**Use of data science to make the difference**

**in the investigation analysis process**

**By Marion CHOUDET (ATR) and Sébastien DAVID (BEA)**

*Marion CHOUDET, ATR Flight Safety Director, is accident investigator in the ATR Flight Safety team, hence heavily involved in safety strategy and supporting the development of the SMS within ATR. She has 13 years’ experience in accident investigation and safety. She has also gained experience in ATR design office on flight control system and continued airworthiness.*

*Sébastien DAVID, BEA Safety Investigator, is a BEA senior safety investigator and coordinator of investigator training and investigation methods. He has participated in many major investigations as IIC or accredited representative, as well as head of working groups, for example for the Human Factors group during the investigation of the Air France A330, flight AF447 accident. He has been type-rated on the Dassault Falcon 7X and has a Masters degree in Human Factors.*

In 2017, digitalisation of our industry continues to grow. Any aviation activity such as operations, maintenance, production, aerodrome, ATM, is a source of data which presents new opportunities for safety.

Three elements are required to talk about data science. The first two are the obvious ones, data and an adapted tool able to process a huge quantity of data from different sources. A competent analyst is the third element required to turn data processing into data science. In aviation, data science has started with Flight Data Monitoring (FDM) to feed airlines safety management systems, and is moving towards “Big Data” projects[[1]](#endnote-1) to support State Safety Programs (SSP) with one main objective to improve safety in a proactive way. Data science is also a powerful tool to feed any systemic approach used to analyse an occurrence (accident, serious incident or incident) and to prevent similar ones.

By giving the opportunity to analyse one occurrence within a global context, data analysis programs further enhance the credibility and the accuracy of both lessons learnt and safety recommendations. The proper integration of data science in the context of a safety investigation requires the use of a structured analysis methodology not only based on explanatory factors as it is commonly performed but also aiming at analysing the accident in a global context, looking at similar occurrences and actual practices.

The objective of this paper is to look into the opportunities associated with the use of data science in safety investigations. The first part considers the BEA analysis methodology and the way to address lessons learnt and safety recommendations. In the second part, examples of flight data analysis contributions to safety investigations illustrate the concept, the evolution of data analysis integration and show the added value for future investigations.

**Analysis methodology, lessons learnt and safety recommendations**

After having identified safety issues and contributing factors posing risks to the conduct of air operations, the challenge for safety investigation authorities (SIA) is to communicate persuasive lessons learnt and safety recommendations to the aviation community and particularly to organizations able to implement actions. As stated in Part IV of ICAO document 9756, those safety communication means must present compelling arguments for safety action to mitigate the risks identified during the analysis process which is at the center of the safety investigation. This process is iterative and has increased in complexity when looking at systemic and organisational issues. It must therefore rely on a structured methodology which must enable to organize, to plot and to share the analysis orientations of the investigation team. As an accident or an incident can be considered as a failure of actual risk control measures (not only regulatory and documented ones), the analysis methodology must also study actual practices and similar occurrences.

The analysis methodology adopted by the BEA is adapted from MINOS©[[2]](#endnote-2). With this approach, safety is considered as the ability of the system to manage perturbations rather than its conformity to a predetermined behaviour considered as safe. An accident or an incident is therefore considered as an escape from a controlled environment. The BEA analysis methodology consists in four iterative steps which can lead to collect additional factual information, hence the need to adapt data science.

The purpose of the first step is to define the analysis framework by determining the concerned operators (front line –one or several- or peripheral), the operational situation they had to manage before and during the accident (or incident) and their preoccupations in such a situation. This step, which seems easy, is very important for the entire process of the analysis by limiting retrospection bias. Furthermore, when searching and working on similar occurrences, this step also allows to select occurrences with events and factors as close as possible as the ones that are investigated.

In the second step of the methodology, the idea is to determine what usually makes the system safe, by listing the safety expectations and provisions relative to the operational situation defined in the first step. Those safety expectations and provisions can be implicit or explicit and correspond to all assumptions dealing with the way safety is normally ensured. The purpose of the following step (third one) is not only to describe the way those safety expectations and provisions performed during the occurrence but also to identify what make them fail (or make them succeed for a positive feedback in safety investigation).

