Improving AI models using Digital Data Marketplaces

Leon Gommans, PhD

Air France KLM Group IT Technology Office / R&D
Guest Researcher University of Amsterdam / Systems & Network Engineering Lab
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”. *

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?

Algorithm Supplier

Data Lake / Data Warehouse

Periodic storage: raw or with enhanced quality

System Engineer

Decision Support Systems

Maintenance Planner

Planning, Prediction, Prevention, Effectiveness, Efficiency, etc.

Real Time using own data

(Near) Real Time Operational Data

Computer science

Data science

Math and statistics

Domain knowledge

Algorithm Supplier

Real Time using own data

How can aircraft operate at maximum safety- and reliability levels at minimal cost?
RESEARCH CONTEXT
ARRANGE ADDITIONAL DATA TO IMPROVE ALGORITHM QUALITY & INNOVATION

Data Exchange

Algorithm Developers
- own or third party

- Computer science
- Math and statistics
- Data science
- Competitive Domain knowledge

Own Organization

Planning, Prediction, Prevention, Effectiveness, Efficiency, etc.

Decision Support Systems

(Near) Real Time Operational Data

Own Organization Data

Data supplied by other organizations

Historic (Big) Data

Periodic storage: raw or with enhanced quality
Five different approaches to B2B data sharing

1. DATA MONETISATION
   - Data supplier
   - Data user

2. DATA MARKETPLACES
   - Data marketplace

3. INDUSTRIAL DATA PLATFORMS
   - Trusted intermediary between data suppliers and data users
   - Data suppliers sell their data to interested data users
   - Revenue is generated from each data transaction

4. TECHNICAL ENABLERS
   - Infrastructure and protocols

5. OPEN DATA
   - Open data standards and formats

Open vs Closed

Difference with Data Marketplaces:
- Governance by a membership organization
- Data is stored outside data platforms to allow multiple platforms to use the same data
- Contracts determine access / use
- Market rules arrange pre-contractual elements
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**
  - Value
  - Cost
  - Benefits
  - Agreements
  - Exchange
  - Trade

- **Legal level**
  - Ownership
  - Access
  - Usage
  - Compliancy
  - Liability
  - Market Rules

- **Data level**
  - Processing
  - Storage
  - Management
  - Transport
  - Transform
  - Security

Worldwide Scale
<table>
<thead>
<tr>
<th>COMMON BENEFIT</th>
<th>GROUP RULES</th>
<th>ORGANIZE TRUST</th>
<th>IMPLEMENT INFRASTRUCTURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Define and agree common benefit no single organization can achieve on its own.</td>
<td>Define consortium rules considering data use, access and benefit sharing</td>
<td>Organize power and trust as a means to reduce risk for participating members</td>
<td>Research operationalization of Digital Data Marketplace &amp; Data Exchange concepts</td>
</tr>
</tbody>
</table>
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?
EXCHANGEWELL™
A Program of SAE ITC

A consortium program to provide the means for industry leaders to access industry experts, develop practical experience from pilots, collaborate on pre-competitive research and establish a strategic path forward to effectively implement data management strategies which positively impact and benefit industry.
SAE ITC

CONSORTIUM MATURATION PATH

We're here!
ExchangeWell

IDEA
EXPLORATORY CONVERSATION
CHAMPIONS
SUFFICIENT INTEREST
DISCUSSION OR STEERING GROUP
SITUATION/TARGET PROPOSAL
ROI/VALUE EQUATION
CONSORTIUM AGREEMENT for STRATEGY GROUP

We're here!
ARINC IA
STAKEHOLDERS

Stakeholders
- Regulatory
- Airline/ Operator
- Airframer/ Integrator
- OEM
- Sub Assembly Manufacturer
- Distributor
- Component/ Part Manufacturer
- Standards Organization
- Industry Review Body
- Auditor/ Mandated Body
- SAE ITC
- Registrar
- Maintenance
- Training Provider
- IT System and Software Tools Provider
- Data Aggregators and Analyzers
- Insurers
- Legal
- Access Authorizing Agent
- Research/ Academics

INTEGRATED PRODUCT CYCLE

Design/ Development

Recycle/ Sunset

Manufacture/ Inspect

Maintenance/ Repair

Integrate

Operations
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
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
- Purpose binding
DIGITAL DATA MARKETPLACE CONCEPT: COMBINED BUSINESS, LEGAL AND COMPUTER SCIENCE RESEARCH

Future Internet Capabilities: Software Definable - No Bandwidth Limitations, On demand, transaction based

- Market rules
- Member admission
- Data suppliers
- Algorithm Developpers
- National Law & Regulations
- Business & Legal Research
- Computer Science Research
- Blockchain/Finance Research
- Dispute Resolution
- Accounting & Auditing

Agreement
Registry
Infrastructure Patterns
Deployment Specification
Data Exchange Infrastructure
DATA EXCHANGE CONCEPT
ENVISAGED GLOBAL EXCHANGE INFRASTRUCTURE

Sovereign Data Owners (e.g. Airlines)

Global Data Exchange Infrastructure

AMdEX

Autonomous Data Science Platforms

Marketplace A

Marketplace B
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? 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 ...

Data Science Platform Layer

Digital Data Marketplace Infrastructure provided by Data Exchange

Data Science Development Platform

Public cloud

Data Transfer & Processing Node

Data Cage

Airline A

KLM Amsterdam

100 GB

10 GB

Amsterdam

Data Owner Layer

Sovereign Data Zones

Airline B

Silicon Valley

Data Transfer & Processing Node

Data Cage

Airline B

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 ...

Data Science Platform Layer
Data Sharing Infrastructure Model Research using Future Internet capabilities

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

GLOBAL RESEARCH INFRASTRUCTURES

GLOBAL DATACENTER INFRASTRUCTURES

Research collaboration

AM3 and AM4 Datacenters
Amsterdam Science Park
SV10 Datacenter
Silicon Valley
Traditional model

DC4 acts as platform:
1: creates potential competitive bottleneck / lock-in.
and
2: raises data owner concerns about risk
DMP provides neutral processing capabilities, which dissolves after Execution.
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.