HSBC: Data for Analytics

Prepared by: **Sam Colyer, Head of Wholesale GRA Data, HSBC**

Tom Richardson, Senior Data Analyst, HSBC

Date: **7th November 2022**

HSBC

Table of Contents

Data Management for Analytics

Introduction and Background

HSBC require time-series internal and external data for:

- ◆ Advanced-Internal Ratings Based PD, LGD & EAD Model Development
- ◆ IFRS9 and Stress Testing forecasting models and execution
- ◆ Model Monitoring (A-IRB & IFRS9)
- ◆ Benchmarking and Reporting

In order to do this modellers need:

- ◆ Access to **live data from production data stores** to replicate live results and carry out monitoring and parallel analysis and testing
- ◆ Consistent time-series data snapshots to maintain **historic data views**
- ◆ Repeatable data refresh to support **model monitoring and reporting**
- ◆ Standardised **data assets** and user space
- ◆ Flexible **application and tooling** that links to underlying data
- ◆ Access to **external data for benchmarking** and model development

Delivering this across systems, jurisdictions and time-periods together with increased regulatory scrutiny, is a significant challenge.
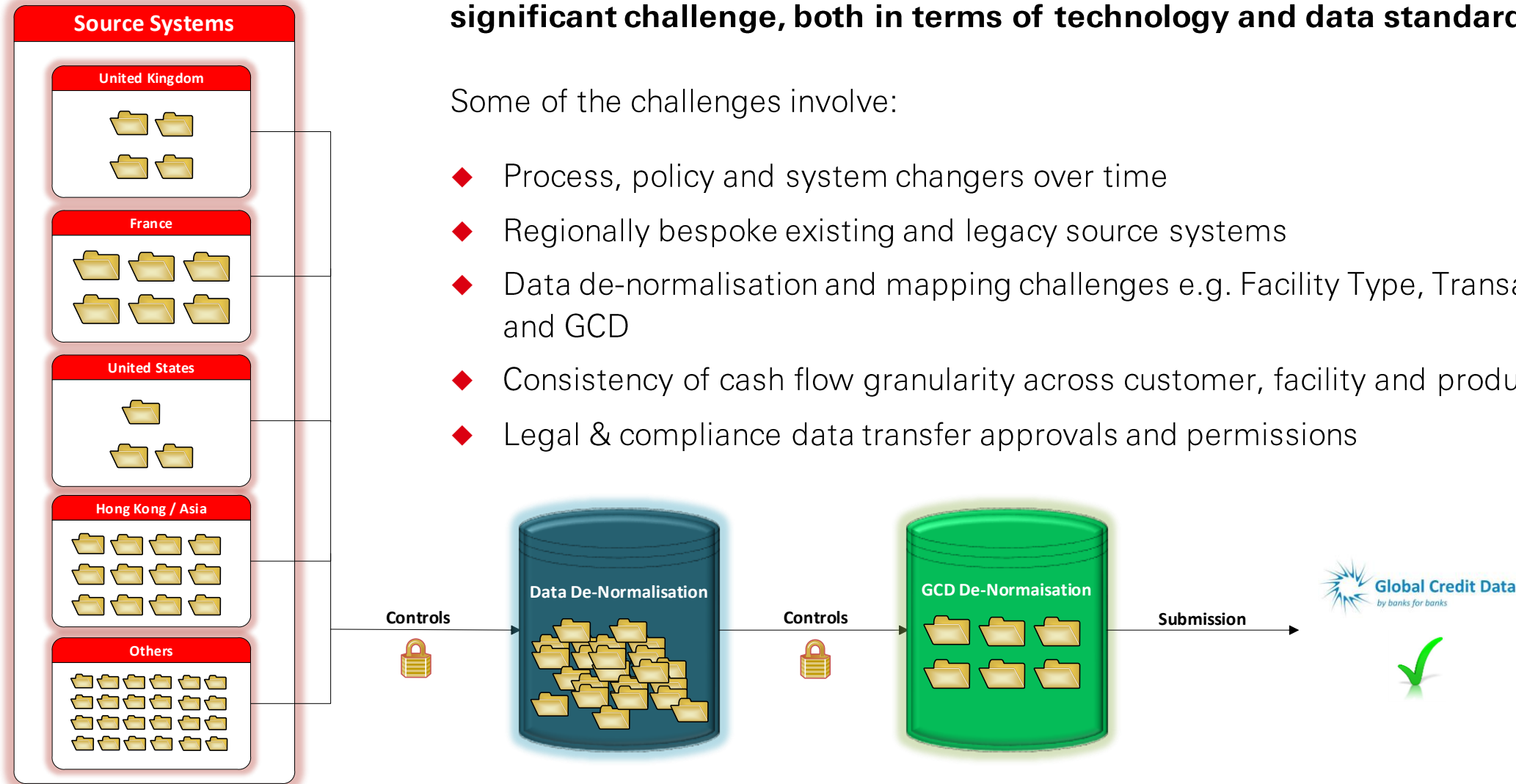
Data Management for Analytics

