Technical challenges in the effective safety analysis

Global Aviation Data Management (GADM), IATA
Hyuntae Jung

To represent, lead and serve the airline industry

Introduction to GADM

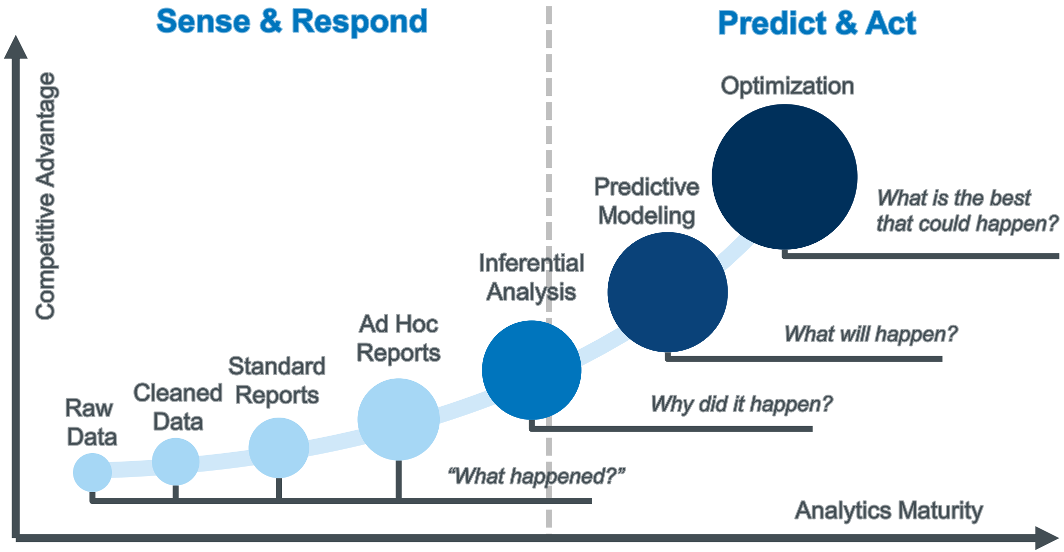
- Global Aviation Data Management (GADM)

GADM is a data management platform integrating several sources of operational data received from different IATA programs.
To provide the industry with comprehensive, cross-database analysis and to support a proactive data-driven approach for advanced trend analysis and predictive risk mitigation.

- More than 470 organizations
- 15 + years of successful delivery
- Multiple databases for comprehensive analysis:
  - Accident / Incident reports
  - Ground damage reports
  - Flight data

Contents

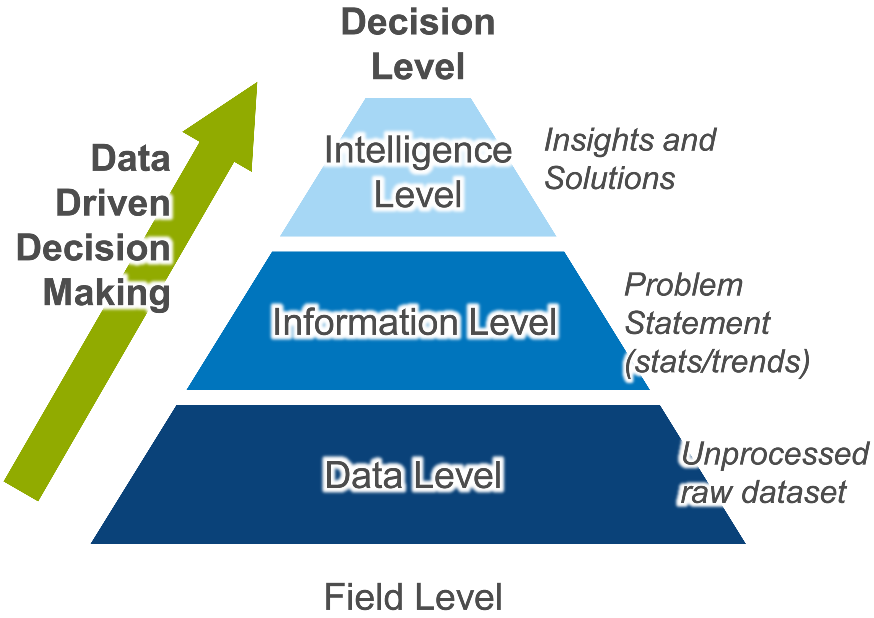
Evolution of Safety Analysis



- **Descriptive Analysis**
  "What is happening?"
- **Inferential Analysis**
  "What is the possible explanation for what is happening now?"
- **Predictive Analysis**
  "How can future decisions be made (and how can finite resources be allocated) based upon what is happening now?

