Training AI/ML models using Digital Data Marketplaces

November 12-15th 2018, Dallas TX

SURF Exhibition Booth #2041

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

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

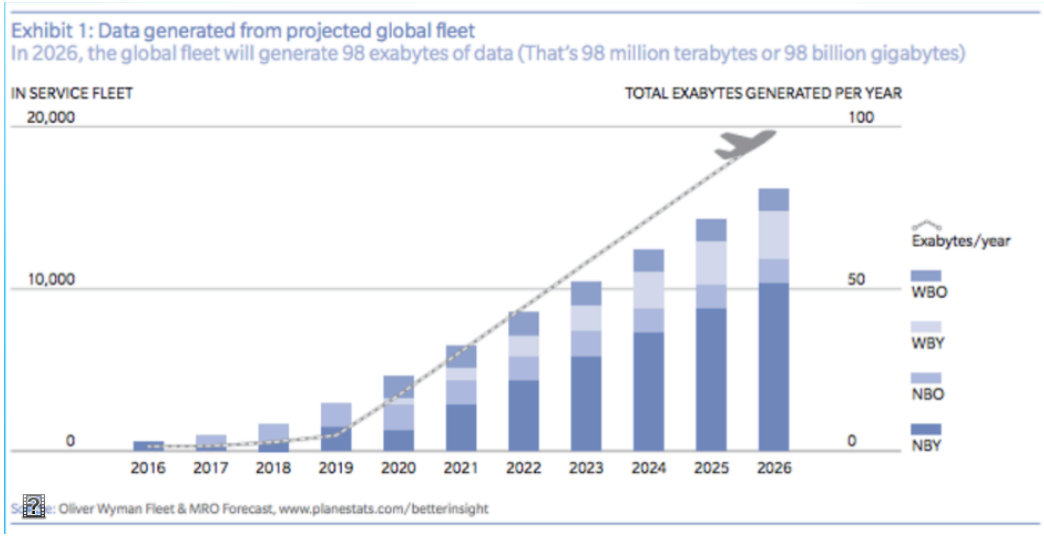
Companies are reluctant to share data when considering involved risk.

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.

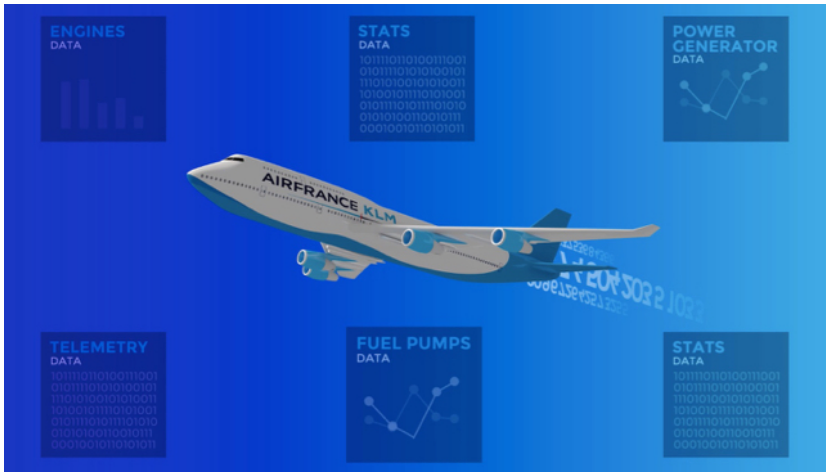
* M. Iansiti, K.R. Lakhani, *Managing our hub economy,* Harvard Business Review, pg. 85-92, Sep/Oct 2017

EXPECTED VOLUMES
OF AIRCRAFT DATA



"Airline operators own the operational data"