GCD submission is a micro component of a wider standardisation challenge

Source Systems

- United Kingdom
- France
- United States
- Hong Kong / Asia
- Others

Controls

Data De-Normalisation

Controls

GCD De-Normaisation

Submission

Global Credit Data
by banks for banks

Consolidation of default & loss data over a time-series in particular is a significant challenge, both in terms of technology and data standardisation.

Some of the challenges involve:

◆ Process, policy and system changers over time

◆ Regionally bespoke existing and legacy source systems

◆ Data de-normalisation and mapping challenges e.g. Facility Type, Transaction Type – internal and GCD

◆ Consistency of cash flow granularity across customer, facility and product level

◆ Legal & compliance data transfer approvals and permissions

Data Management for Analytics

Data is a challenge everywhere……..
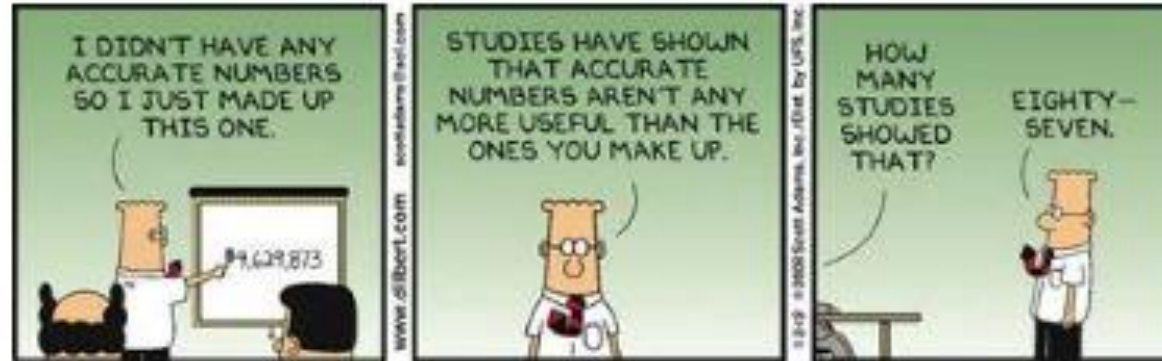
So how can we deliver a data management framework that meets the needs of a dynamic model analytics function, as well as GCD submission?

Data Management for Analytics

Guiding Principles for Model Data Management Framework

How can model users be satisfied that the data it is sourcing is appropriate for its intended use ?

Data provided must be:

- ◆ Robust > Data and IT supported data ingestion processes, including history and periodic refresh

- ◆ Sustainable > Scalable environment with managed data change process

- ◆ Repeatable > Batch processing of snap-shot system data drops appended to historic data

- ◆ Consistent > Documented transformations & lineage including controls on data transfer

- ◆ Convenient > Files & data can be shared between on/off shore teams plus IMR and others

GRADE (Global Risk Analytics Data Environment) is the consolidated data repository and integration programme that is key to implementation of the above principles.

When sourcing data we must ensure :

And is supported by :

1. **Data is ingested direct from source systems, with appropriate data controls**

2. **Available and transparent data lineage at a system and data attribute level**

3. **Data Controls monitored and evidenced, issues raised against data owners**

4. **Data Assets constructed around defined Critical Data Elements (CDEs) & DQ Reqts**

