Training AI/ML models using Digital Data Marketplaces

2nd project meeting
November 28th 2018, Coimbra

Leon Gommans, Anne Savelkoul, Wouter Kalfsbeek, Dirk van den Herik, David Langerveld, Erik IJzermans, Floris Freeman, Brend Dikkers, Cees de Laat, Tom van Engers, Wouter Los, Paola Grosso, Joseph Hill, Reggie Cushing, Giovanni Sileno, Lu Zhang, Ameneh Deljoo, Thomas Baeck, Willem Koeman, Laurie Strom, Axel Berg, Gerben van Malenstein, Kaladhar Voruganti, Rodney Wilson, Patricia Florissi

This project has received funding from the European Union's Horizon 2020 research and innocvation programme under grant agreement No 769288

BUSINESS CONTEXT

Companies increasingly understand how to apply AI technologies to extract business value from data.

The more data the better: algorithm quality depends on data quantity and quality **Knowledge** how to translate such data into reliable algorithms **is competitive**

Companies are reluctant to share data when considering involved risk.

Emerging platform dominance: *"While creating real value for users, these companies are also capturing a **disproportionate and expanding share of the value**, and that 's shaping our collective economic future". **

Sharing data across companies increases the potential of creating business value no single organization can create on its own.

* M. Iansiti, K.R. Lakhani, *Managing our hub economy,* Harvard Business Review, pg. 85-92, Sep/Oct 2017

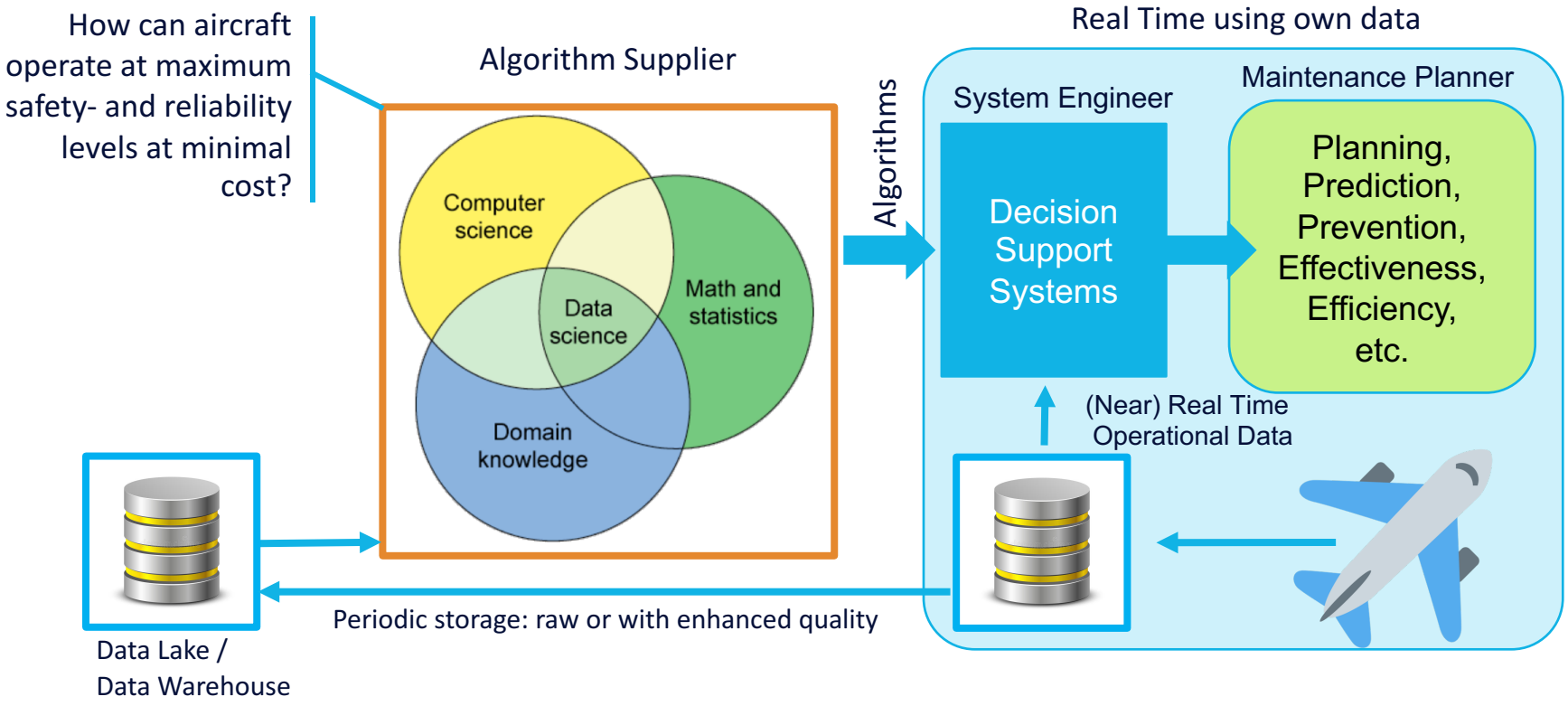
DATA IS INCREASINGLY CONSIDERED AN ASSET

