International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869 (O) 2454-4698 (P), Volume-9, Issue-10, October 2019

Exploring Validation Techniques to Ensure Correctness in Aeronautical Databases

Johnny Marques, Sarasuaty Yelisetty

***Abstract*—Aeronautical Data Chain is a conceptual representation of the path that a set or element of aeronautical data takes from its origination up to end-use. There are classical databases such as navigation or terrain database and configuration files for modularity and portability purposes. These databases can be used to activate or deactivate software items, adapt the software computation to the aircraft configuration, or provide computation data. The RTCA DO-200B establishes three Data Process Assurance Level (DPAL) as the level of rigor, representing the amount of verification and validation tasks performed, during data processing to assure data quality. This paper presents four different and complementary validation techniques to ensure correctness for Aeronautical Databases. This work aims to cover the existing gap in the available standards, presenting and exemplifying some proposed techniques. It also presents some techniques and provides different comparisons and specific intentions. At the end, some relevant needs of qualified tools are discussed to ensure that the use of techniques can be automated, and credits can be sought using tools.**

***Index Terms*—Aeronautical, databases, standards, validation, technique, tool.**

I. Introduction

According to Gao et al. [1] and Woodall et al. [2], due to the huge volume of generated data, the fast velocity of arriving data, and the large variety of heterogeneous data, the quality of data is far from perfect. Poor data quality used by embedded software makes significant effects in the development of safety-critical systems.

The development of safety-critical systems is usually part of a regulated environment. A software development error can directly cause losses of human lives or can have other catastrophic consequences. Some examples include systems that control aircrafts, nuclear reactors, and medical devices. The correctness of such software needs to be demonstrated with high assurance [3].

The society has become more and more reliant on software and database systems. Therefore, efforts to ensure software and database intensive systems must be reliable and safe. The aviation industry has a good track record, but as complexity increases, standardization is also required [4].

According to Hernandes (2013) [5], during architecture definitions of system products, databases are sets of data that influence the behavior of software without modifying executable codes managed as separate items.

According to Xie et al. [6], data validation is an important step to improve data quality. Almost all kinds of enterprises have begun to pay attention to big data validation.

In aviation, there are two types of embedded databases: Airborne System Databases and Aeronautical Databases [6]. The Airborne System Databases are considered Parameter Data items per the RTCA DO-178C [8].

Typically, such databases are always approved by certification authorities every time a new release is available. Aeronautical Databases are databases that are released in a small period and should follow the RTCA Do200B. In some Navigation Databases for Flight Management Systems (FMS), an update is required each 28 days. Thus, it is impossible to have a certification authorities' approval every release.

The RTCA DO-200B [9] is the standard for companies that produce Aeronautical Databases. It is recognized by the AC20-153B [10] and it provides recommended minimum requirements for the processing of aeronautical data. It aims to assist aeronautical data chain participants and regulatory authorities in meeting their responsibilities. The RTCA DO200B is intended to be used by organizations seeking approval of the method(s) they use to process or manipulate data. Basically, the approval is focused on methods, not products.

An initial oversight on verification methods of Databases under RTCA DO-178C and DO-200B was previously presented by the authors within a short paper approved and presented in 36th AIAA/IEEE Digital Avionics System Conference [11]. Now, this work presents validation techniques for Aeronautical Databases developed under the RTCA DO200B.

Each technique will provide a specific assurance and may be integrated to any methods available inside companies that release and distribute Aeronautical Databases. There is no guidance and strategies in standards on how to do it and this paper covers this gap. So, our contribution are the validation techniques to ensure correctness of aeronautical databases (e.g., navigation, terrain, obstacle, airport mapping, and other purposes).

This publication may help aeronautical data suppliers (e.g., data providers, application integrators, etc.), aircraft manufacturers, avionics manufacturers, and operators/end-users.

This paper is organized in another four additional sections. Section 2 briefly describes the standards for processing aeronautical data. Section 3 presents the techniques for ensuring correctness. Section 4 describes considerations about tools usage. Section 5 briefly describes a case study. Finally, Section 6 presents the conclusion.