Oliver Wyman



DATA IS INCREASINGLY CONSIDERED AN ASSET

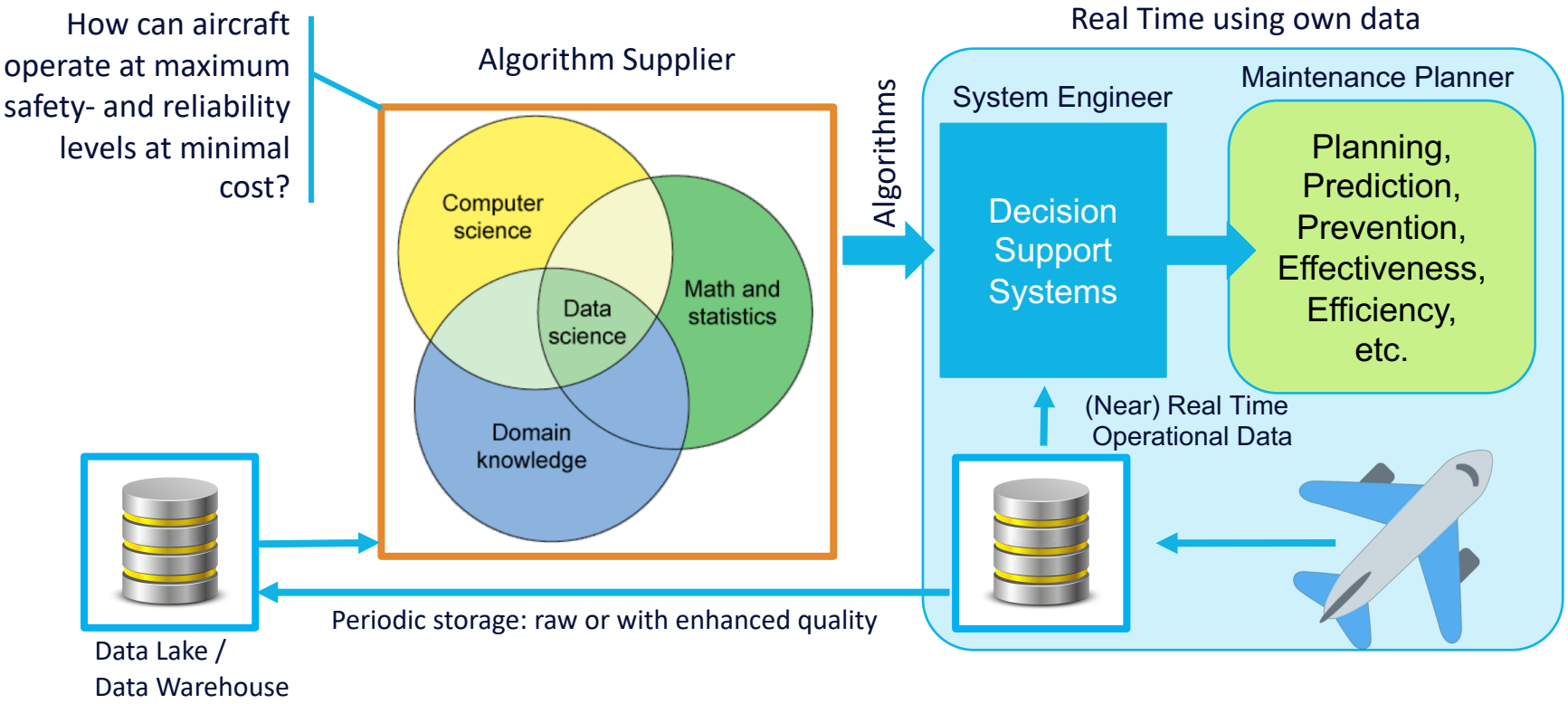
Considering value exchange and involved risk raises the main research question:

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?***

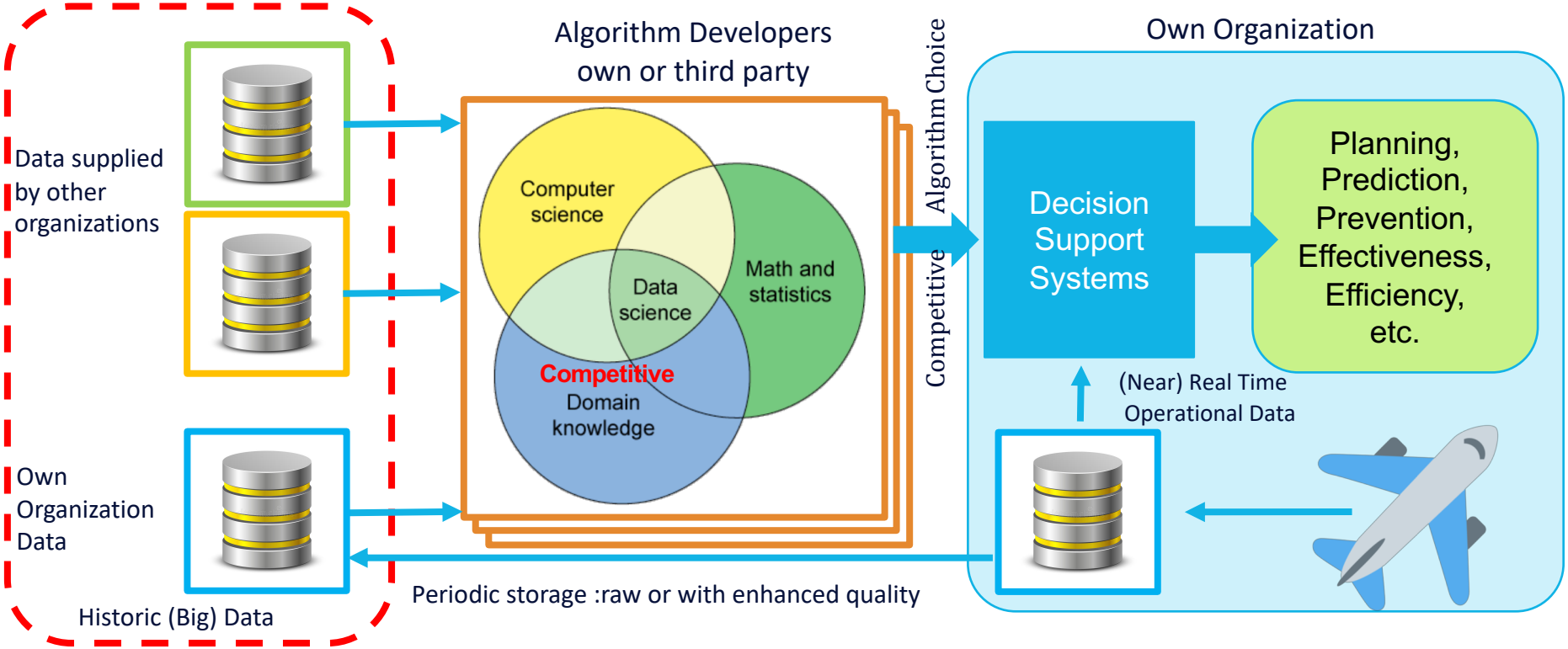
CURRENT ALGORITHM DEVELOPMENT CONTEXT

How can aircraft operate at maximum safety- and reliability levels at minimal cost?