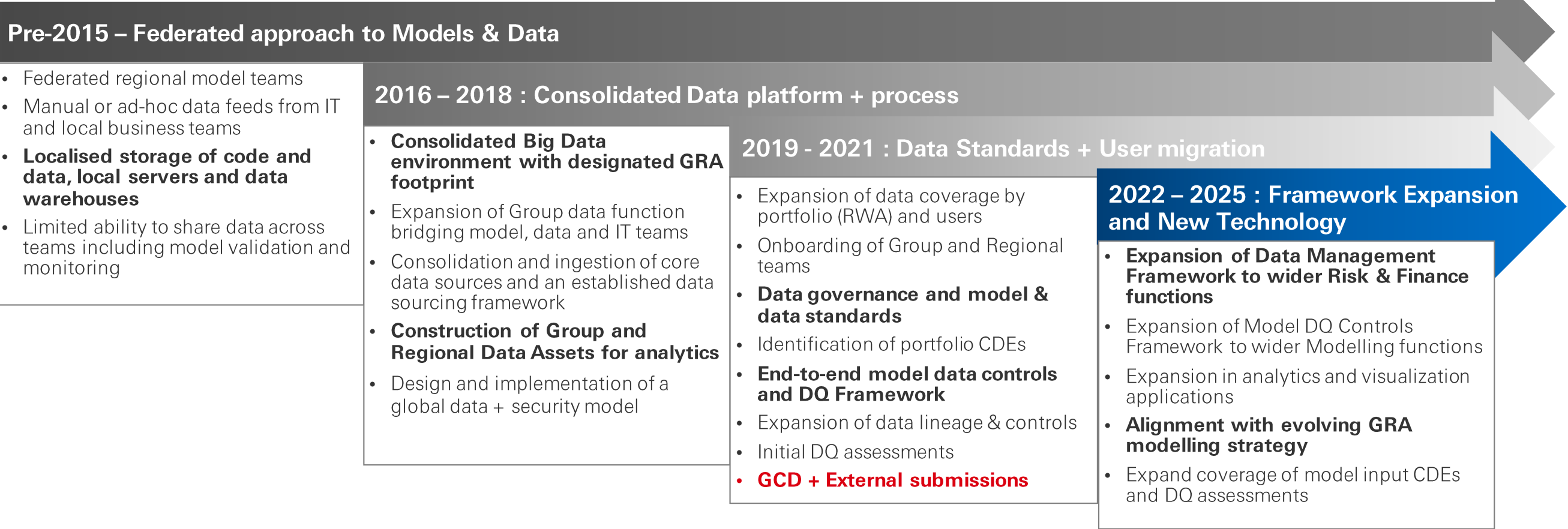**Model Development & Monitoring**

Model & Data Governance & Policy

Data Management for Analytics

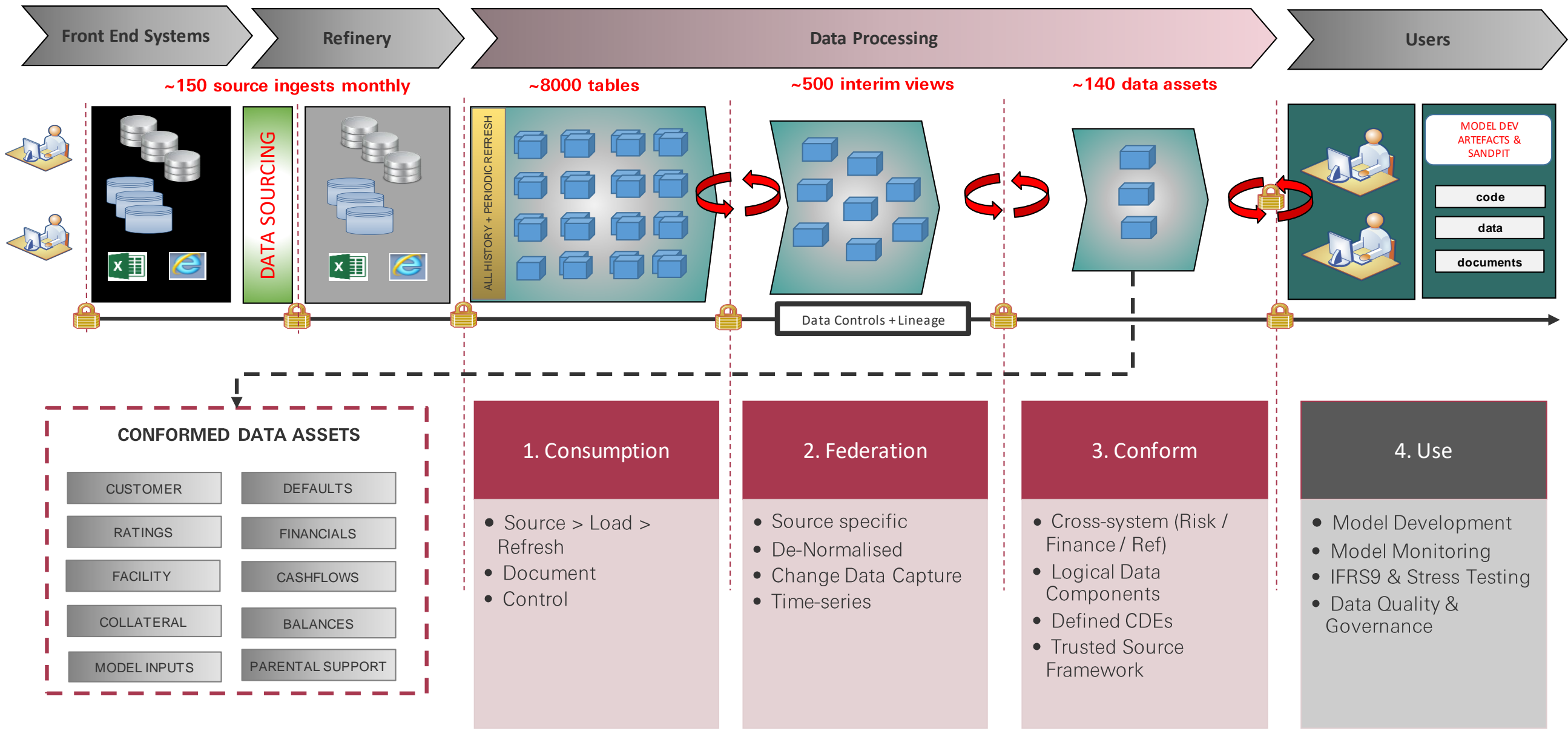## Implementation and maturity of data frameworks is multi-year, multi-functional programme

