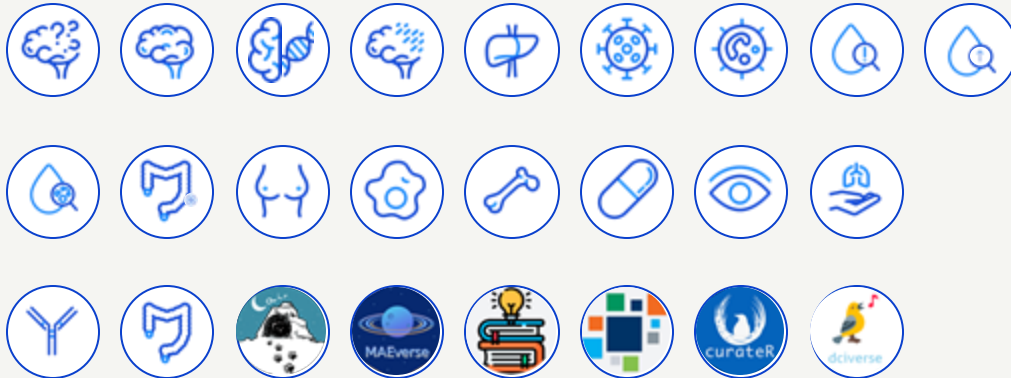


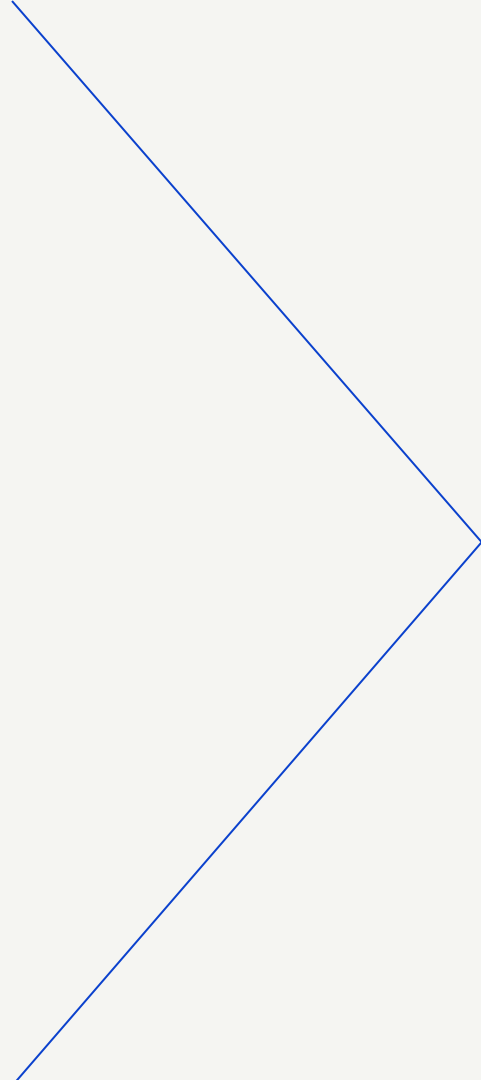
DH09: Database Lock is Only the Beginning: The Long, Winding Journey of Secondary Use of Data at Roche

