forthpoint

DATA MESH AND FINTECH

Neil Carden, COO Forth Point

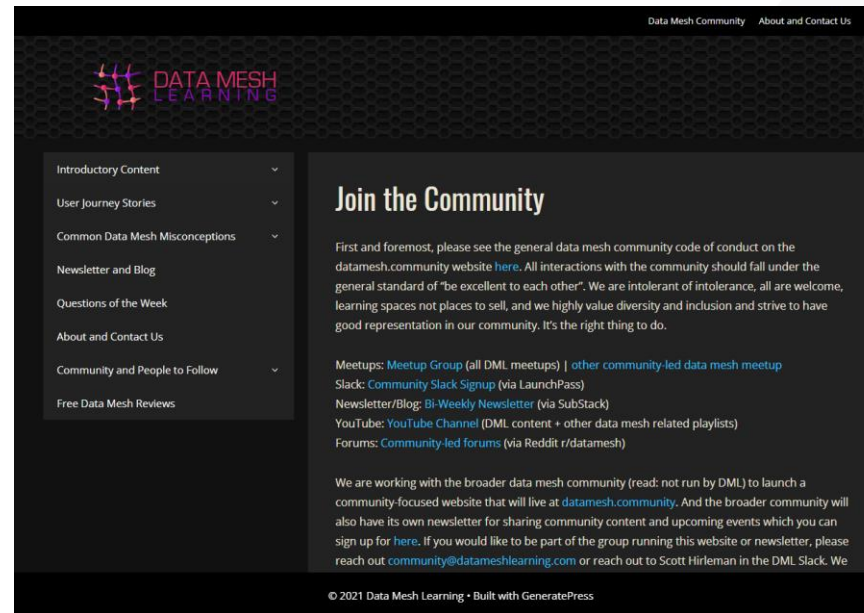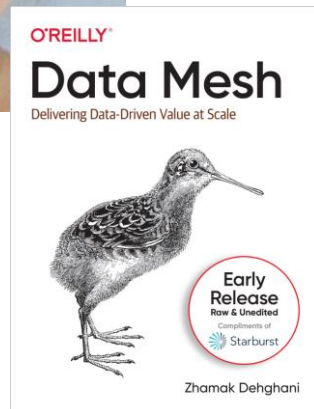# WHAT I'LL COVER

- **Introductions and context**

- **The problem**
  - ○ Why aren't we better at using data?
  - ○ The Swamp and Other Problems
  - ○ The Fourth Industrial Revolution (and Data 4.0)

- **The data mesh**
  - ○ The approach
  - ○ The tech
  - ○ Hype vs reality

- **The FinTech opportunity**
  - ○ The state of financial services data
  - ○ Data Unions

- **Summary and questions**

forthpoint

DATA MESH RESOURCES

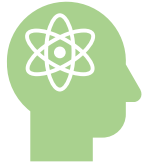## AND SHOUT OUTS

1

INTRODUCTIONS

CAREER TO DATE

FORTH POINT TEAM

MADE UP OF ~~EXPERTS~~ CURIOUS MONKEYS

Data scientists; AI and Machine Learning

Data and cloud engineers; front end and back

Data visualisers; Data story tellers

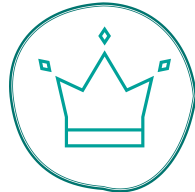Agile delivery leads; Client success
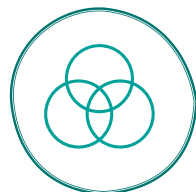
Consultants; translators