Source: ICAO Doc.9859 Safety Management Manual, 4th edition (2018)

Data-Driven Decision Making



- **Data-driven decision making:** Decisions with the data and quantifiable evidences, rather than intuition or based on observation or experience alone.
- including the most effective and efficient allocation of resources.

Using safety data / information to identify and to mitigate safety risks

Data Driven Decision Making (D3M)

Source: ICAO Doc.9859 Safety Management Manual, 4th edition (2018)

Data Collection → Analysis → Identification of high risk/priority → Risk management / Safety Management

Safety Intelligence

- Leveraging safety data and information to develop **actionable insights**.

  More than just stacking lots of data!

- A huge amount of safety data, which, when correctly analyzed, can be transformed in what we call **“Safety Intelligence”**, for example, safety intelligence may be able to identify **precursors** to better address and monitor the safety risks.

IATA | GADM GLOBAL AVIATION DATA MANAGEMENT

- **Finding data in the aviation system is easy, but effectively building knowledge with it is difficult.**

Good Analysis: “Right Data at the Right Time.”

- Large, complex and various kinds of data
  - **Quantitative Data:** Flight parameters (FDA/FOQA), Radar or ADS-B tracks – ***sensor captured***
  - **Qualitative Data:** Accident, incident and risk reports, Audits – ***written in human language***
- Human-Machine Interoperability
  - The data is for the machine, but the analysis is for human.
  - Analysis delivery, as well as understanding the context of the analysis, is important to avoid misinterpretation.
- Limited Resources
  - Clear objectives, scopes, priorities and effective resource allocation, with appropriate competencies, are a must.

Challenges in Safety Analysis

- All kinds of reasons how and why safety analysis can be misconducted:

Data Collection → **Analysis** → **Identification of high risk/priority** → **Risk / Safety Management**

Incomplete Dataset
Poor data collection system, Poor protection, Reporting culture.
- *Insufficient data*
- *Biased data*

Poor Data Quality
Uncertain data source, Lack of quality management.
- *Non-standard taxonomy*
- *Not-fit-for-purpose data*

Poor Scope and Objectives
Mis-identified problems, unknown context, unreachable expectation,
- *Analysis insignificant or not supporting critical objectives.*

Lack of Resources
Lack of manpower, time, skills, tools or funds.
- *Low quality analysis – "No insight delivered."*

Poor Delivery
Miscommunication, Lack of or wrong visualization.
- *Complicated and difficult analysis result.*

Misinterpretation
Level of understanding, Bias, Uncontrollable variables.
- *Right number, wrong conclusion*
- *Manipulation / misuse*

- **Wrong Decision**
- **Distrust of Safety Analysis**
- **Ineffective D3M**

IATA

Data Availability: “Can we get the full picture?”

- Data is like nutrition for safety analysis: it needs to be sufficient and balanced.

Accident, Mandatory and Voluntary Reports

Undesired Aircraft State, Operational deviation, Graded condition

Latent precursors, Threats and errors

Blind spots, Unobservables

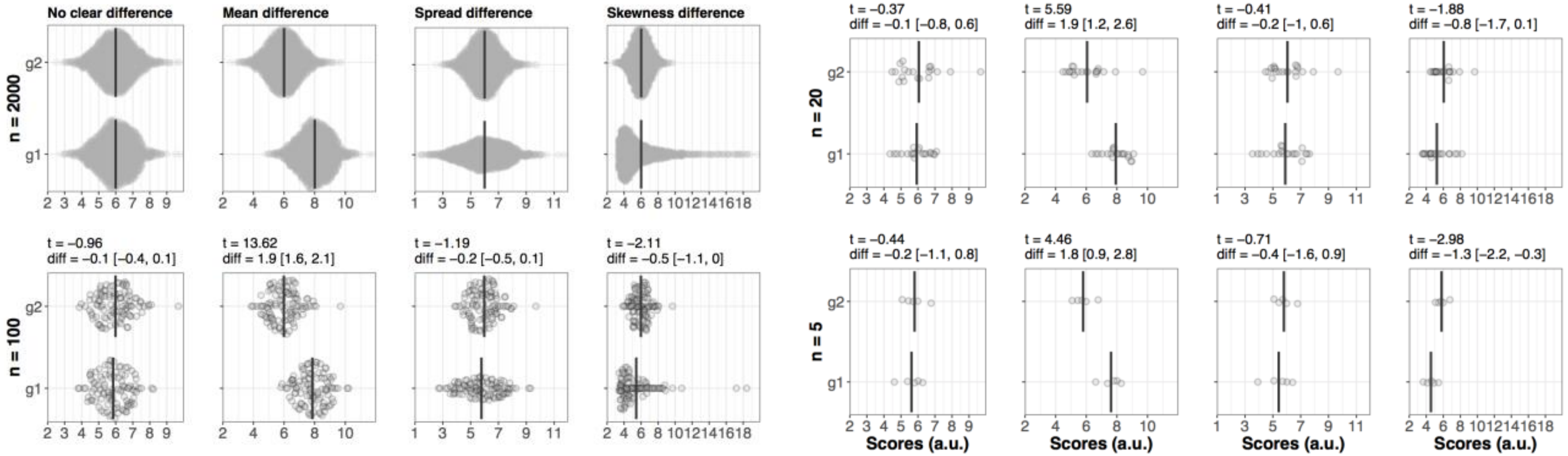
Conditions that we don’t even know.