RESEARCH CONTEXT

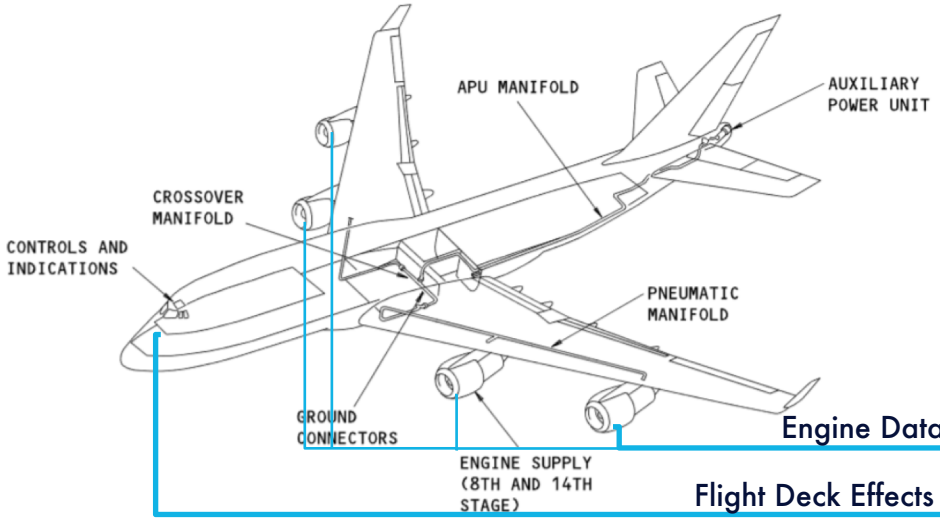
ARRANGE ADDITIONAL DATA TO IMPROVE ALGORITHM QUALITY & INNOVATION



USE CASE:BLEED AIR SYSTEM

RELATING ENGINE DATA TO EVENTS SIGNALLED AT THE FLIGHT DECK

Imagine if data scientist can use historic data from 747 aircraft operated by multiple airlines..

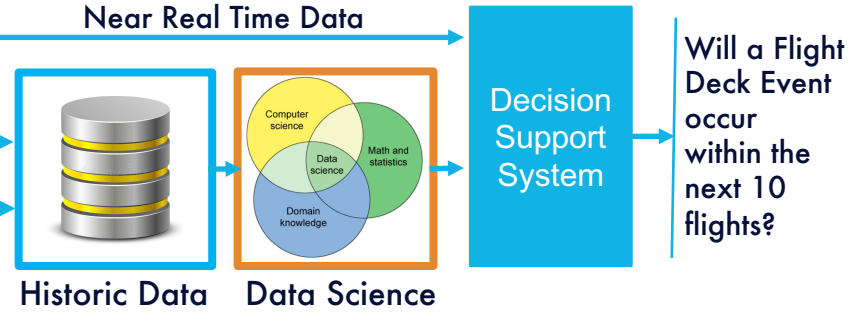


The more Flight Deck Effect occurrences are available, the more likely that a prognostic relation can be learnt

The Bleed Air System regulates pressure and temperature of air from a turbine engine needed by other aircraft systems taking care of:

- cabin pressure
- de-icing
- water pressure
- and more..

Flight Deck Effects indicate system functionality decreases and may trigger maintenance actions.



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 Cost Benefits Agreements Exchange Trade	Ownership Access Usage Compliancy Liability Market Rules	Processing Storage Management Transport Transform Security

MANAGING RISK

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

DEFINE AND AGREE COMMON BENEFIT

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

- Understand need: the more data the better
- Expect: capability that will help transform the 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