MOST AWARDED DATA PROFESSIONALS IN THE UK

Best place to work in data

Data & Analytics leader of the year

Best Diversity program

Best data product service of the year

Best new Talent

Most innovative data management solution

Best data & analytics team

Most innovative use of AI

2

THE PROBLEM

SWAMP
AND OTHER PROBLEMS

PEOPLE PROBLEMS

1. Mostly run on experience, intuition and a bit of spreadsheet work

2. Even those who say they want to be evidence led rarely have the discipline

PROCESS PROBLEMS

1. Risk paralysis

2. Gap between decision makers and data folk

3. Data owners ≠ users ≠ beneficiaries

TECH PROBLEMS

1. Data silos

2. Data swamps

forthpoint

FOURTH INDUSTRIAL REVOLUTION

THE EVOLUTION OF INDUSTRY

Industry 1.0

Mechanization and the introduction of steam and water-power

Industry 2.0

Mass production assembly lines using electrical power
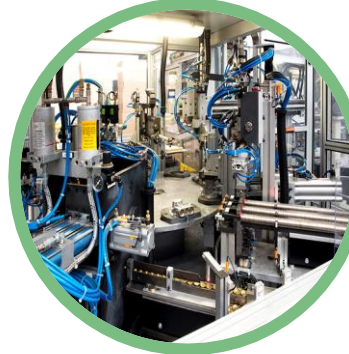
Industry 3.0

Automated production, computers, IT systems and robotics

Industry 4.0

The Smart Factory. Autonomous systems, IoT & machine learning

3

THE DATA MESH

DATA MESH
THE APPROACH
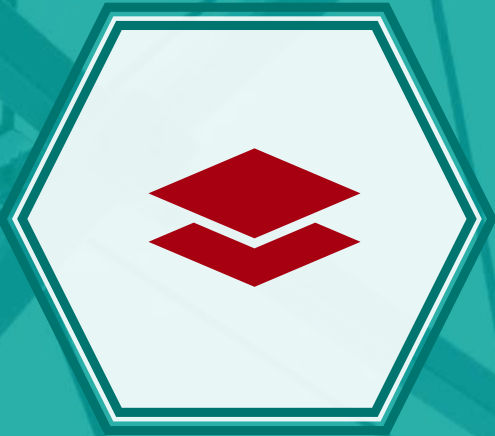
Self serve data infrastructure (data fabric)

Data as a product

Culture and governance
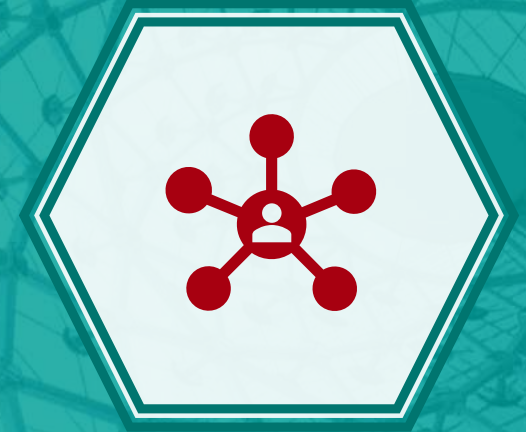
Domain specific/ Distributed

DATA MESH

ADDRESSING DATA OWNERSHIP BY MAKING OWNERS CARE

- **Encryption for data at rest and in motion**

- **Data product versioning**

- **Data product schema**

- **Data product discovery, catalog registration, and publishing**

- **Data governance and standardization**

- **Data production lineage**

- **Data product monitoring, alerting, and logging**

- **Data product quality metrics**

EXAMPLE DATA FABRIC TECHNOLOGY: TRINO

OPEN SOURCE AND ENTERPRISE VERSIONS

A single point of access to all your data

Starburst Enterprise, based on open source Trino (formerly PrestoSQL) is the fastest SQL-based MPP query engine. We give your team a common query tool, abstracted from the systems that store your data, that gives them the ability to run analytics on data where it lives. No data movement or copies required.

Fast.

Lightning fast MPP query performance across all your data, no matter where it lives.

Flexible.

We work in your environment today & tomorrow. We remove the need to move or copy your data.

Open.