Considering value exchange and involved risk raises the main research question:

How can (big) data assets be shared between data suppliers and algorithms developers in

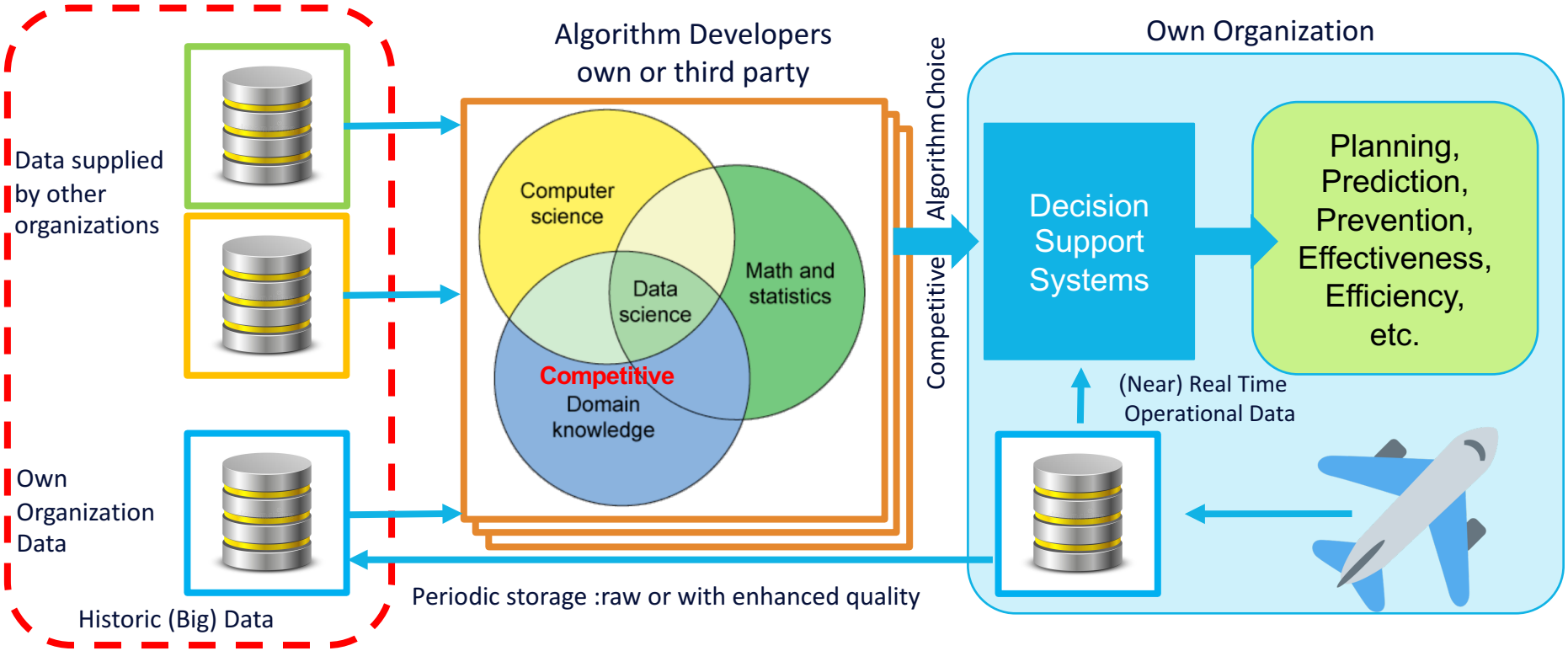
1) ***A fair and economic way,***
2) ***whilst providing adequate means to reduce risk?***