PHUSE EU Connect 2024

Tamara Senior
Senior Data Scientist



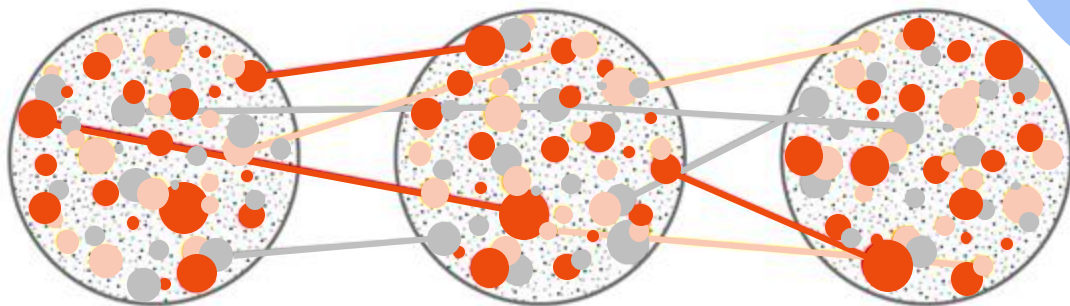
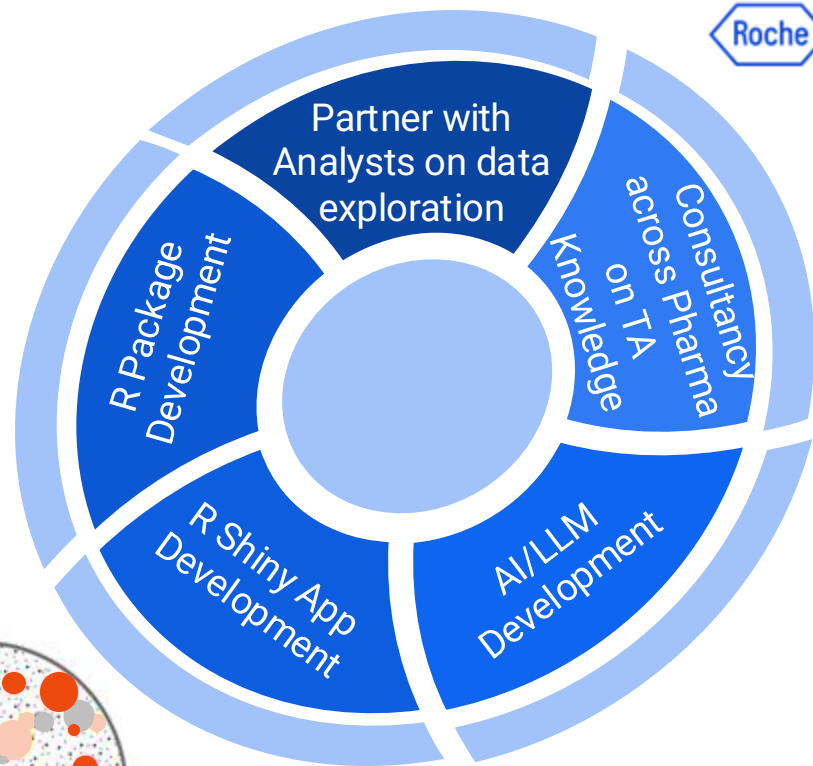
Contents

- 
- A large blue arrow shape pointing to the right, formed by two diagonal lines meeting at a point on a vertical line that runs down the center of the slide.
- Introduction to Data Marts & Secondary Use of Data at Roche
 - Clinical Curation & Integration
 - Molecular Integration
 - Learnings and Impact

Introduction to Data Marts & Secondary Use of Data at Roche

What does our group do at Roche?

Data Curation & Integration (DCI) specialise in creating high-quality data assets known as Data Marts, and in pioneering capabilities that unlock invaluable secondary data for transformative insights.



Data marts are multi-modal data assets

Data marts are harmonised across multiple studies

Clinical Data



Real World Data



Biomarker
Data*



Curation &
Integration E2E
engine



Data marts are diverse, large-scale data assets from patient data integrated across internal trials and external consortia

*This includes low & high-dimensional assay data, Sample Manifest Files (SMFs), images and Master Index (MI) files, videos, etc.