	Data Availability	Data Volume	Processing Cost (Time)	Collected Format
Visible	High (accessible)	Limited	Cheap	Structured
Obscured	Low (difficult)	Large and Complex	Expensive (exponential)	Unstructured

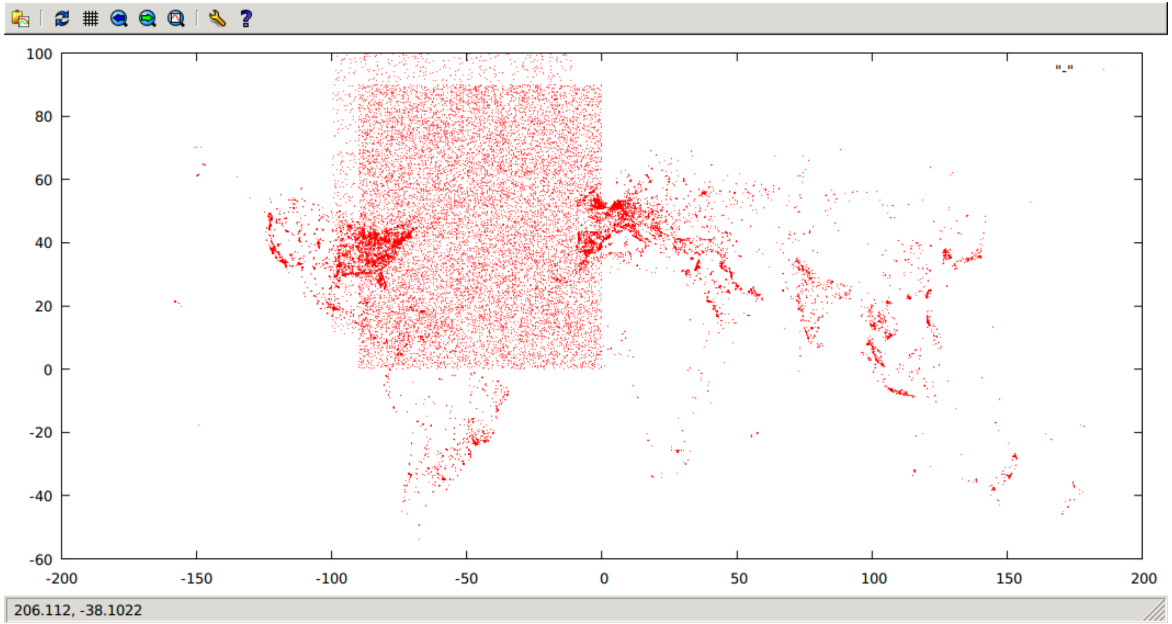
- **We will not have data for every known parameter for normal operations.**
  - ***Insufficient data*** *– lower confidence level*
  - ***Biased data*** *– biased conclusion*

Sample Size and Uncertainty

- Analysis with insufficient sample size has lower confidence level. (Weak Conclusion)



