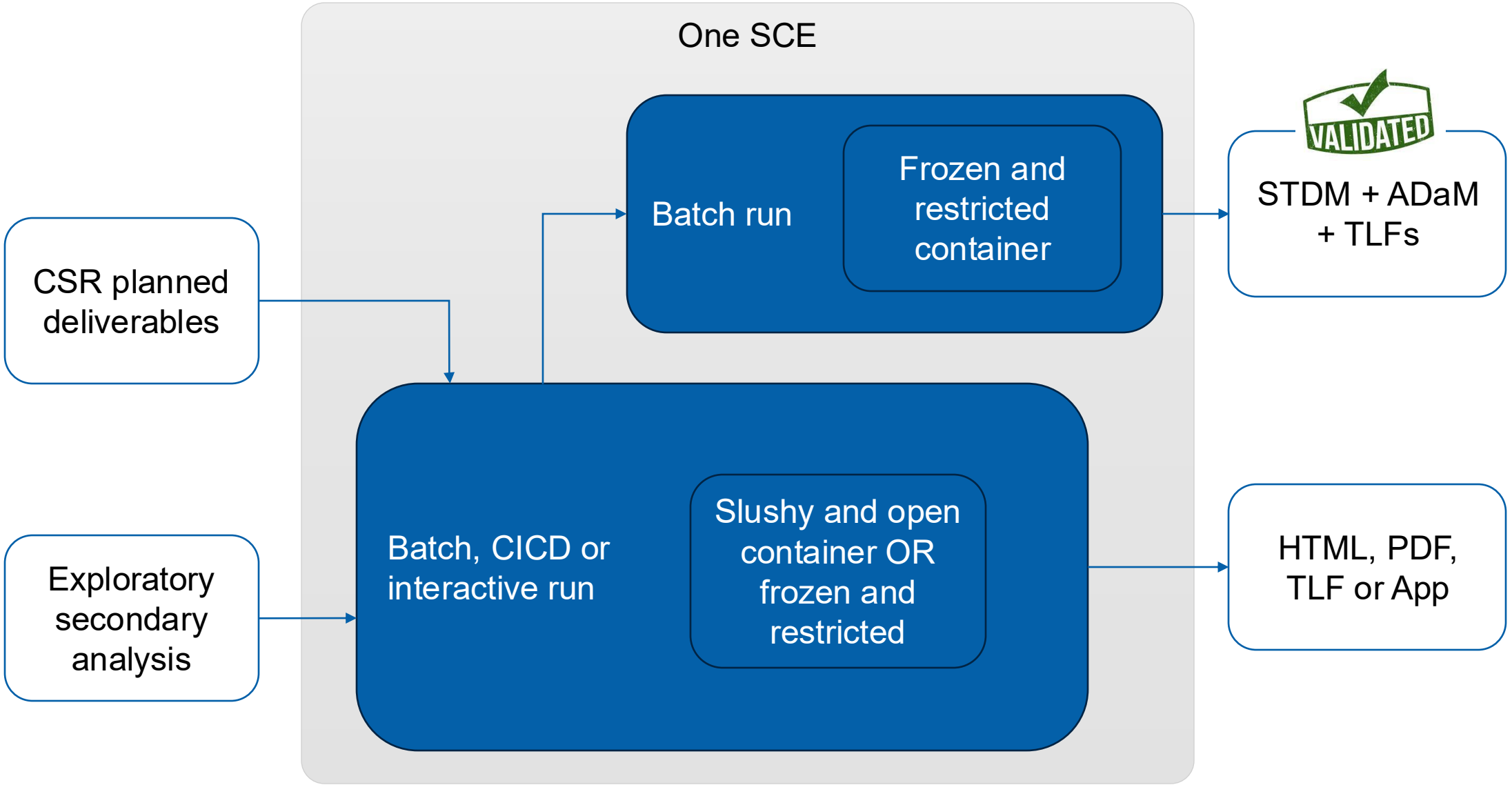
More GxP pre-specified analyses

Less GxP pre-specified analyses

Legacy SCEs: Fragmented workflows, frustrated teams



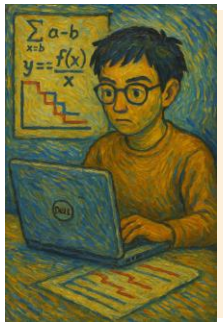
A blueprint for one unified SCE



What they gain: Value of unified SCE

Statistical Programmer

Easily execute GxP deliverables in a controlled, compliant environment without reworking code or moving between tools