Data Marts at a Glance

52 R&D Enabling Insights have been generated from Data Marts since 2020

Data Mart Figures



19 Data Marts



264 Studies



49 Indications



284904 Patients

Insights per Disease Area



Solid Tumor: **14**



Hematology: **8**



I2O: **9**

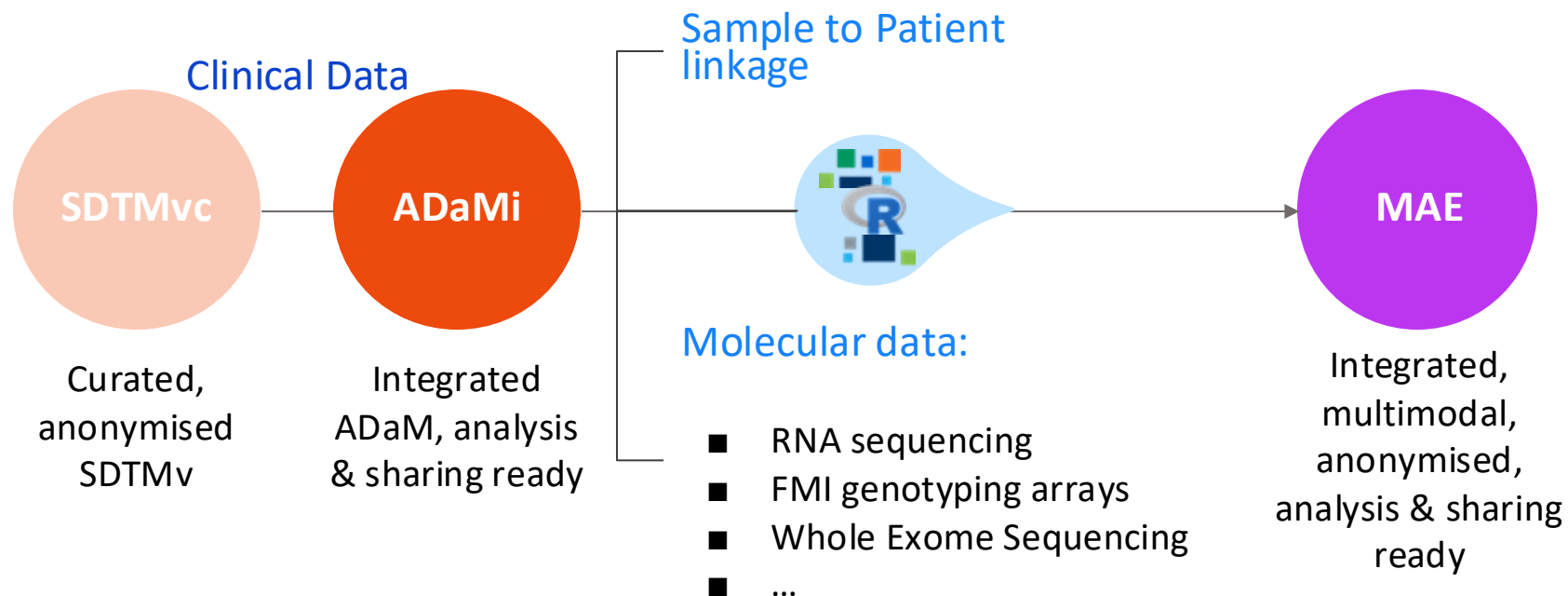


NeuroScience: **21**

Creating Data Marts: Clinical Curation & Integration

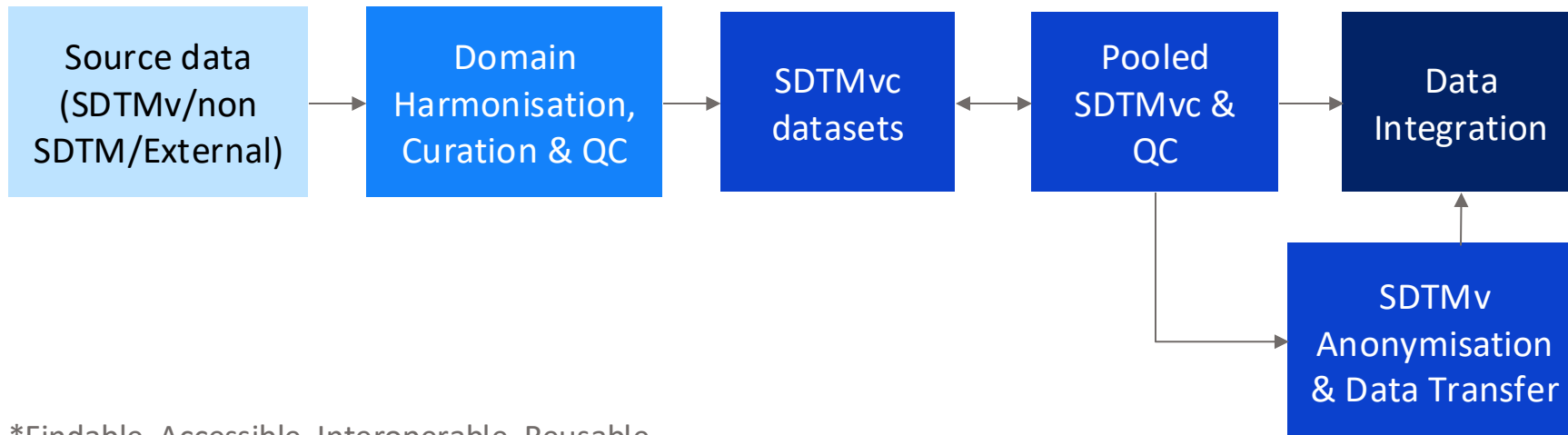
Data Marts are Delivered as 3 Data Models

Each Data Mart Consists of A Set of SDTMvc, ADaMi and MAEs



Dataflow of Clinical Curation

We deliver SDTMvc datasets based on our Clinical Data Curation Guidelines which provide the information required to define what FAIR looks like for the SDTMvc datasets, allowing curation and harmonisation of data from diverse internal and external sources, with varying standards.



*Findable. Accessible. Interoperable. Reusable.