Bolstered by the robust Trino open source community, we help companies gain freedom from data lock-in.
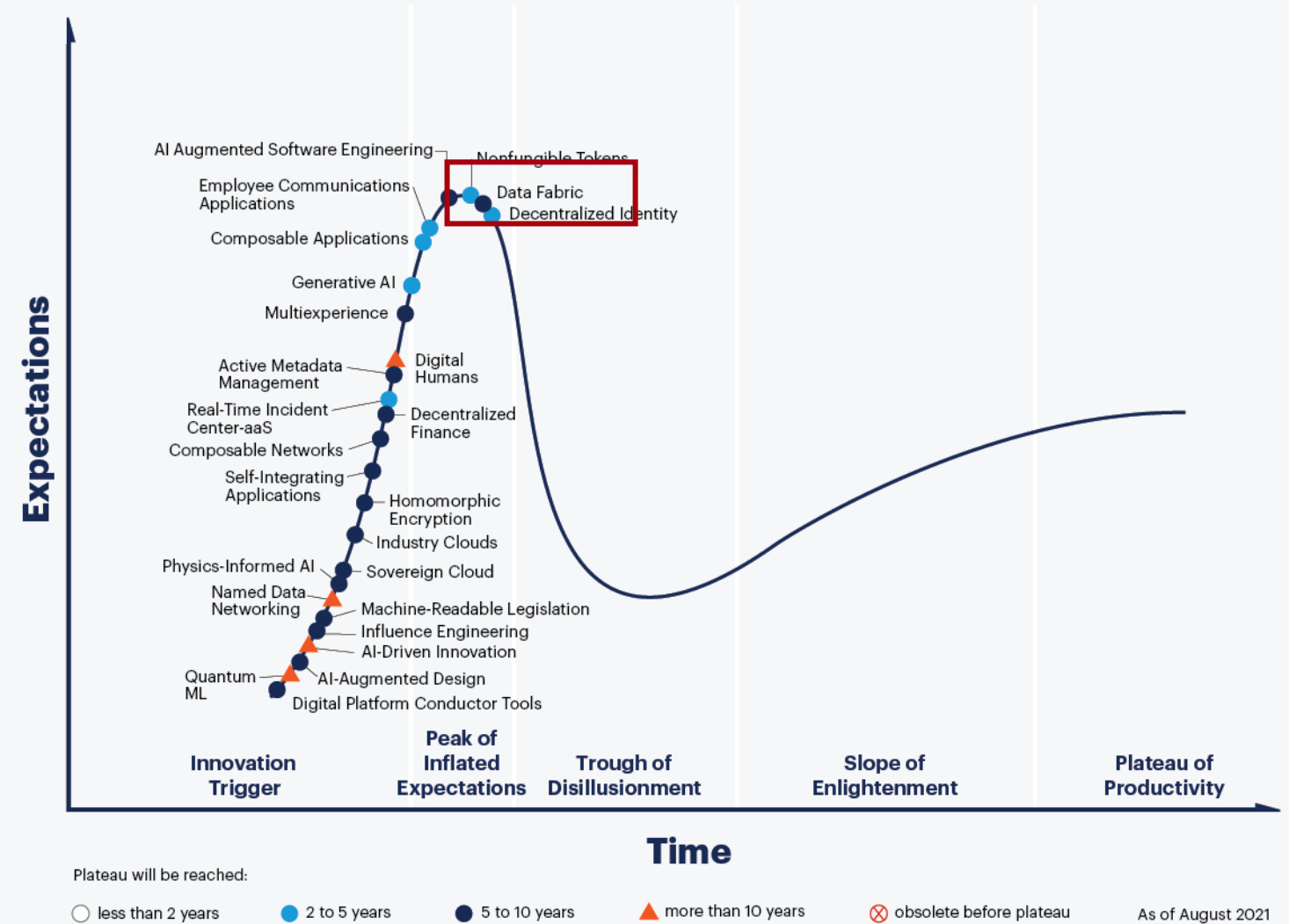
HOW DATA FABRIC ADDRESSES OFFERS TO LET THE DATA REST WHERE IT IS

Data Fabric Layer – Single Point of Access, No Data Movement

| Multiple clouds | Data Lakes | Relational Databases | On Premise |

HOW STARBURST DO IT

Consume / Insights

Data Scientists · Finance · Marketers · Data Analysts

python · Apache Spark · Jupyter · R · DataRobot · dataiku · ThoughtSpot · Looker · Power BI · Qlik Q · +tableau

Analytics Data Layer

Starburst — *Data Fabric Layer — Fast, Scalable, Concurrent — No data move / copy*

Data Masking · Data Encryption · Data Sovereignty · Query Auditing · Global Security · Fine-grained Access Control · Column + Row-level Permissions

Data Sources

Data Lakes
Hadoop · Google Cloud · IBM Cloud Object Storage · DELL EMC ECS · ceph · Azure · BigQuery · MINIO · Delta Lake

Relational Databases
snowflake · ORACLE · IBM DB2 · PostgreSQL · teradata · memSQL · Greenplum · SQL Server · amazon REDSHIFT · Amazon RDS · MySQL

NoSQL Stores
accumulo · elastic · Azure Cosmos DB · mongoDB · redis · cassandra

Publish/Subscribe
kafka · amazon Kinesis · Azure Event Hub

Infrastructure

aws · Azure · Google Cloud · VMware Tanzu · HP Ezmeral · Alibaba Cloud · RED HAT OPENSHIFT

WHO IS DOING DATA MESH?