Johnny Cardoso Marques, Computer Science Division, Aeronautics Institute of Technology, Sao Jose dos Campos, Brazil, +55(12)99152-2804
Sarasuaty Yelisetty, Computer Science Division, Aeronautics Institute of Technology, Sao Jose dos Campos, Brazil, +55(12)99143-9676

II. Standards for Processing Aeronautical Data

The RTCA DO-200B provides minimum requirements for all phases of the data process applicable to the processing of aeronautical data, including quality assurance. This standard provides guidance to assess compliance and determination of the levels of process assurance and supports the development of Aeronautical Databases.

According to the RTCA DO-200B, an Aeronautical Data Chain is a conceptual representation of the path that a set or element of aeronautical data takes from its origination up to its end-use. The RTCA DO-200B establishes three Data Process Assurance Level (DPAL) as the level of rigor, representing the amount of verification and validation tasks performed, during data processing to assure data quality.

For applications integrated into aircraft, the required DPAL is identified based upon the overall system architecture through allocation of risk determined by using a preliminary system safety assessment, as specified in Table 1.

Table 1 - The Failure Condition Categories and Associated DPALS [11]

Failure Condition Category	DPAL
Catastrophic	1
Hazardous	
Major	2
Minor	
No Safety Effect	3

Typically, the Aeronautical Data Chain involves many organizations. Data Providers are organizations responsible for data generated by them. Data Processors are organizations responsible for using data from Data Providers and generating their own data.

In a database system, there is not only recorded data itself, but also the whole definition of some tables' structure. In this structure, it is defined: table names, parameter definitions, storage formats, indexes that were created, and possible restrictions regarding data. All this information is defined in the literature as metadata. In the context of Aeronautical Databases, this information is expressed as Data Quality Requirements (DQR).

All data used to generate Aeronautical Database must meet DQR specified by Data Processors. Specific and generic DQRs are available from the RTCA DO-201A [13]. DQRs shall characterize the data by:

- Accuracy;
- Resolution;
- Confidence that data have not been corrupted while stored, processed, or transmitted (assurance level);
- Ability to determine the origin of data (traceability);
- Level of confidence that data are applicable to the period of the intended use (timeliness); and
- Format.

Figure 1 presents a typical Aeronautical Data Chain with the following 7 phases and, after each phase, some checks are made to ensure correctness:

1) Identify Data Source (IDS);
2) Assemble Data (AD);
3) Translate Data (TD);
4) Select Data (SD);
5) Format Data (FD);
6) Construct Data File (CDF); and
7) Distribute Data (DD).

However, the Construct Data File (CDF) phase is included by authors, basically because the database is generated using an additional intermediate XML format. This modified aeronautical data chain is presented in Fig. 1.