Data Integrity for Data Quality

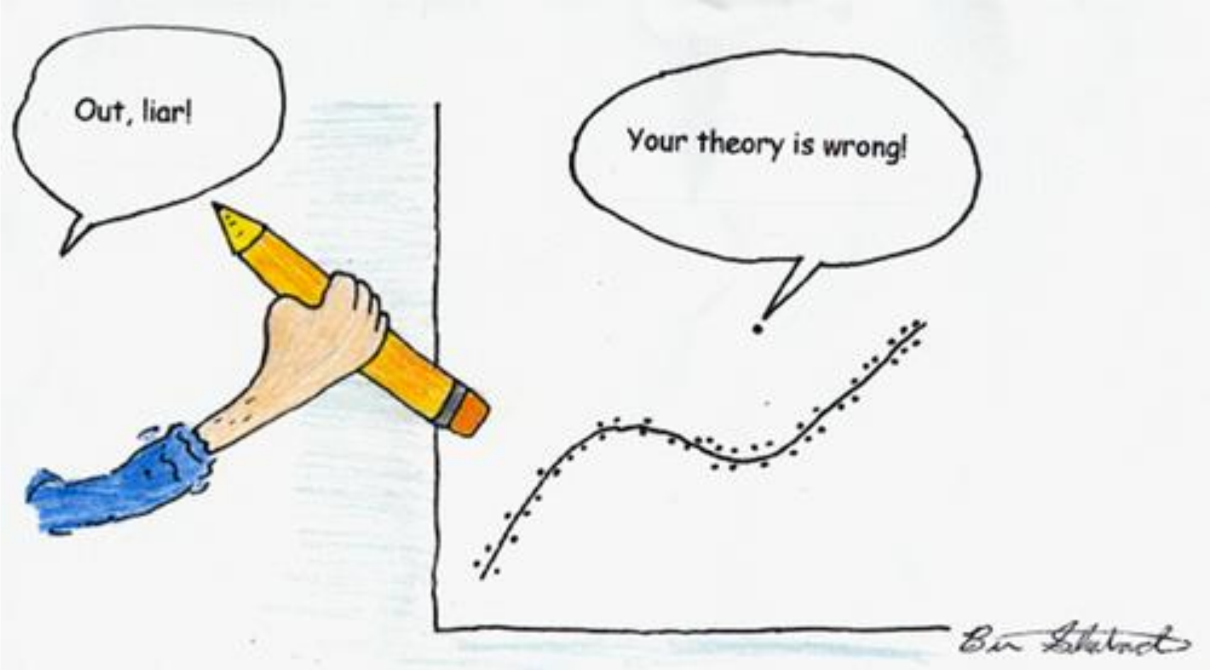


Data Integrity for Data Quality

- **Data Integrity**: the reliability of the information in terms of its physical and logical validity.
  - *Completeness*
  - *Uniqueness*
  - *Validity*
  - *Accuracy*
  - *Consistency*

- **Data Quality**: the reliability of information to serve an intended purpose including planning, decision making and operations.

"Nearly one third of analysts spend more than 40 percent of their time vetting and validating their analytics data before it can be used for strategic decision-making"



- Multiple poor data points can ruin the whole findings!

Data Heterogeneity

- Heterogeneous data are any data with high variability of data types and formats.
  - **Syntactic**: two data sources are not expressed in the same language.
  - **Semantic**: the differences in modelling the same domain of interest. *(logical mismatch)*
  - **Terminological**: variations in names when referring to the same entities
  - **Semiotic**: different interpretation of entities by people.
- Heterogeneous data in safety analysis:
  - *Merging global and multi-source data*
  - *Integrating quantitative data (flight data) with qualitative data (written report)*
  - *Especially data written in natural language form (basically all of the safety reports).*
- **Only limited parts of the dataset would be remained as valuable.**

Challenge – Sample size and heterogeneous data

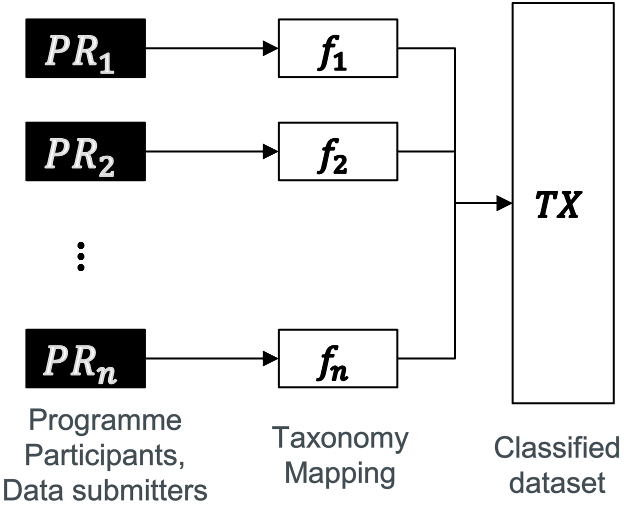
- For aviation safety analysis,
  high data integrity is required but there is only limited data available.

- ICAO world accident rate: 2.42 per 1 Million departures **(0.000242%)**
  Scheduled Commercial flights on airplanes above 5.7 tons only, Source: iSTARS Accident Statistics
- Null Hypothesis: *"there was no accident in 2017"*, ratio of "*correct*" prediction: 99.999758%
  Is 99.999 of accuracy is what we can satisfy?

- Things are getting complicated:
  - Is the data point an outlier to carry on, or a rarely observed but critical value?
  - How can we determine that the data point was occurred just by chance or not?
  - How can we clean poor data, if we are not sure about its reliability?
  - How do we maintain the data quality and confidence level if the sample size is small or the sample is extremely skewed?