JPMorgan Chase & Co.

ao

HELLO FRESH

zalando

Disney+

intuit

THERE'S REASON TO BE SCEPTICAL

PEOPLE PROBLEMS

Enormous culture change required to truly embrace

Magic wand-ness causes scepticism

PROCESS PROBLEMS

What's the business case?

Governance, including access and write controls

TECH PROBLEMS

Still a big piping challenge

Scarce data engineering resource

BUT...

CONCEPTUALLY ELEGANT

ADDRESSES REAL PROBLEMS

GAINING TRACTION WITH WINNERS

PLANCK'S PRINCIPLE

A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die and a new generation grows up that is familiar with it. . . .

An important scientific innovation rarely makes its way by gradually winning over and converting its opponents: it rarely happens that Saul becomes Paul.

What does happen is that its opponents gradually die out, and that the growing generation is familiarized with the ideas from the beginning: another instance of the fact that the future lies with the youth.

4

THE FINTECH OPPORTUNITY

FOR FINTECH BUSINESSES THEMSELVES

DATA MESH ✓

DATA FABRIC ?

THE STATE OF FINANCIAL SERVICES DATA

01

OPEN BANKING

02

LEGACY STRATA

03

REGULATORY REQUIREMENTS

04

TRADING RHYTHMS

forthpoint

BANKING 4.0 USE CASES (AND THE REGULATORS) REQUIRE DATA CONSOLIDATION

X ANALYTICS IN FINANCIAL SERVICES
TWO CRITICAL USE CASES, ONE SOLUTION.

01 THE CHALLENGE

- **The BCBS 239 Challenge**
  Banks have a duty to be able to aggregate all of their data to enable them to make well informed risk management decisions.
- **The Vulnerable Customer Challenge**
  Banks have a duty to protect vulnerable customers by identifying them and treating them appropriately.

02 THE ISSUE

Bank data is typically :
- Disparate (across on premise, operational, legacy platforms and in multiple clouds)
- Difficult to aggregate systematically at scale
- Not structured for data science modelling, scoring and simulation

03 THE SOLUTION

X analytics, powered by Starburst

- Leave the data in source systems
- Create a secure query fabric layer
- Build analytical assets
- Low or no disruption to BAU IT

04 THE OUTCOME

- **The BCBS 239 Outcome**
  Demonstrably timely and comprehensive access to all relevant data & the ability to analyse and query that data to make better informed decisions.
- **The Vulnerable Customer Outcome**
  The ability to build targeting models on all data to proactively identify customers at risk of being vulnerable and then output appropriate scoring into CRM and decisioning platforms such as Adobe, Salesforce and Pega for treatment programmes.