Advantages of Clinical Curation

- ⊕ Data mapping for source data in Non-SDTM data models or old SDTM versions
- ⊕ Data Harmonization across multiple studies in use cases
- ⊕ Improved data quality before downstream Clinical Integration (ADaMi) datasets creation
- ⊕ Highly automated curation and harmonization on standardised data structure (SDTM)

What is Clinical Integration



We pool data that has been curated to a standard format and use SAS and/or R to generate analysis data sets based on Roche standard specifications.



How do we generate analysis ready data?

Plan first - data second

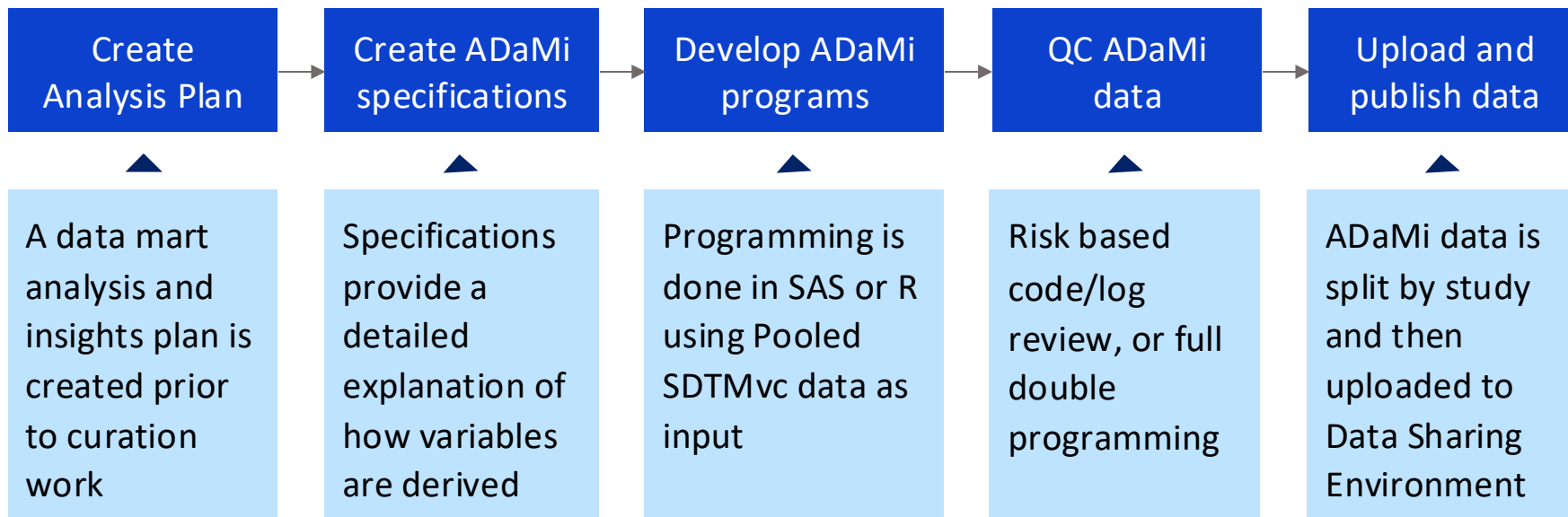
Analysis ready dataset begins with a **clear and unambiguous analysis plan**. Teams create a data mart analysis and insights plan prior to curation work. This plan is used to **create specifications** that will be programmed into analysis ready dataset.