Statistician

Develop and iterate in an open, flexible environment, with a seamless path to validated, production-ready outputs

Bioinformatician or Data Scientist

Work in modern, high-performance tooling without hitting compliance roadblocks when work transitions to production

Consolidation

Simplification

Harmonized Training

Innovation and compliance can coexist in a unified SCE

Development Space

Design and build
in familiar tools

Validated Execution

Promote to GxP with governed,
repeatable workflows

The blueprint is in motion

Abstractions: Scaling without slowing down

Efficiency & Speed	Compliance	Flexibility	Collaboration
Automate project setup	Controlled QC process and production runs	Some users / use cases won't need the abstractions	Ensuring FAIR principles

Why abstractions matter: Experts build, systems control

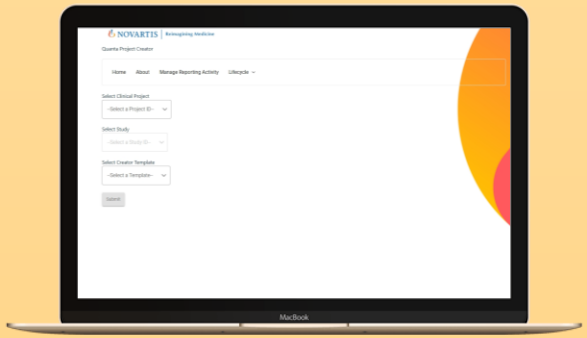
[It's Abstractions All the Way Down... - YouTube]

JD's assertions

- *The single biggest value of data science is it led to code written by domain specific experts*
- *Abstractions will leak, so abstractions should be permeable for debugging*
- *80 / 16 / 4 rule*

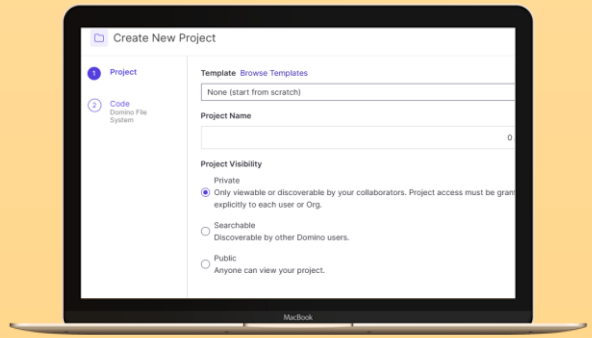
Project creation made simple and governed

Abstraction for Domino Project creation



- Automation (create git repo and project with naming scheme)
- Governance (default to specific validated workspaces)

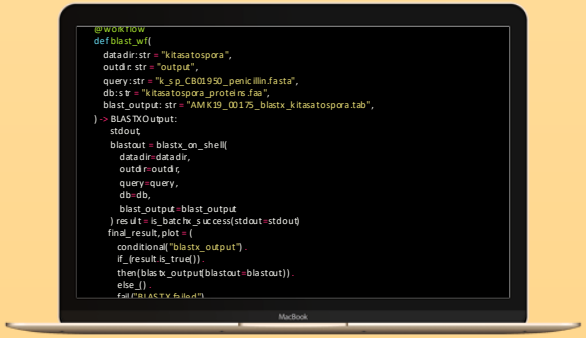
Native Domino Project creator



- Full flexibility to configure projects
- Manually configure data, permissions, etc.