www.starburst.io | www.forthpoint.com

Paper reduction

Complaints attribution

Accessibility and channel personalization

New definitions of wealth and affordability

Individual empowerment through Open Banking

Vulnerable customer identification and treatment

BCBS 239 compliance

Dynamic and federated dashboarding

X analytics – breaking the data silos

External environment and scenario forecasting and simulation

Brand review

Primary relationship attack/defend

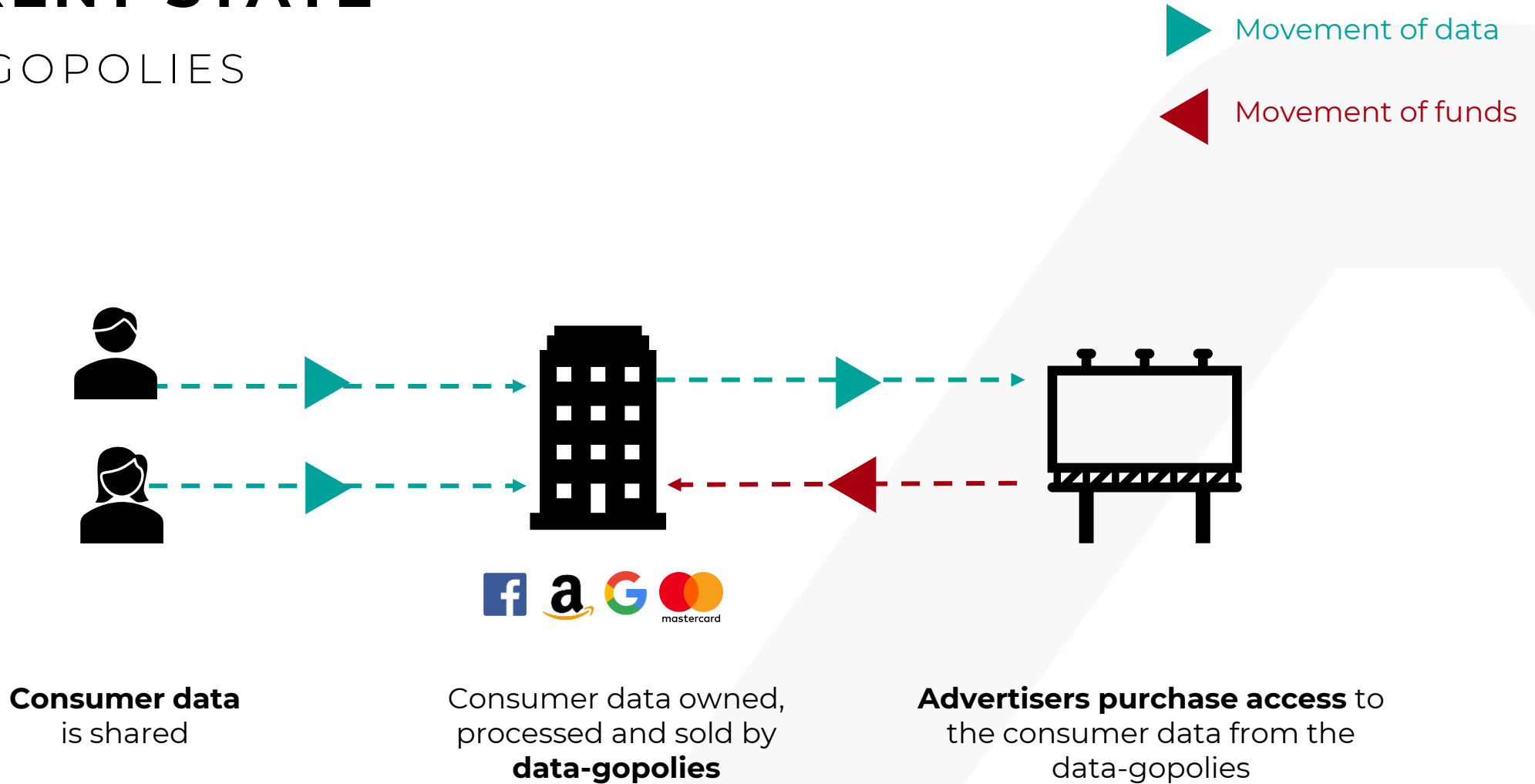Product personalisation

BCBS 239 PRINCIPLES

Principle 2 *Data architecture and IT infrastructure* – A bank should design, build and maintain data architecture and IT infrastructure which fully supports its risk data aggregation capabilities and risk reporting practices not only in normal times but also during times of stress or crisis, while still meeting the other Principles.

Principle 5 *Timeliness* – A bank should be able to generate aggregate and up-to-date risk data in a timely manner while also meeting the principles relating to accuracy and integrity, completeness and adaptability. The precise timing will depend upon the nature and potential volatility of the risk being measured as well as its criticality to the overall risk profile of the bank. The precise timing will also depend on the bank-specific frequency requirements for risk management reporting, under both normal and stress/crisis situations, set based on the characteristics and overall risk profile of the bank.