Approach: Taxonomy

- Defining a structure of classifications of raw data with human supervision

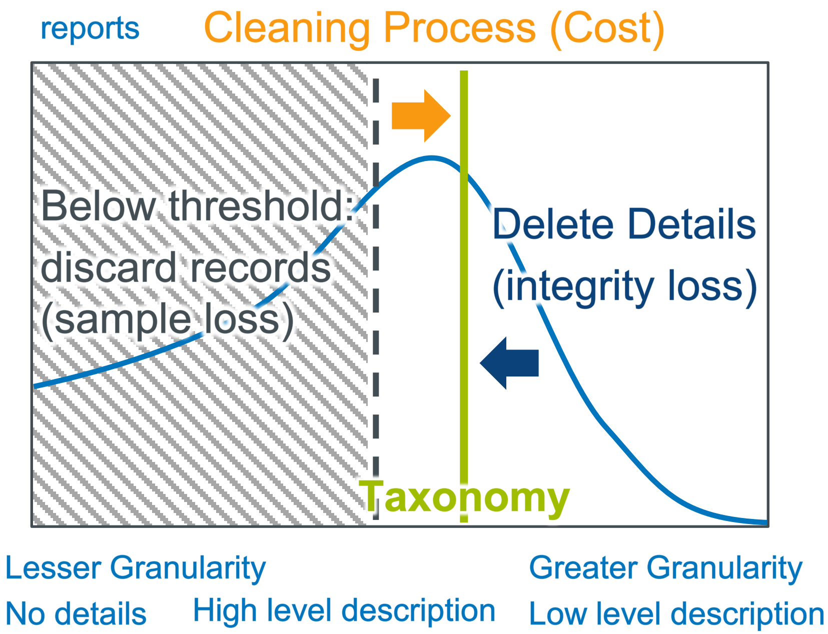


- Can be referred as “data standardization” to achieve, track and manage data quality.
  consistent, complete and unambiguous

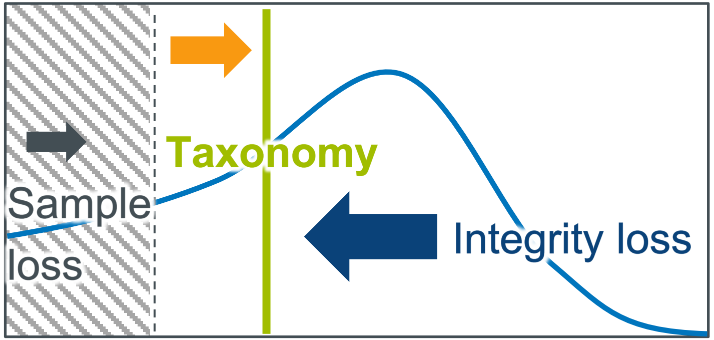
- Centralized rules and procedures from expert’s knowledge.

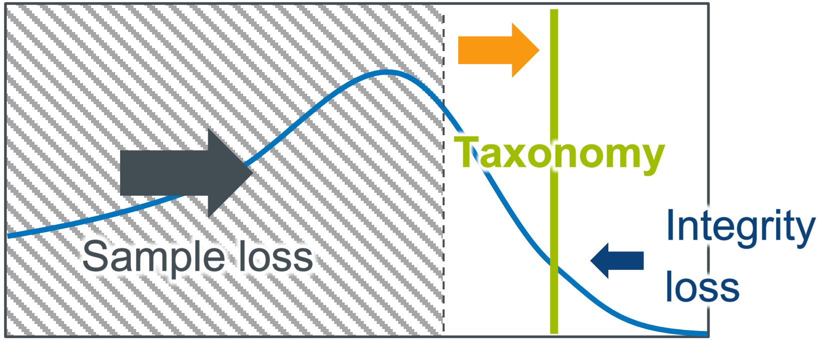
- However, human intervention makes data vulnerable to human errors, thus additional quality check and validation process are required.

Taxonomy and Data Standardization



- **Granularity**
  How detail the valid information each record contains.
- **Cleaning Cost**
  Brining records with less granularities can be fitted into taxonomy after investing resources.
- **Quality Threshold**
  The minimum threshold of granularity for records required to be fitted into taxonomy.
- **Sample Loss**
  Records below the threshold cannot be fitted into taxonomy, and thus, such records are deleted for consistency (data quality)
- **Integrity Loss**
  Taxonomy cannot hold greater granularity of some records, and thus, such records lost some granularity by being fitted into taxonomy





- **Taxonomy in too much high level**
  - Less data sample loss
    More data integrity loss
  - Safety analysis might require greater granularity than taxonomy. (Additional data processing needed).