CURRENT ALGORITHM DEVELOPMENT CONTEXT



RESEARCH CONTEXT

ARRANGE ADDITIONAL DATA TO IMPROVE ALGORITHM QUALITY & INNOVATION



B2B DATA SHARING APPROACHES

AN EU STUDY BY EVERIS JAN 2018

Case studies | Approaches to B2B data sharing

European Commission

everis an NTT DATA Company

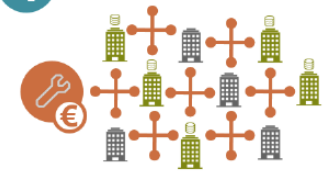
Five different approaches to B2B data sharing

1 DATA MONETISATION

2 DATA MARKETPLACES

3 INDUSTRIAL DATA PLATFORMS

4 TECHNICAL ENABLERS



5 OPEN DATA

Open vs Closed

DATA MARKETPLACES

- ✓ Trusted intermediary between data suppliers and data users
- ✓ Data suppliers sell their data to interested data users
- ✓ Revenue is generated from each data transaction

DAWEX

INDUSTRIAL DATA PLATFORMS

- ✓ Strategic and collaborative partnerships
- ✓ Mutual benefits for all parties
- ✓ Data shared (for free) in a closed, exclusive and secure environment
- ✓ Develop new or improved products and/or services
- ✓ Enhance internal performance

AIRBUS MAN

Difference with 2
Governance by a **membership organization**
Difference with 3:
Data is stored **outside** data platforms to allow multiple platforms to use same data
Contracts determine access & use
Membership rules: compliancy, liability, arbitration,…

DATA SHARING CHALLENGES

WHEN TRAINING MODELS WITH AS MUCH DATA AS POSSIBLE

Many organizations want to keep their historical data in their sovereign data zones.

Many implications need to be considered:

Business level	Legal level	Data level
Value	Ownership	Processing
Cost	Access	Storage
Benefits	Usage	Management
Agreements	Compliancy	Transport
Exchange	Liability	Transform
Trade	Market Rules	Security

OVERCOMMING CHALLENGES

ELEMENTS TO ORGANIZE TRUST AS MEANS TO REDUCE RISK

COMMON BENEFIT

Define and agree common benefit no single organization can achieve on its own.

GROUP RULES

Define consortium rules considering data use, access and benefit sharing

ORGANIZE TRUST

Organize power and trust as a means to reduce risk for participating members

IMPLEMENT INFRASTRUCTURE

Research operationalization of **Digital Data Marketplace & Data Exchange** concepts

EXCHANGEWELL™
A Program of SAE ITC

INTRODUCTION

- Organized by SAE ITC, **ExchangeWell** brings data owners and algorithm developers together in a digital data marketplace that provides the required trust for mutual engagement.
- It enables members to share their data assets in a **fair and economic way** whilst providing an adequate means to **reduce risk**.
- Sharing data enables **digital transformation of the industry** and **business value creation.**

Objective: Help answer key question:

- Will ExchangeWell as proposed provide value to our industry?

COMMON BENEFIT

GROUP RULES

ORGANIZE TRUST

IMPLEMENT INFRASTRUCTURE

DEFINE AND AGREE COMMON BENEFIT

Example: enable data sharing to improve quality of AI/ML innovations

- Understand need: the more data the better
- Expect: capability that will help transform the MRO business in the digital era.

Innovations that will improve air safety, passenger experience and additional cost reductions by:

- avoiding unplanned maintenance
- increasing maintenance planning flexibility
- moving from fixed interval planning to maintenance when indicated
- less network disruptions by avoiding 'Aircraft On Ground' situations

CONSORTIUM MEMBERSHIP RULES:

WHAT KIND OF RULES DO WE NEED?

