Training AI models using Digital Data Marketplaces

Leon Gommans
Science Officer 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”. *


Sharing data across companies increases the potential of creating business value no single organization can create on its own.
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?*
How can aircraft operate at maximum safety- and reliability levels at minimal cost?
RESEARCH CONTEXT

ARRANGE ADDITIONAL DATA TO IMPROVE ALGORITHM QUALITY & INNOVATION

Historic (Big) Data

Data supplied by other organizations

Own Organization Data

Own Organization

Digital Data Marketplace

Algorithm Developers

Computer science

Data science

Math and statistics

Competitive Domain knowledge

Data Exchange

Decision Support Systems

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

(Near) Real Time Operational Data

Periodic storage

Algorithm Choice

Own Organization
Open vs Closed

Considering our research context:
- difference with 2: Governance by a membership organization
- difference with 3: Data is stored outside data platforms to allow multiple platforms to access & use same data
Many organizations want to keep their historical data in their sovereign data zones.

Many implications need to be considered:

<table>
<thead>
<tr>
<th>Business level</th>
<th>Legal level</th>
<th>Data level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>Ownership</td>
<td>Processing</td>
</tr>
<tr>
<td>Cost</td>
<td>Access</td>
<td>Storage</td>
</tr>
<tr>
<td>Benefits</td>
<td>Usage</td>
<td>Management</td>
</tr>
<tr>
<td>Agreements</td>
<td>Compliancy</td>
<td>Transport</td>
</tr>
<tr>
<td>Exchange</td>
<td>Liability</td>
<td>Transform</td>
</tr>
<tr>
<td>Trade</td>
<td>Market Rules</td>
<td>Security</td>
</tr>
</tbody>
</table>
OVERCOMING 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.
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

DEFINE AND AGREE COMMON BENEFIT
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

National Law & Regulations
Market rules
Member admission

Digital Data Marketplace Membership Organization

Agreement
Registry

Infrastructure Patterns

Deployment Specification

Data Exchange Infrastructure

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

Data suppliers

Algorithm Developers

Dispute Resolution

Accounting & Auditing

Algorithm Developpers
DATA EXCHANGE CONCEPT
ENVISAGED GLOBAL EXCHANGE INFRASTRUCTURE

Global Data Exchange Infrastructure

Autonomous Data Science Platforms

Marketplace A

Marketplace B

Sovereign Data Owners (e.g., Airlines)

AMdEX
THE DATA HYPERMARKET

amsterdam economic board
PROCESSING & STORAGE: TRAINING STRATEGIES

CENTRALIZED VS FEDERATED ANALYTICS

**Centralized**
- Raw data transferred from dispersed data zones to a central repository for analysis

**Federated**
- Raw data stays in place. Model trained through orchestration of local (at each data zone) and global computations

- **Centralized Repository**
- **Virtual Computing Cluster**

- **Data-zone Fabric**
- **Data-zone Fabric**
DMP provides neutral processing capabilities, dissolving 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.