The analysis of the failures (or success as seen earlier) of the safety provisions and expectations is performed in the fourth step by three possible means: a “classic” explanatory analysis, an analysis using similar occurrences and an analysis of everyday operations, actual practices. Those sources of information enable to evaluate if the observed failures are specific to the operators involved in the occurrence or if they can be generalized to all the operators of the organization or to operators from several organizations. This evaluation therefore helps to determine if the measures that are supposed to manage actual risks are robust and effective, and therefore, to identify safety issues.

The analysis of the everyday operations also enables to know the way they are performed and the way errors can occur. During this fourth step of the analysis, if similar failures are not observed, the variability may be an exception to usual practices. If not, the reliability of the safety expectations and provisions can be considered as insufficient with risks that will be addressed in lessons learnt and safety recommendations. Data science significantly increases the accuracy of this analysis. The use of these analyses is crucial to make those safety communication means convincing and persuasive to the aviation community.

**Evolution in the use of recorded data in investigations**

The analysis of additional recorded data in the course of a safety investigation is not new.

The investigation of a CFIT accident that occurred in 2013 to an ATR evidenced that the crew selected an altitude which was below the MDA during the approach. An analysis evaluated the conduct of approaches performed on the same airport. The available data for the analysis was limited to the accident flight FDR data. One objective was to identify if there was a standard practice regarding the altitude selection for the final approach. The detailed analysis of the fifteen previous flights to this airport showed first that the accident flight was the only one during which the crew did not select the go around altitude. However, this analysis of the previous fifteen flights also evidenced a lack of standardisation in the conduct of the approach, and the absence of adherence to the published approach chart. This study illustrated organisational issues and behavioural variability.

The development of tools able to process a huge quantity of data, from multiple sources and in a few hours, allows the investigator to look at the data in many different ways. It opens the perspective to further assess and communicate, with an increased reliability, potential contributing factors.

As an example, a FDM landing study undertaken by Brit Air for the operation of its Bombardier CRJ-700 was used during the safety investigation of an overrun which occurred on 16th October 2012. It revealed that 551 out of 1,680 landings on runways where the Landing Distance Available (LDA) is less than 2,400 m, occurred more than 650 m from the threshold. This classification also showed that among the 551 landings, the largest number of overshoots occurred at the aerodrome where the overrun happened.



The TSB of Canada also made references to the airline FDM in the Final Report concluded in March 2017 into the occurrence of an impact with terrain during approach which occurred on 24th February 2015 to a De Havilland DHC-8-102 operated by Jazz Aviation. The examination of this additional information enabled to determine that stabilized approach criteria exceedance raised during the investigation were not specific to the crew involved in the accident.

However, those examples of investigations using data other than the accident flight remain isolated. To widen the scope of safety investigations, data science now represents an opportunity which needs to be taken into account to improve the “risk management” oriented analysis approach. This can be accomplished in the fourth step of BEA analysis methodology by using for example the ATR Flight Operational Analysis described in the following paragraph.

**The use of Flight Operational Analysis (FOA) results in investigations**

The FOA project has been developed by ATR as a preventive tool to identify weak signals and provide operational recommendations. A number of airlines voluntarily cooperate to this project by providing with their flight data. As for any investigation, confidentiality of data is key, hence data is protected and results de-identified.

An increased knowledge in ATR aircraft performance, in system behaviour and in fleet-wide operations is gained from FOA studies. The objective of the FOA is to support the risk assessment in the frame of the SMS and assist the regional safety plan to focus on particular support activities. It can also be used in the design framework to review system reliability or to evaluate the behaviour of a new system. Furthermore, FOA can be very powerful to convey deep statistical analysis in a strong convincing communication (as illustrated below a comparison of speed management in approach within two different airlines).



FOA benefits became apparent during recent investigations. It provides a better understanding of what is considered actual operation with regards to safety expectations. It enables to have a different perspective, by looking at the event in the context of aircraft operations fleet wide. It also enables to compare airlines practices and potentially to identify specificities and organizational issues.