Data requirements to support Risk Analytics functions are complex, and different from standard reporting :

- Detailed requirements often not clear at project initiation - interactive feedback and collaborative working is critical to understand and evolve
- Scalable and flexible technology to support model strategy and regulatory demands
- Evidencing appropriateness and agreeing Trusted Sources requires an understanding of front-end processes, policy and controls
- Time-series data requires an understanding of changes in policy, process and systems over time

Pre-2015 – Federated approach to Models & Data

- Federated regional model teams
- Manual or ad-hoc data feeds from IT and local business teams
- **Localised storage of code and data, local servers and data warehouses**
- Limited ability to share data across teams including model validation and monitoring

2016 – 2018 : Consolidated Data platform + process

- **Consolidated Big Data environment with designated GRA footprint**
- Expansion of Group data function bridging model, data and IT teams
- Consolidation and ingestion of core data sources and an established data sourcing framework
- **Construction of Group and Regional Data Assets for analytics**
- Design and implementation of a global data + security model

2019 - 2021 : Data Standards + User migration

- Expansion of data coverage by portfolio (RWA) and users
- Onboarding of Group and Regional teams
- **Data governance and model & data standards**
- Identification of portfolio CDEs
- **End-to-end model data controls and DQ Framework**
- Expansion of data lineage & controls
- Initial DQ assessments
- **GCD + External submissions**

2022 – 2025 : Framework Expansion and New Technology

- **Expansion of Data Management Framework to wider Risk & Finance functions**
- Expansion of Model DQ Controls Framework to wider Modelling functions
- Expansion in analytics and visualization applications
- **Alignment with evolving GRA modelling strategy**
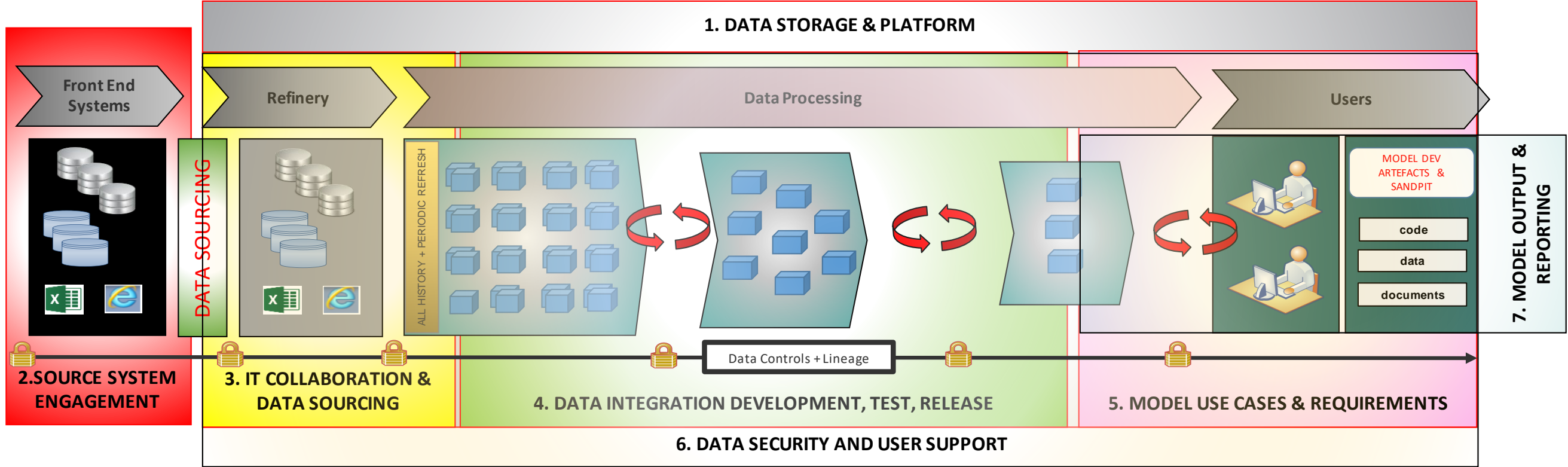- Expand coverage of model input CDEs and DQ assessments