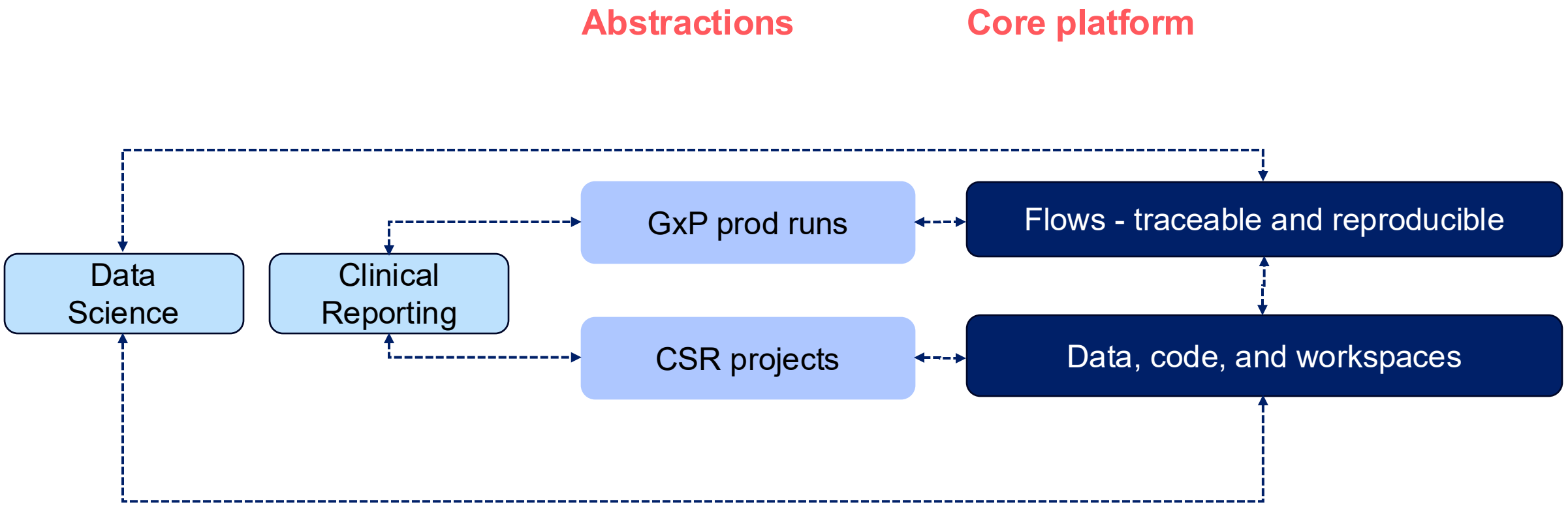
Orchestration

Native Flow (Flyte)

