CHALLENGES ON DATA SHARING
AND HOW RESEARCH HELPS ADDRESSING THEM

Researching consortium driven Data Exchanges for AI/ML development

Leon Gommans, PhD
Science Officer
AIRFRANCE KLM GROUP
Professor Data Exchange Systems,
University of Amsterdam, SNE
What is Trust?

Context: two sided markets for AI development

What are the challenges to overcome?

- Platform archetypes in the light of trust
- Consortium creating challenges

How can trust be organized?

How to implement trust?

Examples of solving some challenges.
WHAT IS TRUST?
COMPLEX CONCEPT, MANY THEORIES

Dimensions (Bachman):

(Inter-) personal trust: Formed by the interaction between persons and growth with experience.

Rooted in the tacit understanding of personal trust he also recognizes impersonal trust, with sub-categories:

System trust: e.g. safe operation of a plane as a system with oversight from aviation authorities

Institutionalized trust: Organizations interacting based on rules, standards, code of conduct established by trade organizations, industry forums, standards bodies, or a dominant player.

A trust model (Mayer)


Trust is the willingness of a trustor to be vulnerable to the actions of a trustee based on the expectation that the trustee will perform a particular action important to the trustor.
CONTEXT: DATA SHARING FOR AI DEVELOPMENT

A MODEL STUDIED

Data suppliers → Trustworthy two-sided market → Data Consumers

Airline Operator A → enabing access and use → AI Developers

Airline Operator B

Own Airline → Periodic storage: raw or with enhanced quality

Historic (Big) Data

Computer science

Math and statistics

Competitive

Domain knowledge

Decision Support Systems

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

(Near) Real Time Operational Data

Algorithm User

Algorithm Choice

Periodic storage: raw or with enhanced quality

Data Consumers

AI Developers

Data suppliers

Airline Operator A

Airline Operator B

Own Airline

Historic (Big) Data

Competitive
# Platform Archetypes for Sharing Data

Driven by Different Ways to Organize Platform Governance

<table>
<thead>
<tr>
<th>Driver</th>
<th>Trust</th>
<th>Self interest</th>
<th>Common interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust</td>
<td></td>
<td>Trust a Single Party</td>
<td>Organize Trust</td>
</tr>
<tr>
<td>Driven by</td>
<td>Archetype</td>
<td>Goal</td>
<td>Role IT</td>
</tr>
<tr>
<td>Existing enterprise</td>
<td>Internal platform offered externally</td>
<td>Be the best in your environment</td>
<td>Efficiency: Digitizing as means</td>
</tr>
<tr>
<td>Investors in new enterprise</td>
<td>Centralized Platform</td>
<td>Create shareholder value (typically at the expense of the existing environment)</td>
<td>Support disruption with Agility: Move fast / break things</td>
</tr>
<tr>
<td>Alliance / Consortium (with specific aim)</td>
<td>Distributed Platform</td>
<td>Extend reach for suppliers and/or offering to a client using a common environment</td>
<td>Support standards to integrate with a common alliance / consortium governed infrastructure</td>
</tr>
<tr>
<td>Federation (with holistic aim)</td>
<td>Federated Business Ecosystem</td>
<td>Arbitrary collaboration with environment often for the common benefit of the environment</td>
<td>Support creation of new business models by integrating own services with standardized, neutral federation services &amp; Infra</td>
</tr>
</tbody>
</table>
Through the design and implementation choices, encapsulated in a business model, a company chooses its stakeholders and its importance (i.e. its bargaining power) in the ecosystem.

**Enabling sovereignty of choices is a key: Requires neutrality of exchanges.**

Competition is formed and defined between business ecosystems.

Business ecosystems are believed to be capable of better explaining **how multi-sided businesses evolve** (demand/supply side). Data can have a demand and supply side.
## DIGITAL DATA MARKETPLACE APPROACH

ORGANIZING TRUST USING TRUSTWORTHY INFRASTRUCTURE FOR SPECIFIC GOAL

<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 <strong>means to reduce risk</strong> for participating members</td>
<td>Research operationalization of Digital Data Marketplace concepts</td>
</tr>
</tbody>
</table>

**Consortium activities**

- **Benevolence**
- **Integrity**
- **Competence**
- **Risk management**
Many organizations want to keep their historical data in their sovereign data zones.

Many implications need to be considered, to name a few:

**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
Imagine if data scientist can use historic data from specific aircraft types operated by multiple airlines.

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.

The more Flight Deck Effect occurrences are available, the more likely that a prognostic relation can be learnt.
JOURNEY OF THE DATA SCIENTIST / ENGINEER

ROLE OF THE DIGITAL DATA MARKETPLACE (DDM):

DDM ENABLING DATA SCIENTIST TO EXPLORE AND USE MORE- AND MORE DIVERSE DATA

ROLE DDM:
1. Explore sample data
2. Industrialize with real data
3. Offer AI products

Offering seamless integration with AI sandbox environment

What is the end purpose? How will it add value? Use case, data scope and stakeholders identification

Collect and explore data, research data science model Prototype, go/nogo to production

Deliver the solution in production environment Product GO live

Feedback analysis, A/B testing, and performance monitoring New product releases, business insights