- **Taxonomy in too much low level**
  - More data sample loss
    Less data integrity loss
  - Higher processing cost is be needed to keep more sample, but sample loss is inevitable. (Analysis will has lower certainty)

- **Finding the right balance is important: data sample, integrity and resources.**
- **The taxonomy shall keep up with changing industry.**

IATA Safety Incident Taxonomy (ISIT)

↗ Better integrate the data and have a common language among aviation safety communities

Program Participant

- Participating airlines incident data submission

Data Submission

ISIT

- ATM & Flight Operations
- Cabin
- Engineering/Maintenance
- Ground
- Security
- Common
- Occupational Health and Safety
- Why taxonomy

(developed with cooperation with over 100 industry professionals)

Data Processing

Incident Database

- **Incident Data Exchange (IDX)**
  STEADES & GDDB
- Safety Analysis
- Query of de-identified database
- Identification of safety issues

Track data qualities

Submission / Taxonomy Improvement

Safety Community

- Operators & Service Providers (Airlines, ANSP, Airports)
- Manufacturers
- Safety Agencies
- SMS Software Provider

Analysis Methodology

- Process of building insights from processed dataset

Effective Analysis using right methodologies

Data Level	Information Level	Intelligence Level
"What happened"	*"What is going on"*	*"Why and How did it happened"*
• Processed data (taxonomy classified data) for analyst to begin safety analysis.	• Using statistics and indicators to show what is going on. (trends)	• Analyst diagnose data in-depth to identify contributing elements below the surface. *(e.g. threat and errors)*

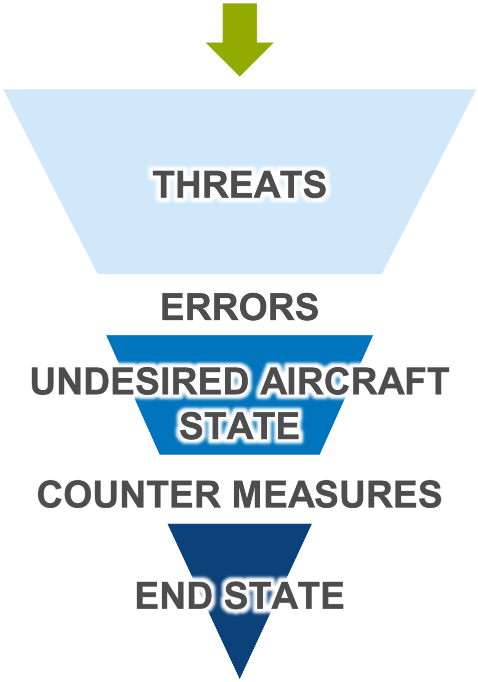
Analysis Methodology

- **Threat Error Model (TEM)**

  Threat and Error Model enhances the classification system used by IATA to determine contributing factors in incidents and accidents.

- **Subject Matter Expert (SME) Interview**

  SME can provide information or hints to analyst which might be invisible in dataset even that they might not even be conscious of knowing them.



Analysis Methodology

Safety Performance Indicator (SPI)

Period of Review [2018-Q2]	IATA	Non IATA
Rate (Accident per 1 Million Sectors)	0.66	1.96
Number of flights per accident	1,505,636	510,953
Number of Accidents	4	10
Fatalities	0	113
% of sectors flown	54%	46%
% of passengers (Total ASM)	82%	18%

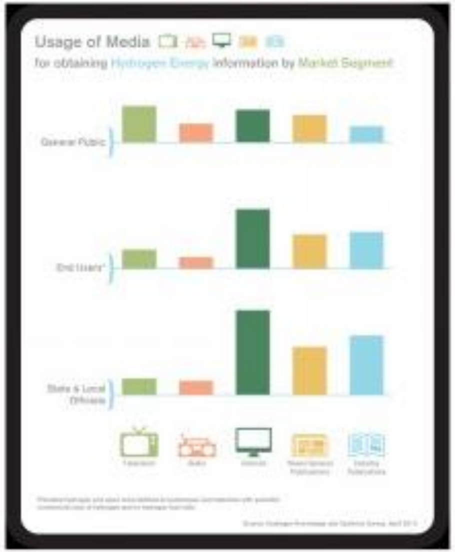
Risk Matrix

Safety Risk		*Severity*				
Probability		*Catastrophic A*	*Hazardous B*	*Major C*	*Minor D*	*Negligible E*
Frequent	5	5A	5B	5C	5D	5E
Occasional	4	4A	4B	4C	4D	4E
Remote	3	3A	3B	3C	3D	3E
Improbable	2	2A	2B	2C	2D	2E
Extremely improbable	1	1A	1B	1C	1D	1E

Note.— In determining the safety risk tolerability, the quality and reliability of the data used for the hazard identification and safety risk probability should be taken into consideration.

