Data platforms
A requirement for the modern AI methods
Data is the new oil

You are already paying with (personal) data
Producing data does not mean you can capture or store it
Sitting on tons of data would not give you any value

Data needs to be extracted, transformed, loaded, analysed
The one generating data rarely has those skills
There are two alternatives

Create your own Data Science team:
- Higher investment, risk
- Added complexity
- It is not necessarily going to be a better team

Contract a specialized 3rd party:
- Need to handle contracts
- Limited control over results
- Data leaves the company, it requires trust

or a mix, which has the best and the worst of both possibilities

Anyway, this solves the problem, or doesn't it?
Modern AI methods demand Big Data

https://xkcd.com/1838/
Value and veracity are also considered as part of Big Data but not as widely accepted as volume, variety and velocity.
The B in Big Data

**Volume, Variety and Velocity**

Modern AI models devour data they require to be fed by huge amounts of it.

In general high Volume increases **precision**.
For modern AI-models it is not enough to have a huge amount of data, it also needs to be representative of the phenomenon under study.

In general variety increases accuracy.
The B in Big Data

Volume, Variety and **Velocity**

Most socio technical systems are not **stationary**, therefore AI-models need to be continuously re-evaluated with new data.
The B in Big Data

Volume, Variety and Velocity

Without the three components it is very unlikely to develop and deploy modern AI applications

volume - precision, variety - accuracy
Variety is impossible in isolation

Volume and velocity can be reached in isolation by a data producer, but variety...
Variety is key

Variety is key for learning, imagine learning what a cat look like with this information:

it is simply impossible, no matter how many pictures or how recent they are
Runway exit prediction

Feature importance when the prediction is performed at 2NM from threshold and Normalised

Confusion matrix of the classification.
Conclusion

Variety is as important as volume if not more.

AI applications should not be developed in isolation.

Data platforms are a necessity for AI applications.

Are data platforms enough?
DataBeacon
Beyond the data platform, the open AI platform
What is DataBeacon?

DataBeacon is not just a data platform but an open AI platform for aviation.
DataBeacon leverages your data to create big data in this sense it is a data platform
By using a proprietary cryptographic solution Secure Data Fusion (SDF) DataBeacon merges your data with others securely and privately, but still useful for building AI applications.

A global Data Protection Governance model extends the technological solution by a series of Data Protection Agreements (DPAs) and Annexes.

Your data is protected twice!
But, just sitting on (big) data won't give much insight
DataBeacon is not just a data platform

(Teams of) Data Scientists can join DataBeacon as well and develop AI applications within the platform.
DataBeacon an open-AI platform

DataBeacon offers a **full-stack solution** for Data Scientists, not just data

Data clean and ready to be consumed in a **Secure Data Frame** format

Under a global **data protection governance** model

Securely private **sandbox environments** to analyse the data

**Data is never distributed, stays always in the cloud**
DataBeacon offers a full-stack solution for Data Scientists, not just data

On demand cloud-based scalable clusters to cope with demanding AI models

Full development environments with customizable data science toolsets

Collaborative work based on Git and agile scrum

Communication of results through interactive dashboards

Accessed anywhere just from a (secure) web-browser
The DataBeacon open AI platform for aviation
The architecture

How is DataBeacon built?
DataBeacon sits in the middle between stakeholders and analysts

Privacy by design

Data engineering - Security, Scalability, Availability

Integrated technology platform & governance model

Designed for AI applications

Secure Data Fusion sandboxed environments
AI is a collaborative effort

Who **builds** a data platform?

**Infrastructure engineers** - design & deploy a cloud infrastructure

**Operational stakeholders** - bring data, design/decide use cases

**Domain experts** - collaborate with data scientists

**Technology partners** - data management, infrastructure
AI is a collaborative effort

Who operates a data platform?

Data Scientists - research, new analytic concepts

Data quality engineers - explore, prepare, validate data

Data analysts - prepare reports, dashboards

Software engineers - develop, optimize and deploy
Integrated technology platform & governance model
DataBeacon today

Tested as the (closed) AI platform for the Safeclouds.eu project

- 10 data providers/operational users
- 5 technology partners
- Infrastructure ready to be used
- Governance model streamlined
- Collaborative AI environment ready

*as of October 2018*
DataBeacon tomorrow

Airprox
Separation with terrain
SafeRunway
Level bust
Hard landing
Wake vortex separation
Congestion monitoring
Unstable approach

Collaborative AI-app market for aviation
Thank you!

Samuel Cristobal
sc@innaxis.org, @tadorea.com
www.tadorea.com/databeacon
What is DataBeacon?

DataBeacon is the open-AI platform for aviation.

Why do I need to join?

Because AI isolated efforts have a limited chance of success.

If you are a **stakeholder**, you need to join DataBeacon to break the isolation:
- make your data really big, by adding variety to it
- reach out a community of **analysts** and **AI-developers**

If you are **analysts** and **AI-developers**, you are smart I will let you figure it out.
Conclusions

Data Platforms - not longer an option but a necessity to carry ML/AI work
Focus on high impact case studies to keep up the momentum
Start with existing data and involve more stakeholders to expand
Deploy analytics to solve real problems engaging with operations
Embrace agile methodology through a product owner
Democratize data with the help of strong data protection
Data Platforms are a gargantuan tasks - Leverage on DataBeacon
Our focus

Secure and streamlined data analytics process
A data protection agreement that brings confidentiality and privacy
De-sensitised / de-identification secure fusion of data
Verifiable restricted access to known partners
Clear, agreed and auditable use of the data
Cost and time efficient AI/ML projects - solving the "cold start" problem
Collaborative, "open", and secure environment