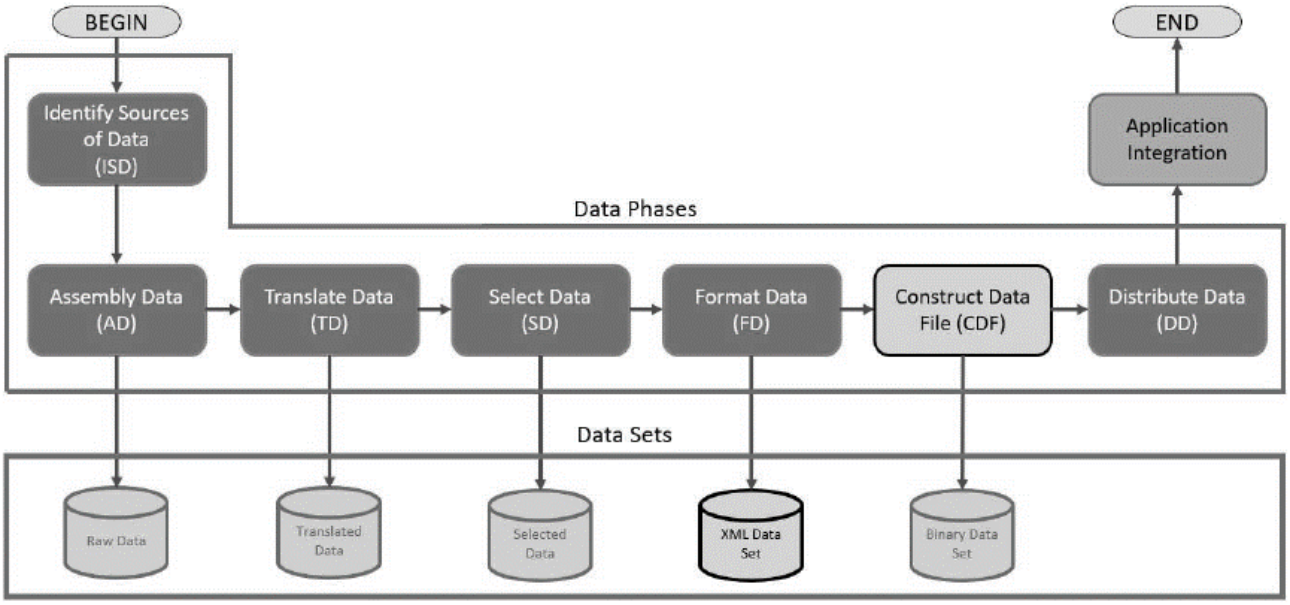

Fig. 1. The Aeronautical Data Chain

The IDS phase involves the identification and selection of sources of data that can support the Aeronautical Database Development. The AD phase involves the collection and collation of data from one or more suppliers. The TD phase involves the changes on how information is expressed. The SD phase selects the data to a specific configuration. The FD involves converting the selected data subset into a format that is acceptable to the next functional link in the chain, in this

paper, an XML format. This may take the shape of a published exchange standard format for the transmission of data, a proprietary format for loading in a target application, or another agreed format. The CDF phase uses the XML format and generate the Data file. The DD phase completes the processing data model and forms part of the transmission link of the Data file. It involves the delivery of the formatted data set to users. If errors or omissions are identified, they are reported to the appropriate participant within the data chain and procedures are followed to ensure that deficiencies are corrected and recorded for potential notification to data end-users.

According to the RTCA DO-200B, the purpose of the tool qualification process is to obtain confidence in the tool functionality. The tool qualification effort varies based upon the potential impact that a tool error could have on the system safety and upon the overall use of the tool in the software life cycle process.

The risk of a tool error adversely affecting system safety introduces the rigor required for tool qualification. The tool qualification process applies to one function, to a collection of tools, or even to all functions. Only tools that can insert or fail to detect an error in the aeronautical data process require qualifications. Error Detection Tools verify aeronautical for correctness. Typically, such tools could theoretically fail to detect an error.

Tools can be used to eliminate, reduce, or automate the activities associated with an aeronautical data chain. In this case, outputs not verified need to be qualified under the RTCA DO-330 [12].

III. Techniques for Ensuring Correctness

This paper presents four techniques used for ensuring the correctness of the Aeronautical Databases validation:

- Semantic Evaluation Technique (SET);
- Logical Consistency Technique (LCT);
- Feedback Check Technique (FCT); and
- Independent Redundancy Technique (IRT).

For each technique we defined different research questions:

- Are there attributes for each parameter respected?
- Are there parameters with dependencies respecting the mathematical rules?
- Are there two Data Sets, in different formats, equivalent?
- Do two different paths produce the same Data Set?

To appropriate explore the four techniques, we used a Database Structure with parameters, types, units, dependencies, and ranges, according to Table 2.

A. The Semantic Evaluation Technique (SET)

The main objective of the Semantic Evaluation Technique (SET) is to compare parameters included in the database to an expected value or range of values. This technique provides assurance that parameters defined contain appropriate values. This technique may involve tools that evaluate the Data File and parameter by parameter, to ensure the appropriate characteristics for a specific configuration. One characteristic is a set formed by type, unit, resolution, and range.

Basically, one of the 7 phases mentioned in Section 2 is the Data Phase that produces a Data Set. The SET compares a Data Set produced from a Data Phase with appropriate characteristics for each parameter included in the Data Set. Fig. 2 provides an example of such technique.

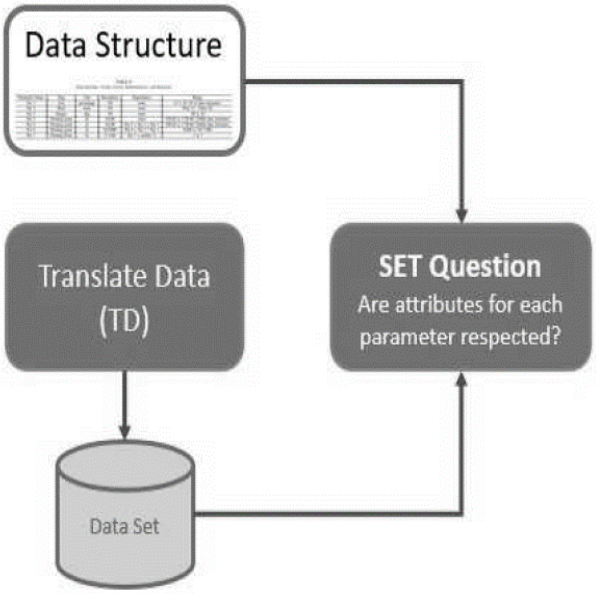

Fig. 2. An Example of the Semantic Evaluation Technique (SET)