- **Provides strengths that offset the weaknesses of both quantitative and qualitative research, and thus, more complete and comprehensive context.**

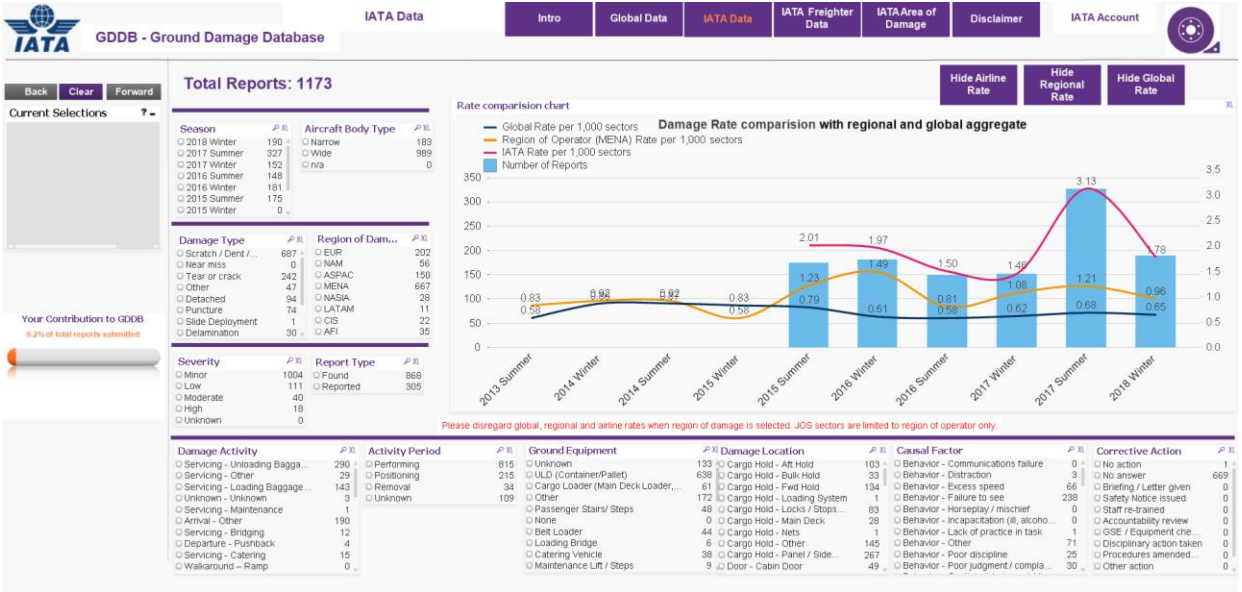
Visualization for Analysis Delivery



- Clear Understanding
- Enables Story Telling
- Stronger Delivery

Dashboard: Self-Service Analysis

- User Customized – Interactive Benchmark and Query Tool



- Users can perform taxonomy-based analysis easily, with their objectives and intentions.
- Basic analysis by end-users: better information accessibility with less analysis cost.
- *How do we prevent misinterpretation?*
  User Experience Design

Effective Use of Safety Intelligence

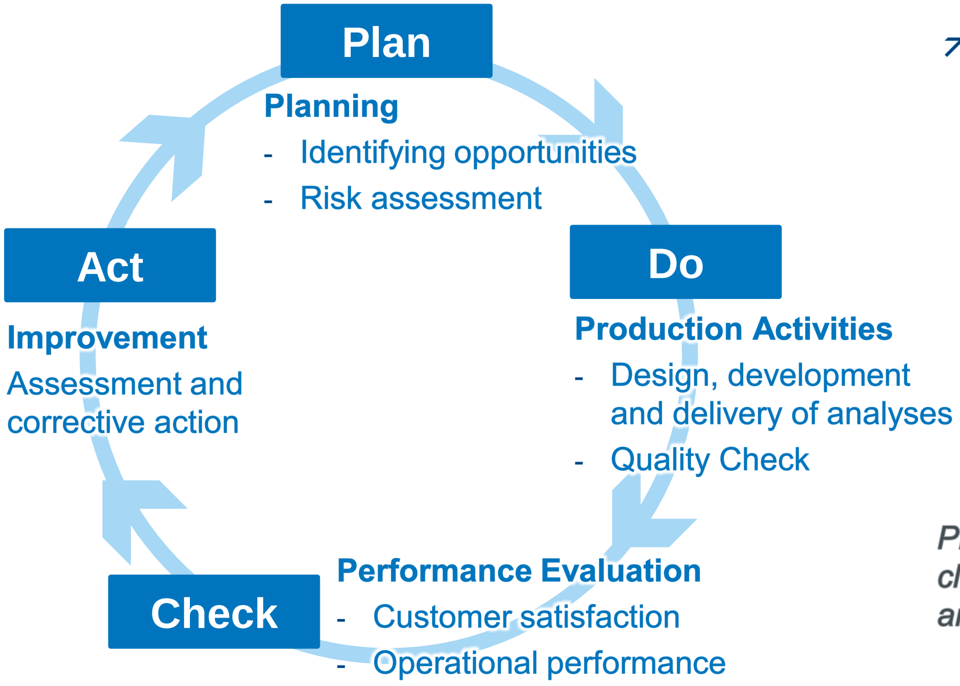
- **Confidence by decision makers** in the accuracy of a safety analysis is the key element for its effective use.