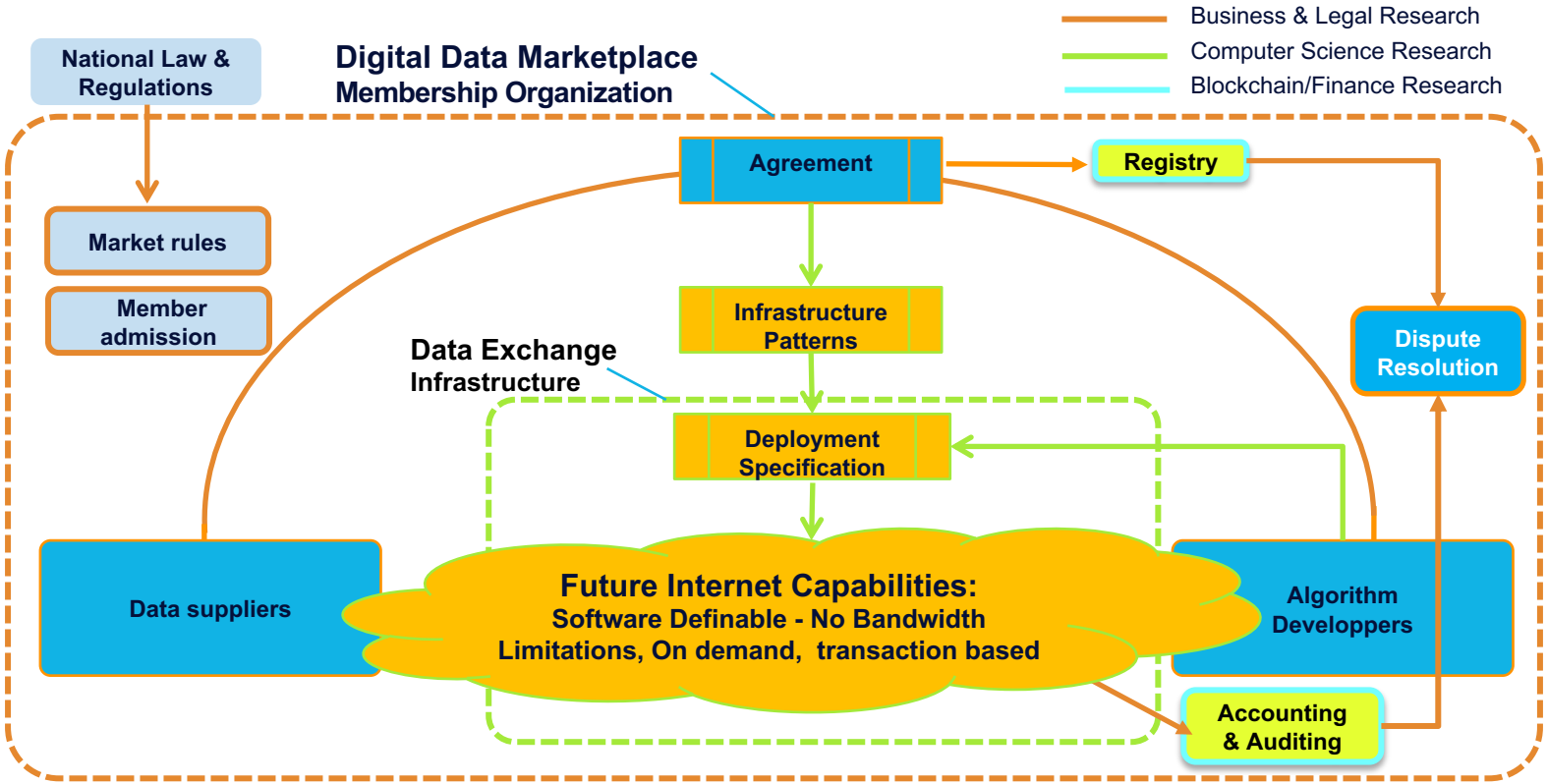
Legal research topic's for discussion:

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

EXCHANGEWELL™
A Program of SAE ITC

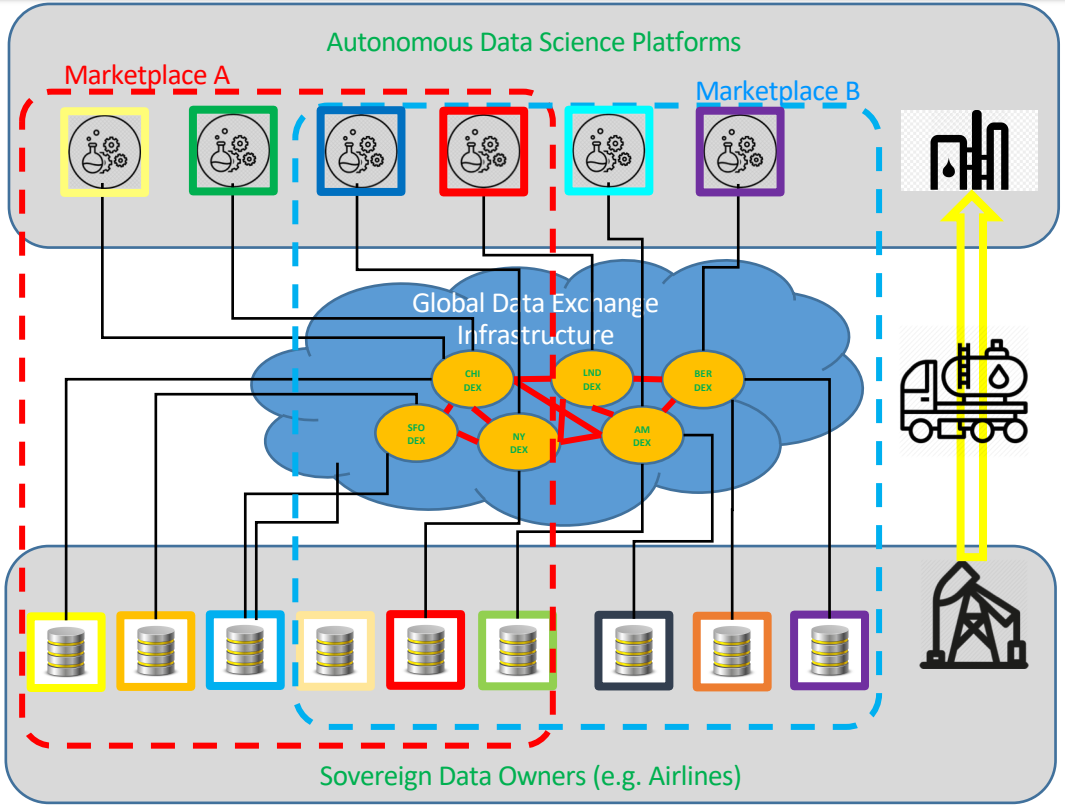
DIGITAL DATA MARKETPLACE CONCEPT:

COMBINED BUSINESS, LEGAL AND COMPUTER SCIENCE RESEARCH



DATA EXCHANGE CONCEPT

ENVISAGED GLOBAL EXCHANGE INFRASTRUCTURE

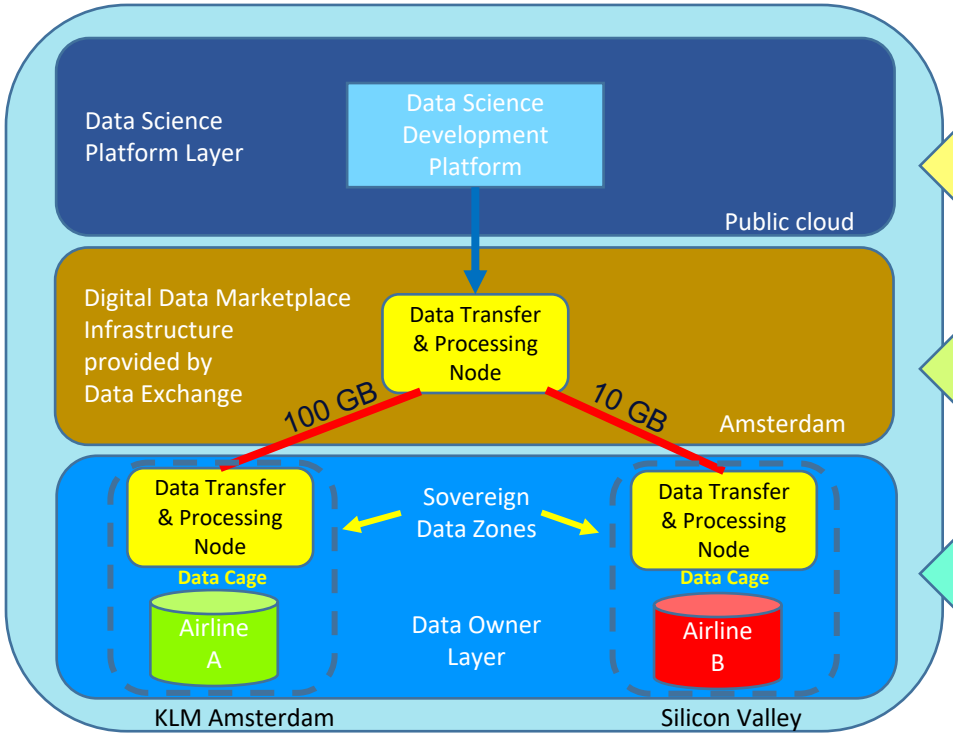


amsterdam economic board

AMdEX
THE DATA HYPERMARKET

RESEARCHING EXCHANGE ARCHITECTURES

ALSO SEE CIENA BOOTH #2847



Trust Modelling:

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

See CIENA booth 2847 and demo

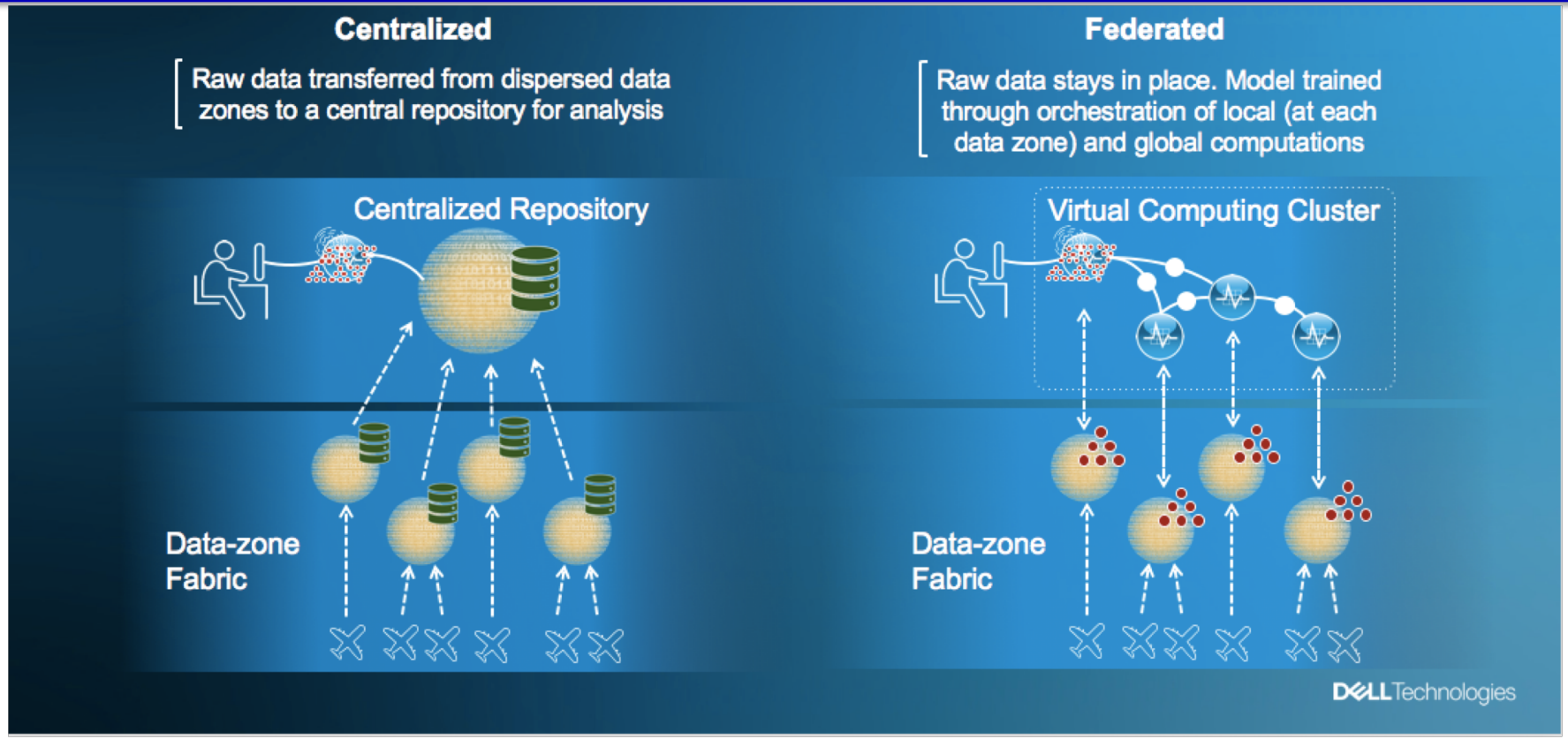
Processing Models:

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

Marketplace Reference Architecture:

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

PROCESSING & STORAGE: TRAINING STRATEGIES



COMPARING TRAINING STRATEGIES

- → Historical training data randomly split in 3 data zones
- → Random Forest classifier
- → Same hyper-parameters and features used in centralized and federated training
- → Both models tested on the same test set

Centralized Training

Single random forest built on combined data from the 3 data zones

0.19	0.59
Average Test Set Precision	Test Set Area Under the ROC Curve

Federated Training

Local random forests built in each data zone and combined to form a global random forest

0.19	0.60
Average Test Set Precision	Test Set Area Under the ROC Curve

Federated accuracy performance close to centralized accuracy performance while minimizing data movement

DELL Technologies

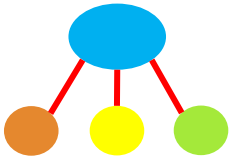
RESEARCHING PHYSICAL IMPLEMENTATION

INVOLVING BOTH RESEARCH AND IT INDUSTRY

GLOBAL RESEARCH INFRASTRUCTURES

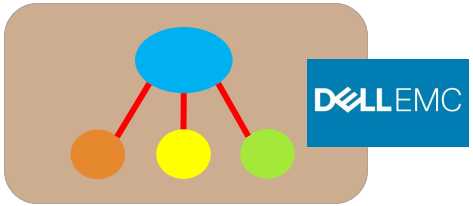
Data Sharing Infrastructure Model Research using Future Internet capabilites

UNIVERSITEIT VAN AMSTERDAM



Research collaboration

GLOBAL DATACENTER INFRASTRUCTURES

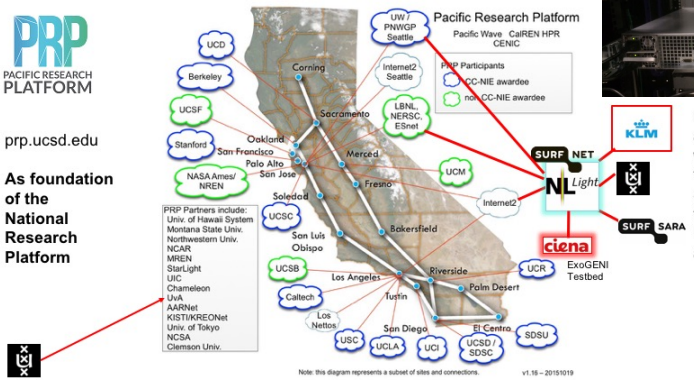


How to create a Global Digital Data Market Ecosystem via Data Exchanges

PRP PACIFIC RESEARCH PLATFORM

prp.ucsd.edu

As foundation of the National Research Platform



Data Transfer Node at KLM fieldlab with 100 gb/s link to enable SDMP research thanks to UvA, SURFnet and Ciena

AM3 and AM4 Datacenters Amsterdam Science Park

SV10 Datacenter Silicon Valley

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.