Principle 8 *Comprehensiveness* - Risk management reports should cover all material risk areas within the organisation. The depth and scope of these reports should be consistent with the size and complexity of the bank's operations and risk profile, as well as the requirements of the recipients.
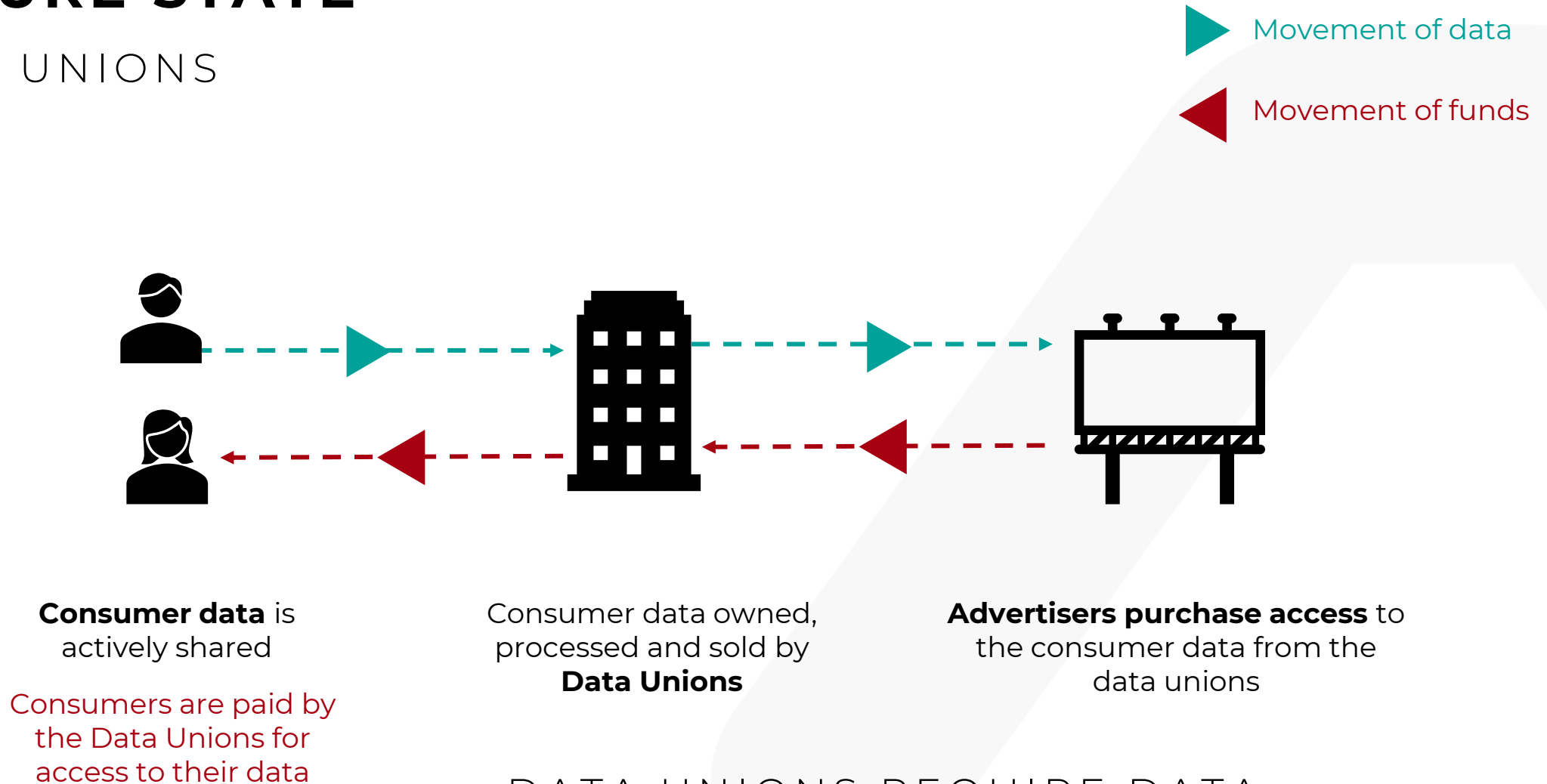
FUTURE STATE

DATA UNIONS

Consumer data is actively shared

Consumers are paid by the Data Unions for access to their data

Consumer data owned, processed and sold by **Data Unions**

Advertisers purchase access to the consumer data from the data unions

DATA UNIONS REQUIRE DATA MESH TYPE THINKING AND TECH

5

SUMMARY AND DISCUSSION