Trust is considered as a means to reduce risk

Defining consortium membership rules is a starting point

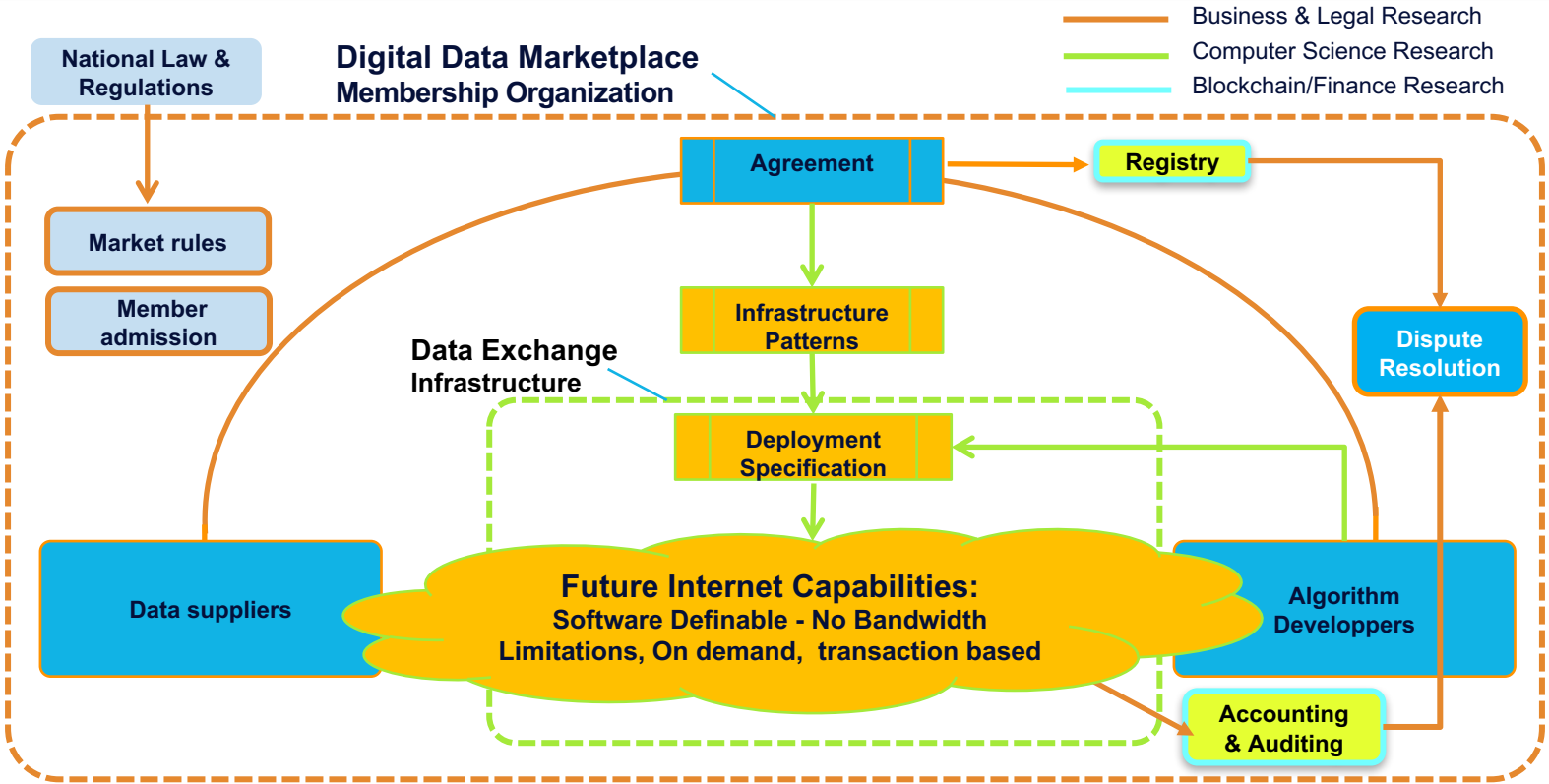
Legal research topic's for discussion:

- Data asset ownership
- Data access & usage
- Liability of owner & user
- Non-compliant behavior
- Market rules