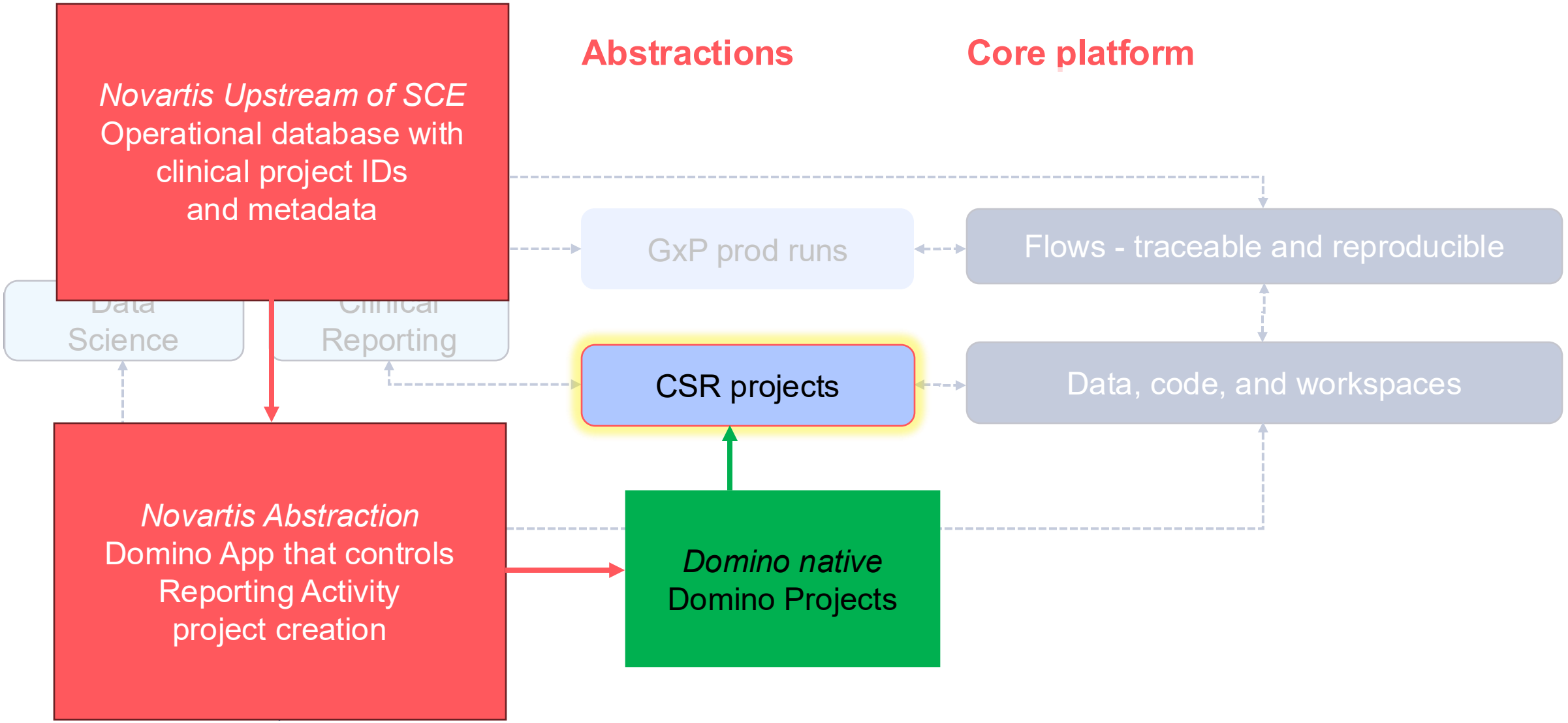


- Powerful and extensible
- Requires basic understanding of Python
- Expectation users conceptualize DAGs

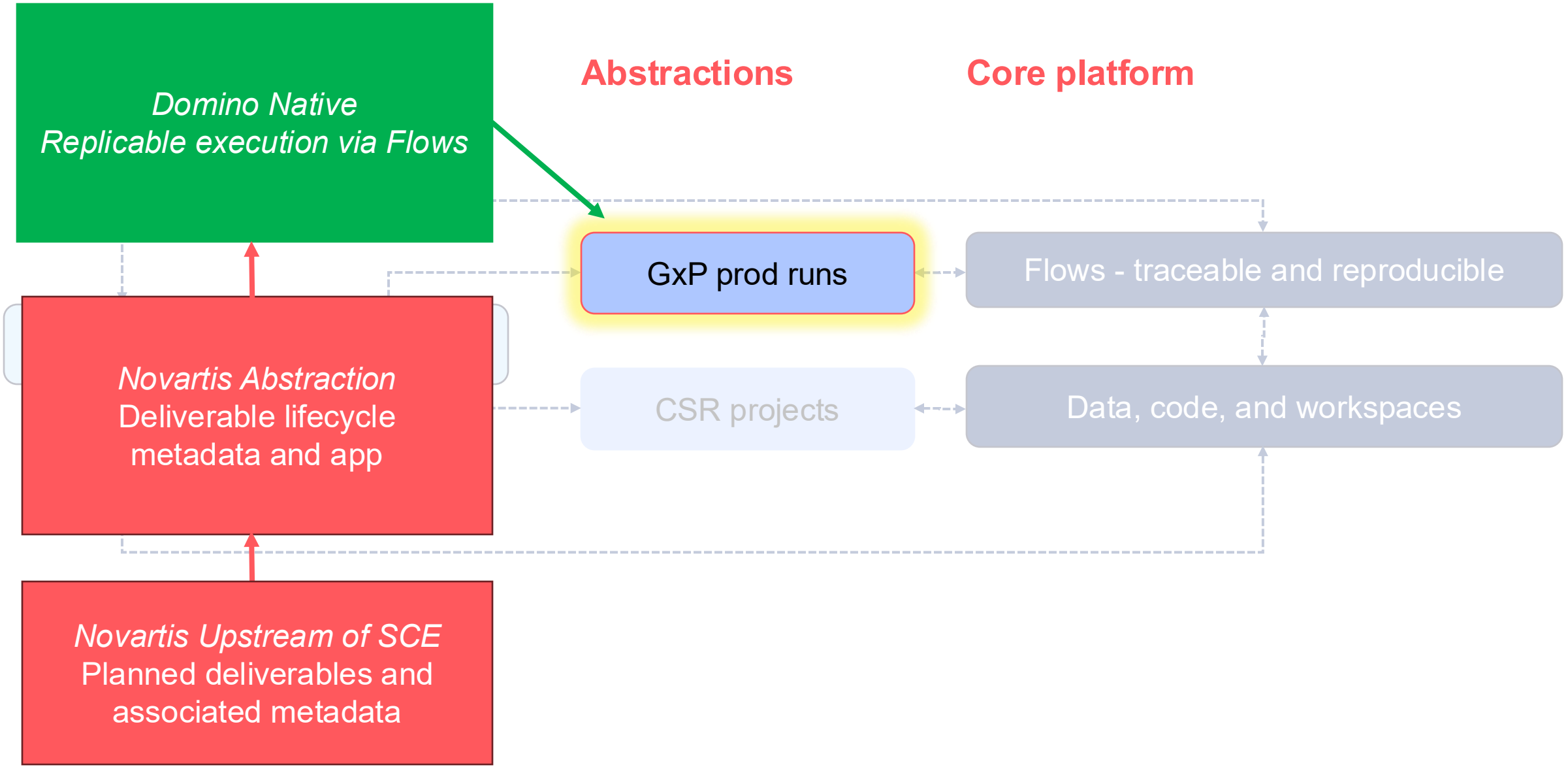
How abstractions enable ‘data science under one roof’