Data Management for Analytics

Sources contain all history and periodic refresh, covering existing and legacy systems

| Front End Systems | Refinery | Data Processing | | | Users |

~150 source ingests monthly   **~8000 tables**   **~500 interim views**   **~140 data assets**

DATA SOURCING

ALL HISTORY + PERIODIC REFRESH

Data Controls + Lineage

MODEL DEV ARTEFACTS & SANDPIT

code | data | documents

CONFORMED DATA ASSETS

CUSTOMER	DEFAULTS
RATINGS	FINANCIALS
FACILITY	CASHFLOWS
COLLATERAL	BALANCES
MODEL INPUTS	PARENTAL SUPPORT

1. Consumption
- Source > Load > Refresh
- Document
- Control

2. Federation
- Source specific
- De-Normalised
- Change Data Capture
- Time-series

3. Conform
- Cross-system (Risk / Finance / Ref)
- Logical Data Components
- Defined CDEs
- Trusted Source Framework

4. Use
- Model Development
- Model Monitoring
- IFRS9 & Stress Testing
- Data Quality & Governance

Data Management for Analytics

A hybrid Data, Model and IT support model is critical to respond to changing requirements

Step	Component	Description	Who
1	Data Storage and Platform	Big data platform, data storage, data + analytics tooling, user sandbox area	Data, IT, Architects
2	Source System Engagement	System + resource prioritisation, architecture approvals	Data, Local IT / Grp IT
3	IT Collaboration & Ingest	Grp and local system owners database connections, system / scope, ingestion reqmts	Data, IT, Big Data
4	Integration Dev, Testing & Release	ETL development, testing, user engagement, change management	Data
5	Model Use Cases & Requirements	Model development & monitoring, IFRS9 & Stress Testing, data governance & reporting	Data, Model Analytics
6	Data Security and User Support	Data transfer compliance, access permissions management, maintenance, periodic review	Data, IT
7	Model Output & Reporting	Model development datasets, IMR reviews, Model Monitoring reports, DQ + reporting	Data, Model Analytics, Reporting

Data Management for Analytics
End-to-end Data controls complement the Model Life-Cycle

Data support and control artefacts are aligned with the model life-cycle, to provide an end-to-end data controls framework that is central to the fundamental principles outlined

- Data sourcing controls monitoring availability of data sources, including coverage and completeness of data transfer

- Central collaborative association with model teams, ensuring consistency of methods and approach

- Standardised Model data scorecards and aligned model & data governance

- Critical Data Elements identified, defined and managed under Data Governance

- Key Risk Indicators monitoring thematic system, process and policy issues

- Model Data Quality measurements and reporting

- Data sourcing support and consistency between model development, monitoring and implementation

Key to success is a recognition of the need for consistency in approach across Group and Regional model teams

Data Management for Analytics

Core components are required in combination to build a successful data analytics function
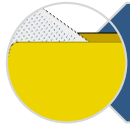
Strategic Data Vision

Stakeholder buy in and engagement

Dedicated hybrid data function

Collaborative Data, IT and Model functions

Iterative model and data development cycle

Aligned Model and Data governance and policy

Scalable analytics environment and technology

Questions

HSBC