EXCHANGEWELL™
A Program of SAE ITC

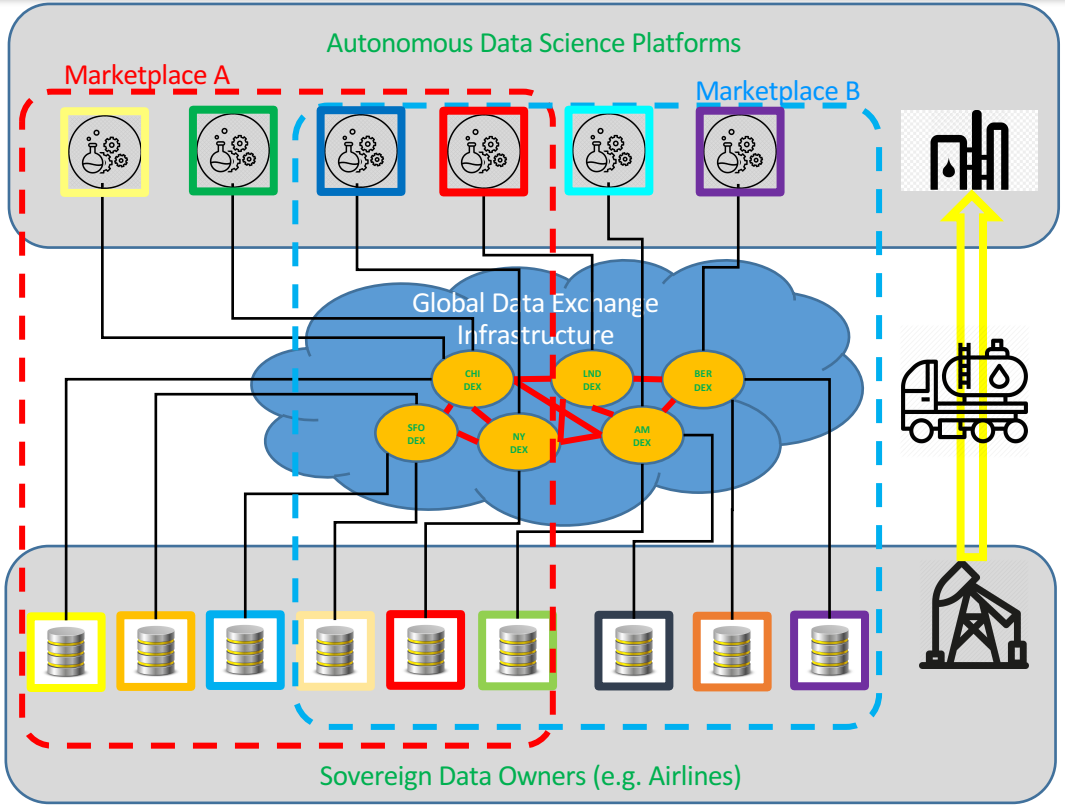
DIGITAL DATA MARKETPLACE CONCEPT:

COMBINED BUSINESS, LEGAL AND COMPUTER SCIENCE RESEARCH



DATA EXCHANGE CONCEPT

ENVISAGED GLOBAL EXCHANGE INFRASTRUCTURE

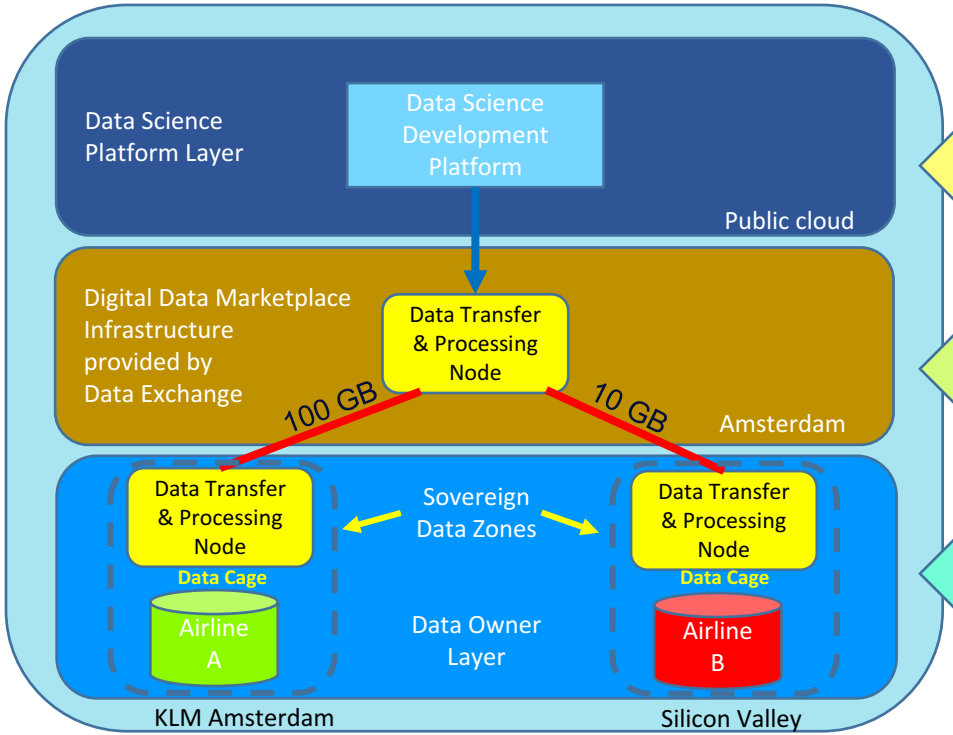


EQUINIX

amsterdam economic board

AMdEX
THE DATA HYPERMARKET

RESEARCHING EXCHANGE ARCHITECTURES

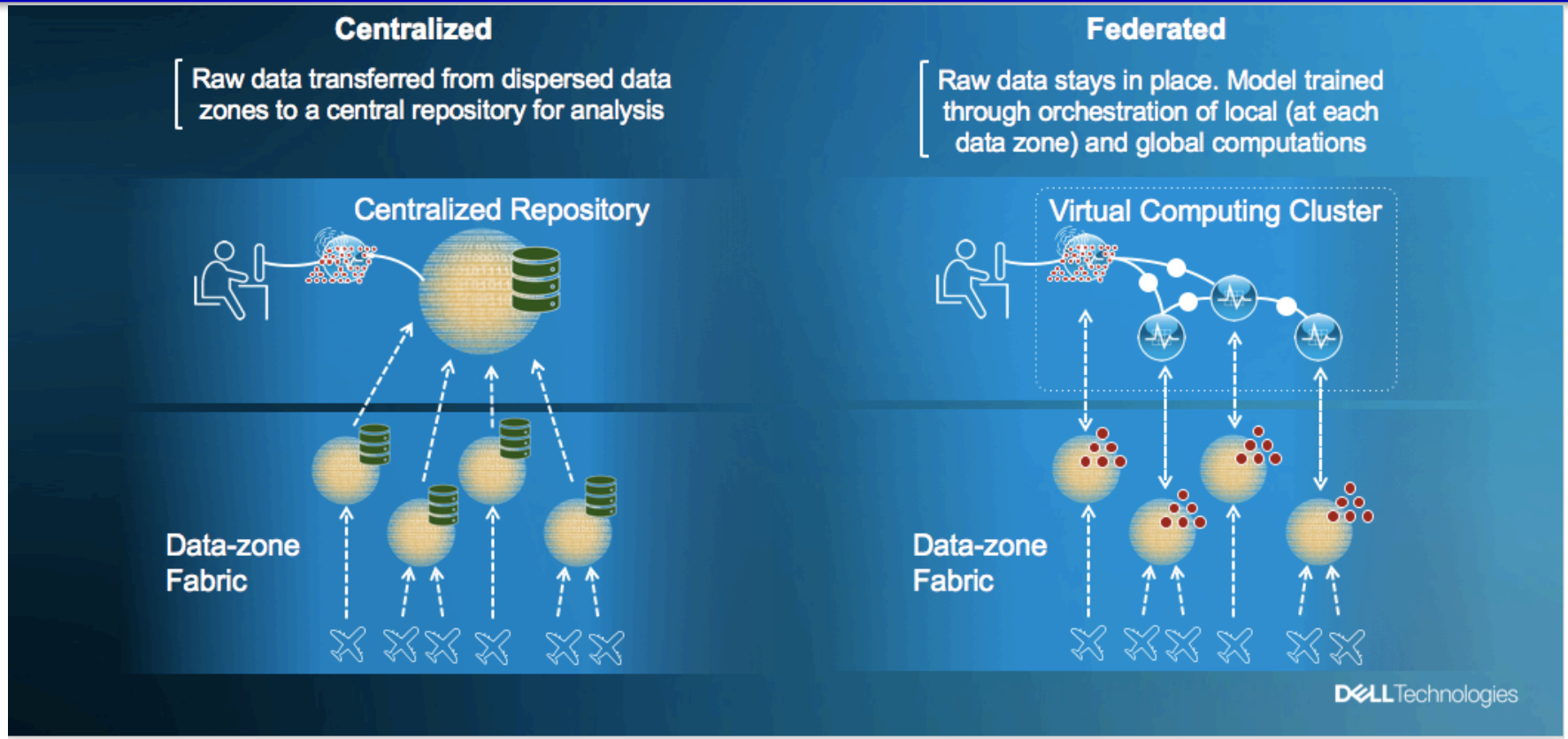


Trust Modelling:
What is the optimal infrastructure archetype, describing storage and processing locations and their relationships, which best suit member requirements when considering risk?

Processing Models:
What are the implications of distributing data processing across membership organization owned infrastructures in terms of achievable model accuracy and processing performance using federated/distributed models vs centralized models, whilst avoiding as much a possible complexity for the data scientist.