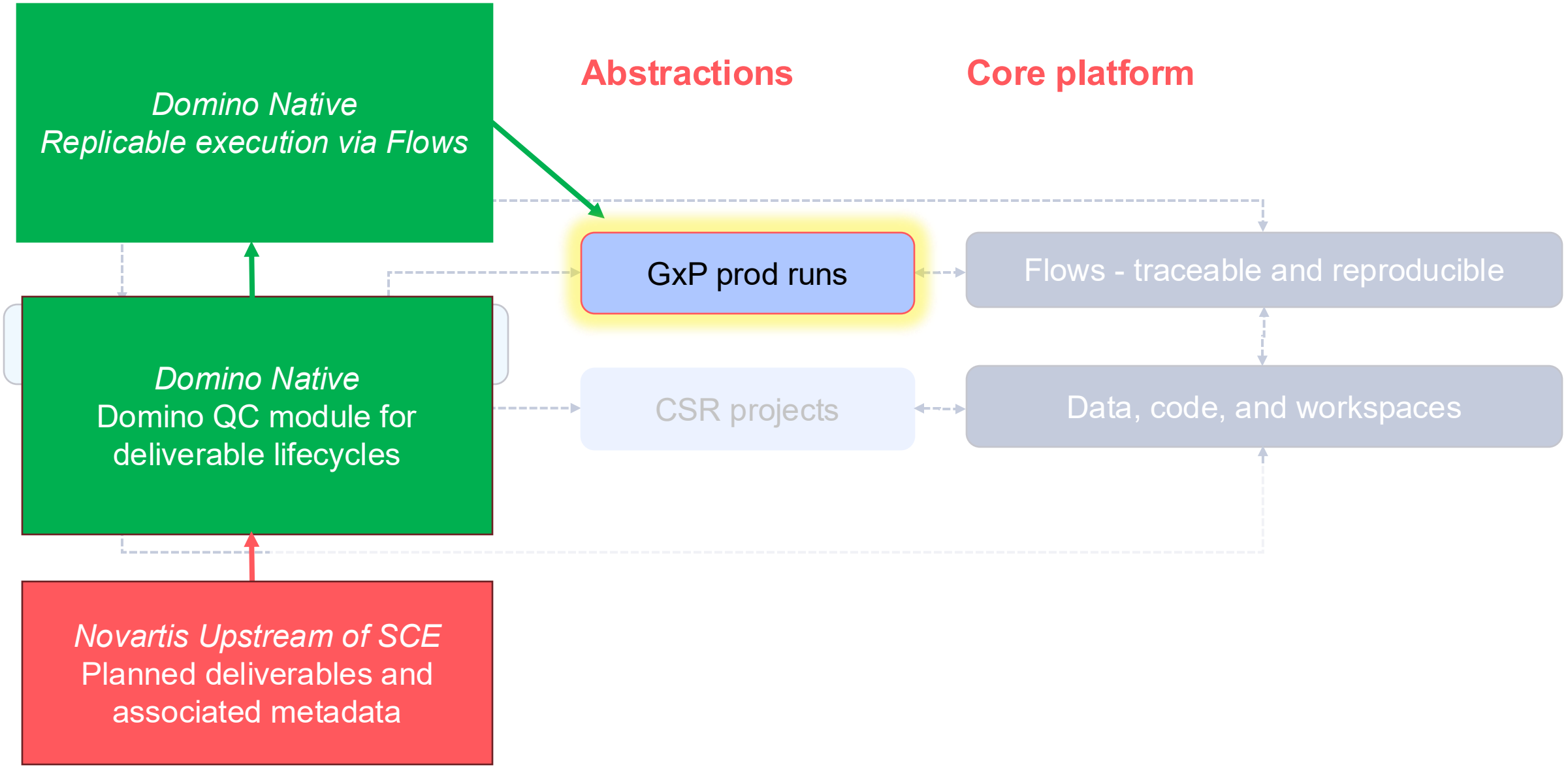
How abstractions enable unified workflows