  If decision makers do not believe in the analysis, for any reason, they will not use it.

- It is critical to assure that the analysis result is clear to decision makers, and that there is no chance for **misinterpretation**.

- These risks can be mitigated by involving end-users (analysis customer and/or decision makers) from the beginning to the end of the analysis cycle.
  - Inception Interview: to better identify what end-users want.
  - Validation: to ensure that the analysis progress is aligned with the original objective and scope of users.
  - Customer Feedback: to improve the analysis cycle and thus, better reliability of the deliverables.

Quality Management System



ISO 9001:2015 implemented with PDCA Cycle

- **Plan**: Design or revise business process components to improve results
- **Do**: Implement the plan and measure its performance
- **Check**: Assess the measurements and report the results to decision makers
- **Act**: Decide on changes needed to improve the process

Proactive management to achieve high standards of quality, clarity and reliability of deliverables, so that it can understand and fulfill the needs of our customers.

Process Capability & Maturity

Optimising
Continuous process improvement (incremental and innovation)
Common causes of variation are identified and improved
Processes are agile and "best in class"

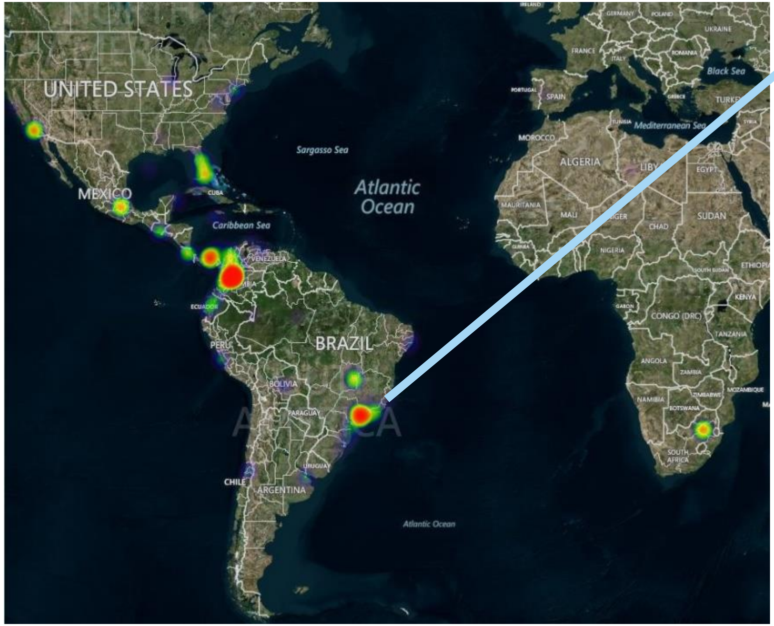
Quantitatively Managed
Targets, standards and measures are used
Special causes of variation are identified and corrected
Performance is predictable

Defined
Organisation-wide focus
Value chains are identified
All processes are documented
Measurements are defined

Managed
Departmental & Team focus
Some organised processes
Performance is repeatable

Initial
No organised processes
Ad hoc and reliant on "heroics"
Performance is not repeatable

Success Analysis Story



Sao Paulo TMA

- IATA + airlines presented to Brazilian authority (RAISING AWARENESS)
- Analysis of TCAS points and SIDs/STARs
- SIDs/STARs modified
- Key point: issue known previously by individual airlines, but aggregate data drove change

Another Problem: Overwhelming Data

- Big Data Era
- Data volume beyond human capabilities (especially where human intervention is required: taxonomy classification and validation process)
- Proposed solution: data-driven automation tools

Next Step: Automation and A.I.

- Automation – replace repetitive works
- Intelligence Augmentation (IA) – support decision-making without calculating everything!

- A.I. experiment in GADM

https://projector.tensorflow.org/?config=https://raw.githubusercontent.com/HTjung/embeddings/master/config.json

Hyuntae Jung, Assistant Manager
jungh@iata.org

To represent, lead and serve the airline industry