In the context of an investigation into severe propeller vibrations that occurred in 2014, it was evidenced that this event and other similar occurrences always happened in a similar operational context with the aircraft in descent, speed close to VMO and power levers retarded to flight idle. Although the related investigations could not yet establish the exact contribution of the operational context with the observed technical failure, it assessed that this context was necessary to the development of the phenomena at the origin of propeller system overload. The FOA was then used to evaluate the exposure of the fleet to this operational context. The analysis showed this specific operational context was encountered during 2% of the flights, which, extrapolated to the number of flights accumulated by the concerned fleet over a year (based on 2015 data), would correspond to 20,000 flights per year. The difficulty was then to interpret this data. One interpretation was that over 20 years of operation of this propeller type, there were 7 reported failure occurrences that looked similar while the specific operational context would have been encountered during 250,000 flights since the entry into service of this propeller. Hence, although the operational context appears necessary to generate the event, the occurrence rate compared to the exposure is very low. Therefore the analysis considers other factors contribute to the event in a more significant way than the operational context.

When adapting data science to a safety investigation, the investigators have to be aware of the risks associated to data bias and possible misinterpretation or improper extrapolation of data analysis results. A strong communication and common understanding of data sources and study objectives between the data analysts and the investigators are required, which is consistent with the first step of BEA analysis methodology. In the previous example, the analysis conducted was based on data with predominance of one airline. A policy to avoid VMO exceedance was implemented within this airline with a reduced maximum airspeed, hence potentially impacting the FOA results. This data bias was evaluated with the analyst and it was considered that the results were conservative based on the intended use of the results in the frame of the investigation.

The FOA was also used in the context of an on-going investigation to acquire a better understanding of the effort applied by pilots to manoeuvre the aircraft in the pitch axis during standard operation. The investigation revealed large efforts applied by the crew members in opposite directions. The only value of pilot effort that could be referred to in the literature is the maximum permissible effort given in CS/FAR 25.143 which is used for certification purpose. While the certification requirements specifies that the pilot effort to manoeuvre the aircraft shall not exceed 34 DaN, on the pitch axis (short term effort application), the FOA results showed that, for 95% of the flights analysed, the maximum effort applied out of take-off or landing phase, does not exceed 14 DaN. For comparison purpose, the maximum effort applied at take-off does not exceed 21 DaN, and at landing, it does not exceed 30.5 DaN. The maximum effort applied on the pitch axis, as recorded during the event, was 69 DaN and is significantly far from what the FOA shows as “standard” pilot input in flight.



For the same investigation, in addition to the analysis on pilot effort, an FOA study was conducted to identify the frequency of dual input on the fleet. A total of 25 dual opposite input occurrences were identified within 53,000 flights analysed, which led to consider an occurrence rate of 4.7 10-4 per flight cycle. Beyond the pure statistical results, the FOA also enabled to further analyse the circumstances of these events. For example, 85% of these events occurred during the approach and 60% were observed during turbulence encounter. This study is currently used by ATR to build up the appropriate materials for training, in order to provide feedback to crew members with regards to the threats associated with dual input and abrupt manoeuvre.

**Conclusion**

In the frame of safety investigations, data science provides high value factual data which should be systematically taken into consideration. A new task for the investigators is to identify all the potential sources of data and understand how to efficiently integrate it within a safety investigation. The profession needs to invest in these skills if we are to realise the potential of data science. The association of data science and a structured systemic analysis approach gives a different perspective on the occurrence in its context and facilitates the identification of organizational and systemic factors. This approach makes lessons learned and safety recommendations, not only more credible, accurate and relevant, but also more widely and easily accepted to be sure investigations really make a difference.

1. One example is the European Big Data Programme named “Data4Safety” [↑](#endnote-ref-1)
2. Méthode d’Investigation Organisationnelle et Systémique, developed by Dedale company (http://www.dedale.net/). [↑](#endnote-ref-2)