How abstractions enable unified workflows



How abstractions enable unified workflows



A path forward

- Bridging GxP and exploratory work in one SCE

Focusing controls and validation on the flow from CDR to decision ready output

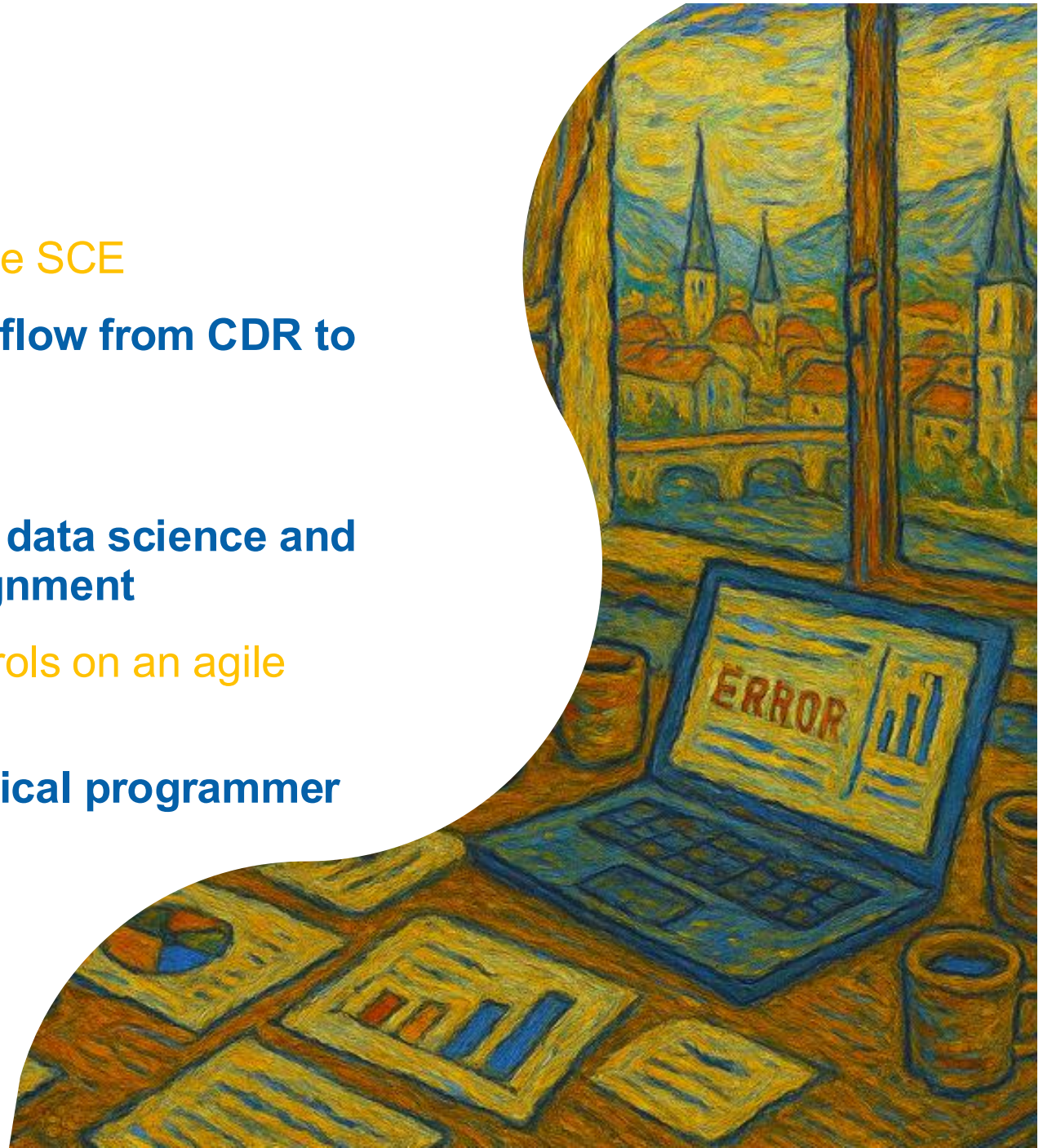
- Speaking the same language

Building shared understanding between data science and quality teams to accelerate trust and alignment