You must have an understanding of the protocol and data collection when pooling data. Without this, the data may be misinterpreted.



Clinical Integration Workflow

ADaMi is the DCI integrated version of the standard ADaM data sets. The structure is the same, however, the focus is on questions that drive insights beyond the protocol

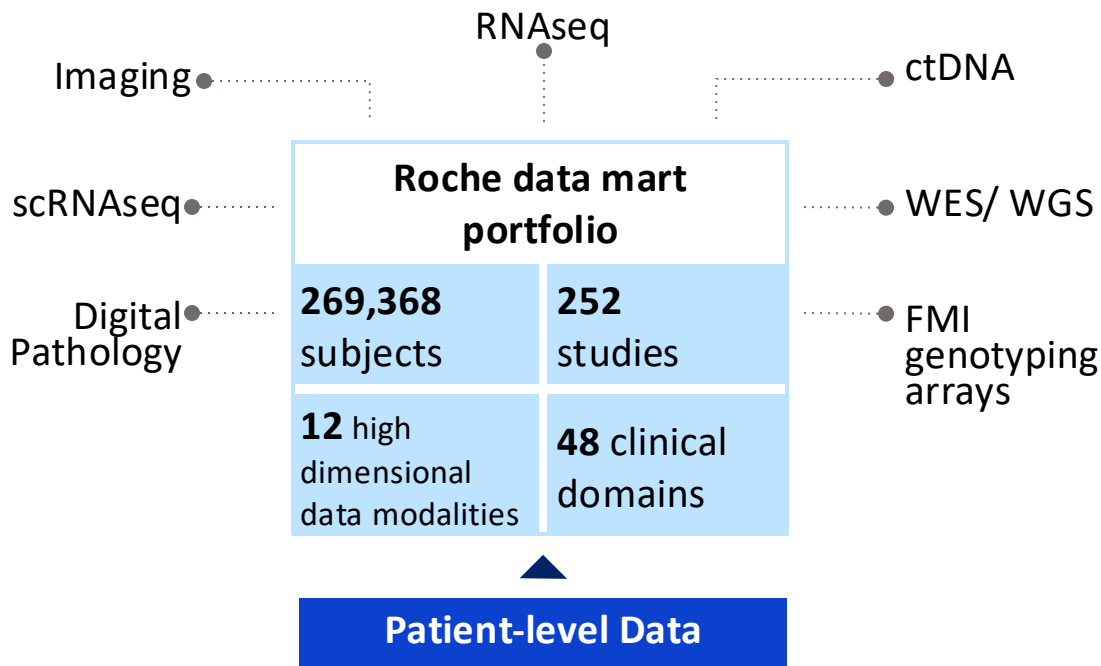


Advantages of Integrated ADaMi data

- ⊕ Pooled, harmonised, standardised analysis data
- ⊕ Harmonised Data Mart-specific Endpoints
- ⊕ All Adverse Event and Medical History terms are coded with the latest MedDRA version and Standardised MedDRA Queries (SMQs) and Customised Queries (CQ) have been added to support safety analysis
- ⊕ All Concomitant Medications are coded with the latest WHODrug version and dictionary derived or customised groupings are added for analysis