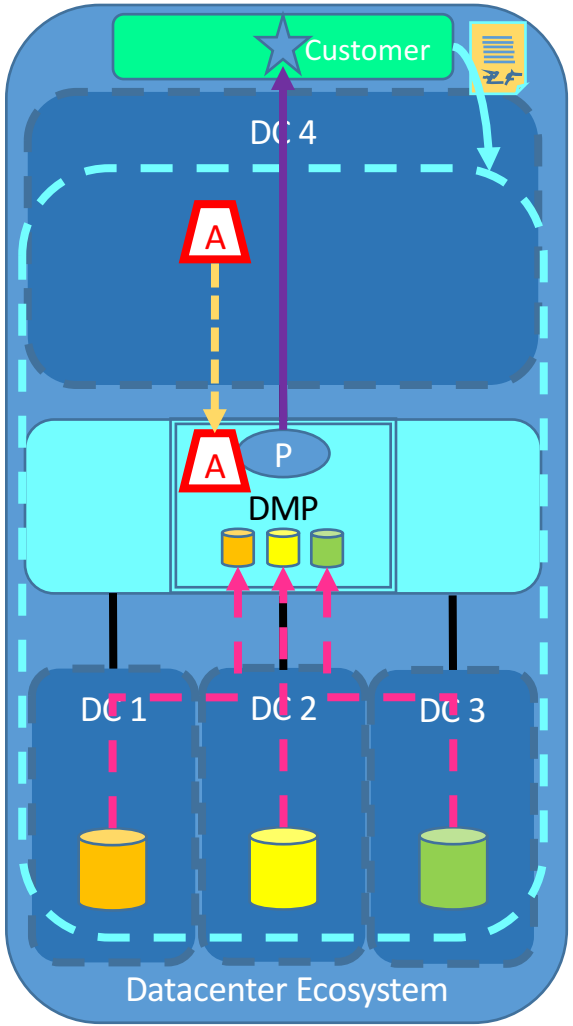
Marketplace Reference Architecture:
What constitutes a marketplace? Researching needed functions, personas, flows, credentials, contracts & rules, conflict resolution, and much more …

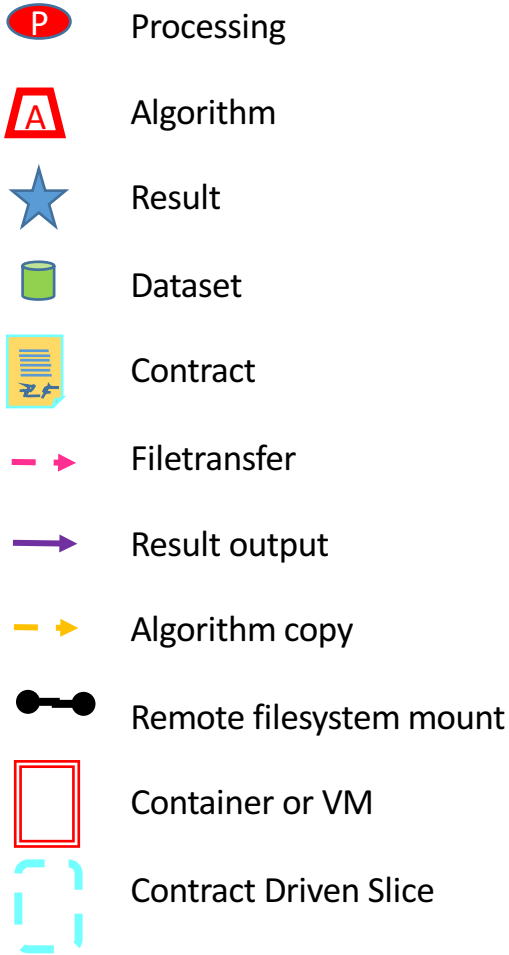
PROCESSING & STORAGE: TRAINING STRATEGIES