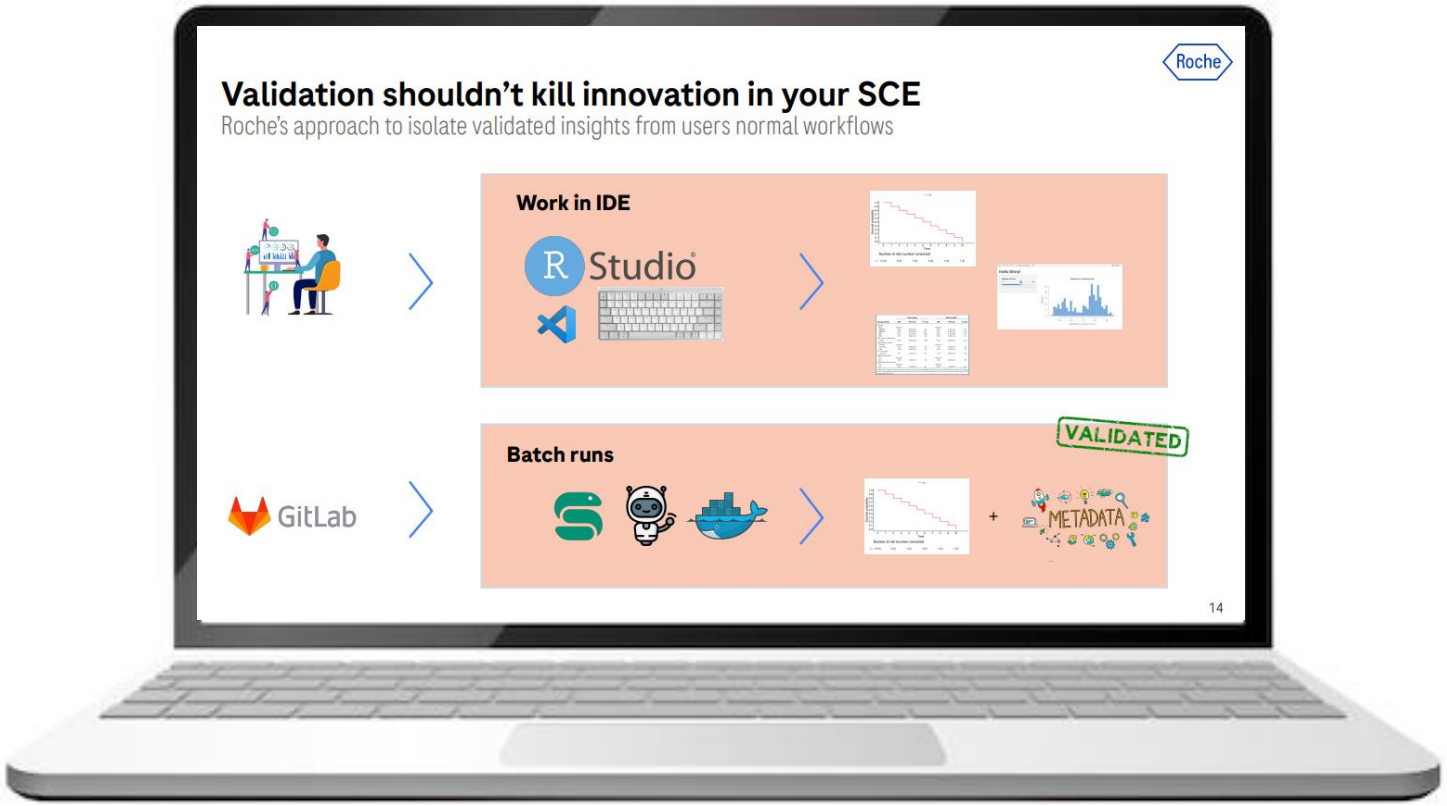
- Abstractions provide a way to layer controls on an agile platform

Enabling a bioinformatician and a statistical programmer

to happily co-exist in one SCE.



A de-coupling of development from GxP output creation exists



Source: The importance of the SCE in enabling our shift from proprietary programming to open-source data science (James Black) R/Pharma 2023

Key takeaways

- A unified SCE unlocks both innovation and compliance
- Abstractions make scale possible without adding burden
- Focus validation where it matters most

James Black
james_alexander.black@novartis.com

Q&A

NOVARTIS | Reimagining Medicine