Creating Data Marts: Molecular Integration

What is Molecular Integration?

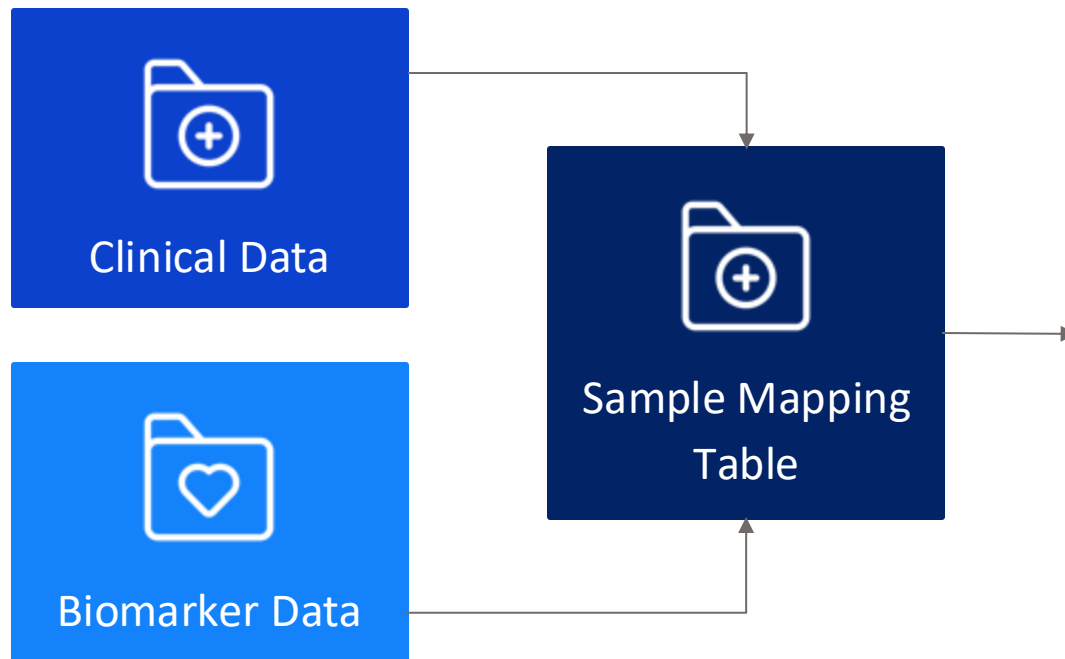


Multiple data modalities are integrated in the data marts

Harmonization and consistency are essential for analysis across studies, including clinical and biomarker data

And the list continues to grow, with novel data types...