TEAMWORK

SELL / USE AI PRODUCT

JOURNEY START

IDEA

EXPRESS

flexibility

Feasible Model

BUSINESS CASE
go/nogo

Go live

PRODUCT

11. PRODUCT

10. SETUP MONITORING

9. FEEDBACK LOOP

8. CONTINUOUS INTEGRATION

7. AUTO TRAIN

6. DATA FLOW

5. VALIDATE

4. MODEL

3. EXPLORE

2. COLLECT

1. DEFINE

DEFINE (1-3)

EXPERIMENT (4-5)

INDUSTRIALIZE (6-10)

IMPROVE (11-12)

SHARE

What is the end purpose? How will it add value? Use case, data scope and stakeholders identification
ESSENTIAL INFRASTRUCTURE ARCHETYPES
MANY VARIANTS: FOCUS ON CONSORTIUM DRIVEN APPROACH TO ORGANIZE TRUST

Centralized
Bring data to the algorithm

Distributed
Bring algorithm to the data

Federated
Using trusted infrastructure

Data owners
Developer

Consortium
Data owners

consolidate
DIGITAL DATA MARKETPLACE ARCHITECTURE
RESEARCHING IMPLEMENTATION OF ESSENTIAL ELEMENTS WITH IT INDUSTRY

Digital Data Marketplace Membership Organization

Market rules

Member admission

National Law & Regulations

Agreement

Registry

Data Exchange Infrastructure

Infrastructure Archetype

Centralized Distributed Federated

Data Science Transaction

Research Testbed

Global Digital Data Market Infrastructure

Data supplier(s)

Algorithm Developer

Accounting & Auditing
SAE ITC EXCHANGEWELL SUPPORTS DDM CONCEPT

https://www.sae-itc.com/program/exchangewell
FEDERATED ARCHETYPE IMPLEMENTATION
CONSORTIUM BUILDING USING DATA SHARING COALITION CANVAS & APPROACH

Federated Trust
Using trusted infrastructure

Data owners

Cross-Domain Trust Framework

Consolidate

DDM orchestration & authorization

Data service transaction agreement

Data service transaction

Data service provider

Data service consumer

Federated Trust
Using trusted infrastructure

Developer

DATA SHARING COALITION

Business
- Roles & responsibilities
- Context and goals
- Fee structure

Legal
- Relevant rules & regulation
- Contracts
- Governance

Operational
- Service levels
- Operational governance
- Incident management
- Tooling

Functional
- Functional scope
- Interaction model
- Privacy

Technical
- Technical specification
- International standards
- Security standards
Demonstrated Digital Data Marketplace as Prototype

AMS-IX Project AMDEX as Facilitator to Enable DDM’s

AMdEX members supporting the establishment and operation of DDM’s defined by consortia customer journeys
Supporting need for consortium neutrality with neutral and trustworthy (software definable) infrastructure.

Governance

Setup

Member Admission

Asset Registration

Trade Agreements

Transactions

Clearing & Settlement

Auditing & Security

Universiteit van Amsterdam
USE-CASE: DATA SHARING FOR AI DEVELOPMENT
USING A DIGITAL DATA MARKETPLACE GOVERNED BY A MEMBERSHIP CONSORTIUM

Demand: AI Developers

- Historic (Big) Data
- (Near) Real Time Operational Data

Data supplied by other organizations

Own Organization Data

Competitive

Domain knowledge

Computer science

Math and statistics

enabling access and use

Periodic storage: raw or with enhanced quality

Algorithm User

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

Two sided DDM

Algorithm Choice

Decision Support Systems

Periodic storage: raw or with enhanced quality
Research creating a TRL 5 prototype performed within DL4LD project lead to product
DDM ORCHESTRATION & AUTHORIZATION PROTOTYPE

AI flow Data/Algorithm

Choose Algorithm

Choose Dataset

Execute Training, Model creation, Prediction

Log tracking, execution tracking
Start with PoC/ MVP building involving a few visionary members seeing the need to develop standards. Value is expected to attract new members and add new business as the consortium grows.

Start with consortium building, defining working groups to work on creating value by designing solutions. Selecting new business & solutions as the consortium growth.

Value delivered to members

Role of AMdEX

Membership fee required to organize meetings

Not achievable

# of consortium members
Marktplaatsen

AMDEX Fieldlab

AMDEX Fieldlab

Open exchange infrastructuur

AMDEX Infrastructuur

Exchange operatie

Governance (Vereniging AMS-IX)

Uitvoering & Beheer (AMS-IX B.V.)

Use cases SAE ITC GAIA-X

UvA DDM research with Industry
Equinix, Dell, Nokia, Ciena, KPN, ...

Use cases SAE ITC GAIA-X

AMdEX Industry Lab

Technology provider Y

Marktplaats services

DEXES

SURF

Usecase Madaster

Usecase RDX

Open service architecture
Consortia development
Identify ‘why’ (use-case benefits), rulemaking, organize trust, policy scheme’s

Normative reasoning
Implement trust via automated compliance monitoring, request assessment, enforcement using legal concepts

‘Data Space’ Infrastructure Development
Data sharing architectures & archetypes for streaming, transactional, historic data, AI development (DDM)

Data Exchange Systems
Applied Industrial Research Lab using Future Internet concepts to create software definable global data exchanges