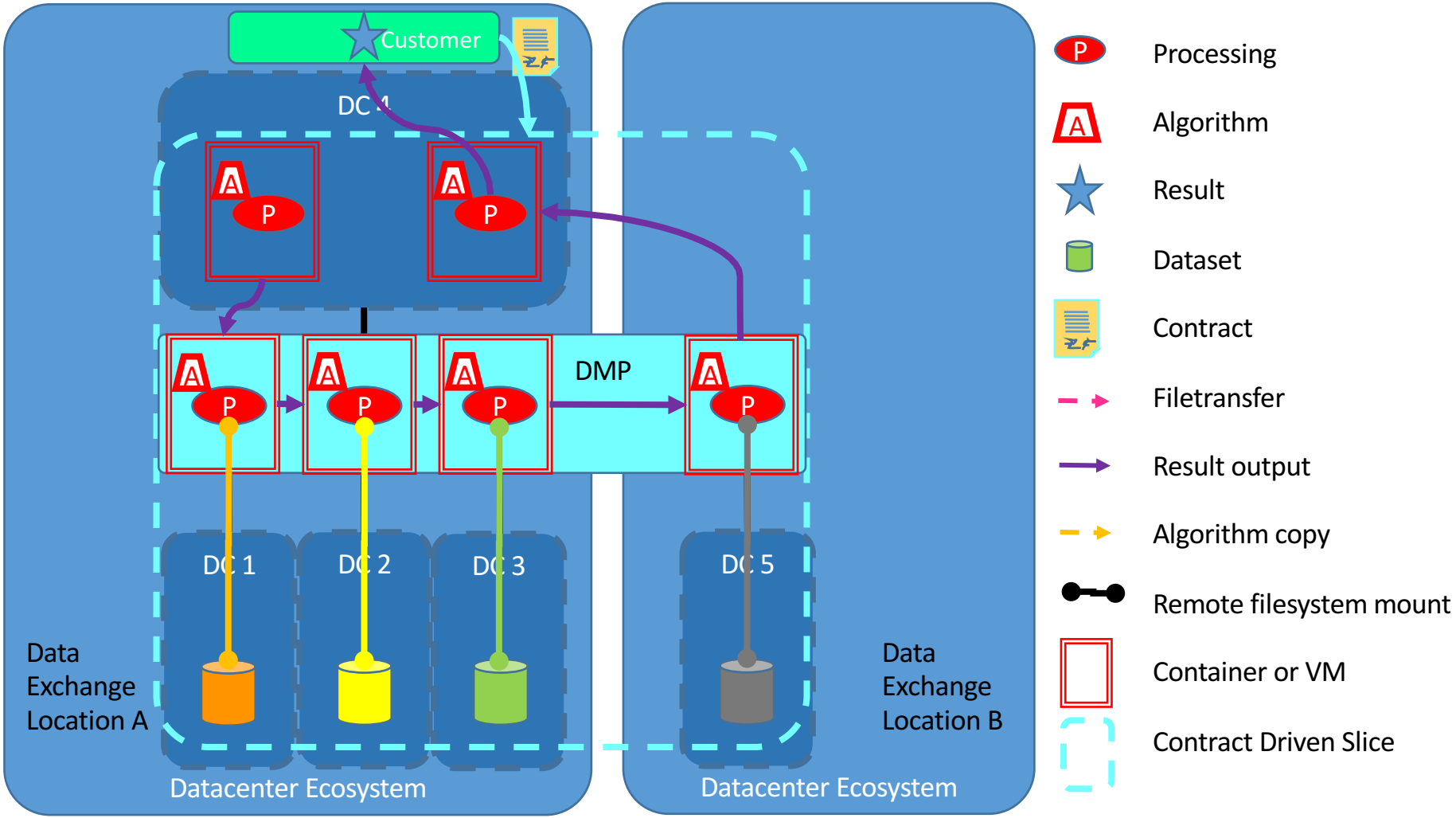


DMP provides neutral processing capabilities, which dissolves after Execution.

Note: multiple incarnations possible depending on need to have all data available at execution time or can be made sequentially available e.g. storing intermediate results.






Customer
DC 4
A
P
A
P
A
P
A
P
A
P
DMP
DC 1
DC 2
DC 3
DC 5
Data
Exchange
Location A
Datacenter Ecosystem
Data
Exchange
Location B
Datacenter Ecosystem
Processing
Algorithm
Result
Dataset
Contract
Filetransfer
Result output
Algorithm copy
Remote filesystem mount
Container or VM
Contract Driven Slice

SUMMARY

Enterprises join a membership organization to achieve a common goal *no single enterprise can achieve on its own*

Membership rules are defined by rulemaking & standards processes, subsequently execution, enforcement and judgement is organized by membership organization as *a means to reduce risk.*

Members arrange data sharing and processing via *agreements deployed in an infrastructure*, provided by a secure digital market place owned by its members.

Members *achieve common benefits in a transparent way*. Members trust its operation based on use of accounting & auditing mechanisms, relying on market dispute resolution mechanisms.