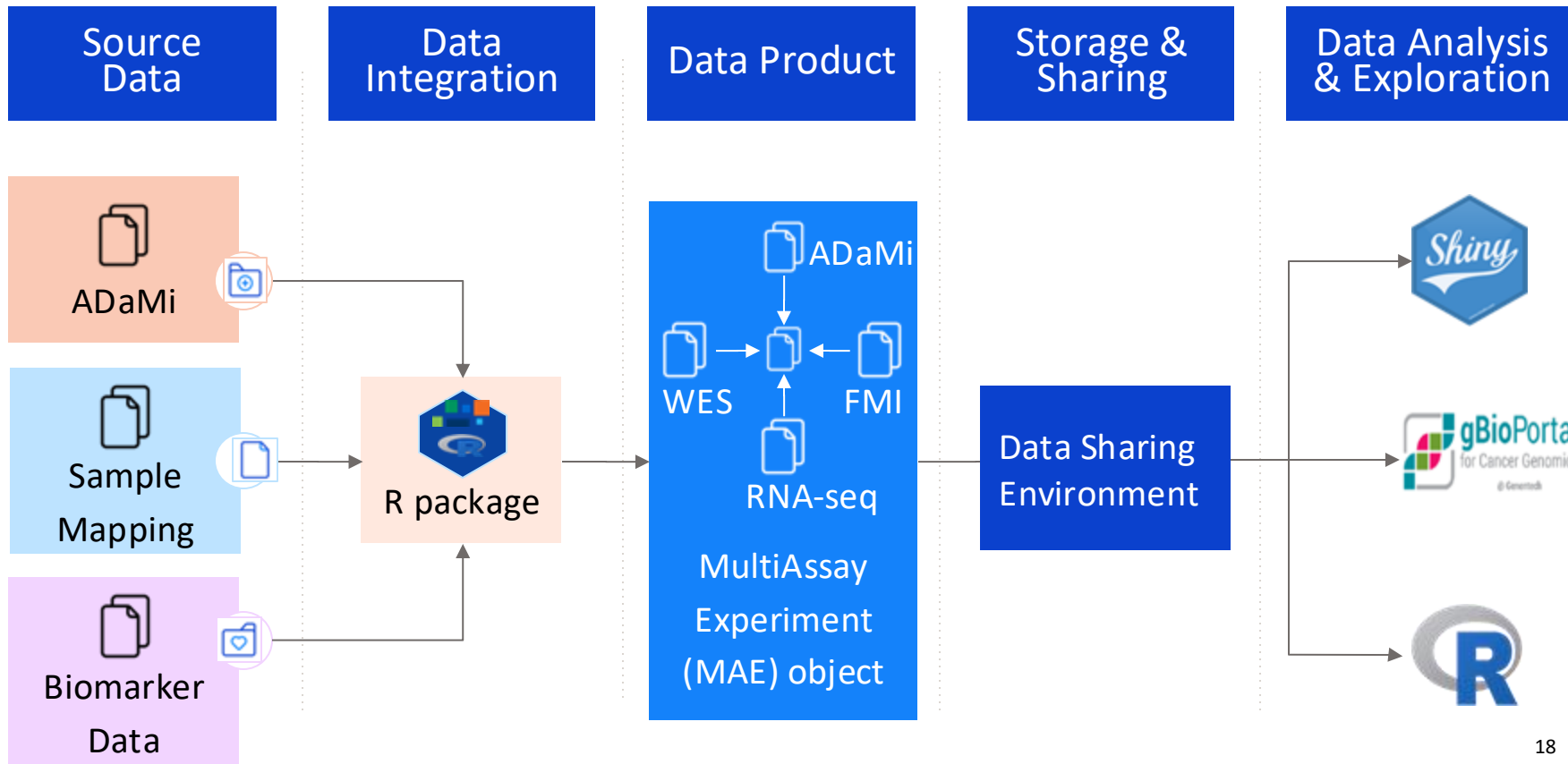
BioSample linkage enables integration with other data modalities



Integrated multi-modal datasets

Having sample linkage information, allows us to integrate clinical data with other data modalities and form multimodal datasets that are important for analysis and insights.

Molecular Integration Pipeline



Advantages of MAE data model

- ⊕ This open source data model is the data standard for high-dimensional biomarker data
- ⊕ The use of a 'common data model' enables automation of the pipelines and simplifies integration with other solutions
- ⊕ Enables analysis across data modalities through the linkage of patient to samples IDs
- ⊕ Simplifies data handling for analysis

Learnings and Insights

Lessons Learned Through Data Mart Creation

Data Standards are the only path to automation and interoperability



Context around the data is just as important as the data itself

Consideration during protocol development and every subsequent step along the way



Consistency across identifiers (e.g. Sample ID) is critical for linkage

Metadata about datasets is critical to reusability and needs to be part of the entire E2E lifecycle



Automation is key to scalability and efficiency



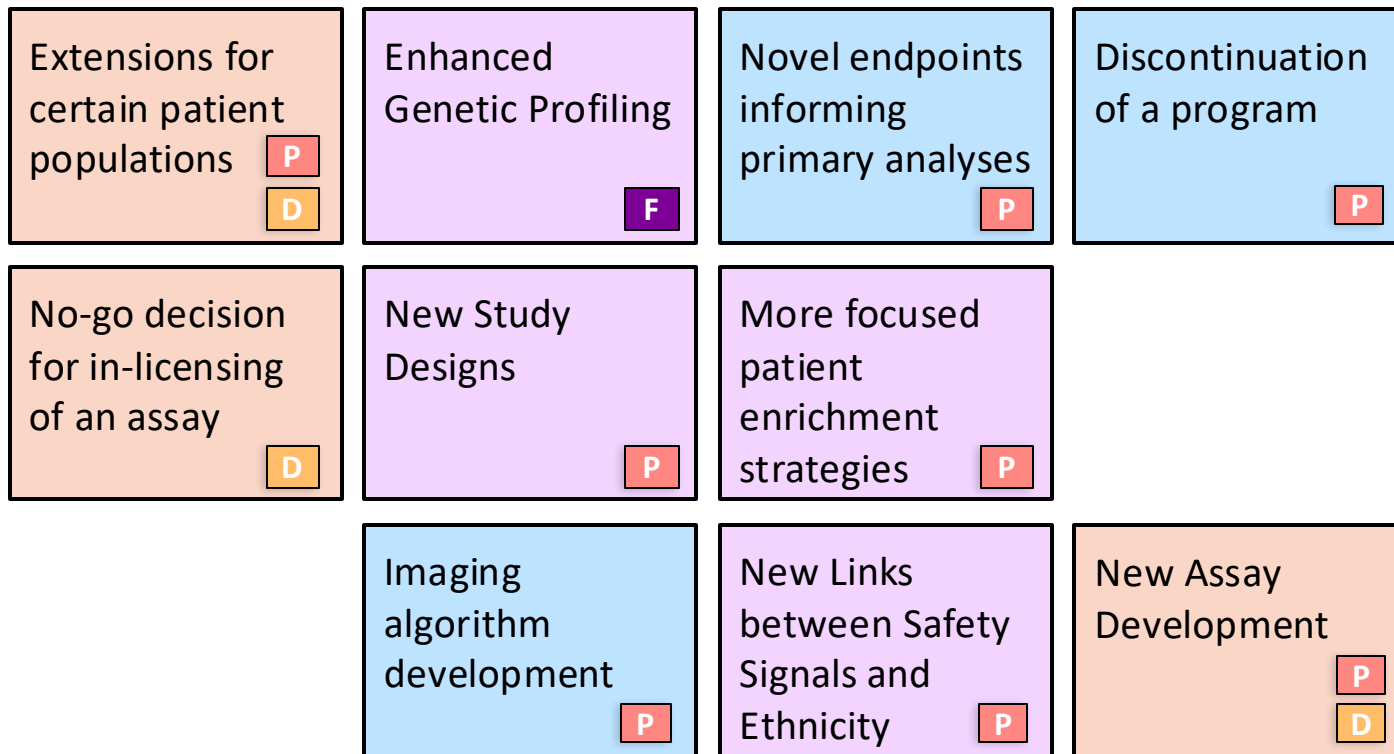
DCI Remit



Non-DCI Remit

Data Mart Insights

20 EDIS Data Mart R&D Enabling Insights generated in 2023



Participating Groups

P Pharma

D DIA

F FMI

Therapeutic Areas

I2O

Neuroscience

Oncology / Heme



Q&A



Contact Details

Presenter Name: Tamara Senior

Email: tamara.senior@roche.com