As an example, parameters values are checked to ensure characteristics presented in Table 2. In this example, we compare some valid configurations with appropriate ranges, but SET may include other consistency checks as: i) presence or absence of data; ii) field and character context; iii) use in the declared time period of validity; among others. Table 2 presents one valid configuration evaluated by the SET.

The Par 3 has failed because value is out-of-range. Although there is a dependency between parameters Par 3 and Par 5, the last one is inside of the range and is considered PASS but is based in an unappropriated value of Par 3. Another error is on parameter Par 6, that uses Par 5 in calculation, but Par 6 is inside of the range and the SET did not catch this error. In this case, the dependency among parameters can be evaluated using the Logical Consistency Technique (LCT) described as follows.

B. The Logical Consistency Technique (LCT)

The Logical Consistency Technique (LCT) is used to validate parameters by comparing two different data sets or elements and identifying inconsistencies between values based upon operative dependencies between parameters. Although this method cannot completely validate data, as there is the possibility that different data sets include the same error, independence of data sets substantially improves the effectiveness of this type of validation.

In the example presented in Table 2, parameters Par 5, Par 6, and Par 7 are dependents from other parameters. So, to ensure that the calculation of these parameters is corrected

according to Attributes, a comparison among different Data Sets will provide assurance that the calculation is correctly performed. Fig. 3 provides an example of such technique.

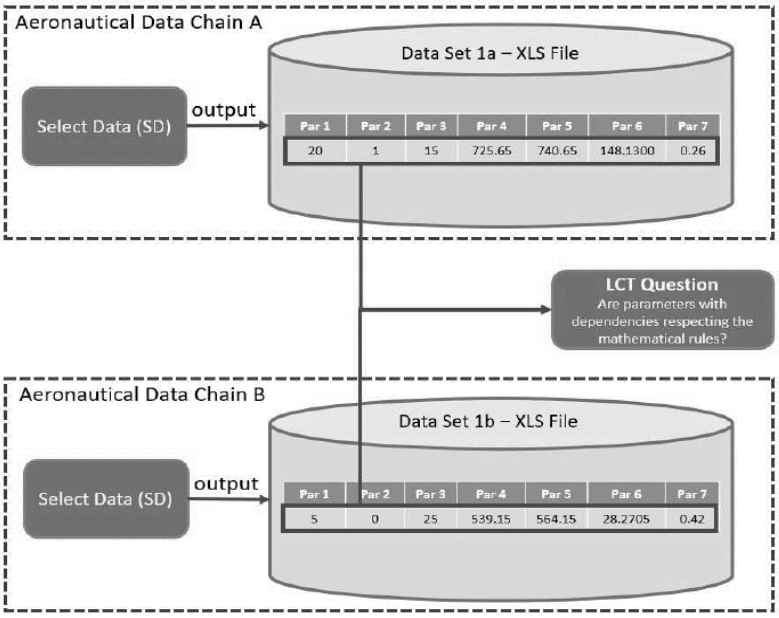

Fig. 3. An Example of the Logical Consistency Technique (LCT)

C. The Feedback Check Technique (FCT)

The Feedback Check Technique (FCT) involves the comparison between data sets. Basically, one Data Set 1 is used as an input of a Data Phase and a Data Set 2 is the output. A common method of FCT is done by manual confirmation, but in some case, when the amount of data is too big such technique may be considered unfeasible. To avoid this bottleneck, the usage of a qualified error detection tool can be a good decision.

Using parameters presented in Table 2, we present an example to compare two equivalent Data Sets from different formats. Basically, Data Set 1 is an

Excel XLS Spreadsheet with values for a configuration. This Data Set 1 is the input for the Format Data (FD) Phase. The output is the Data Set 2 that produces an equivalent configuration using an XML format. This example is presented in Fig. 4.

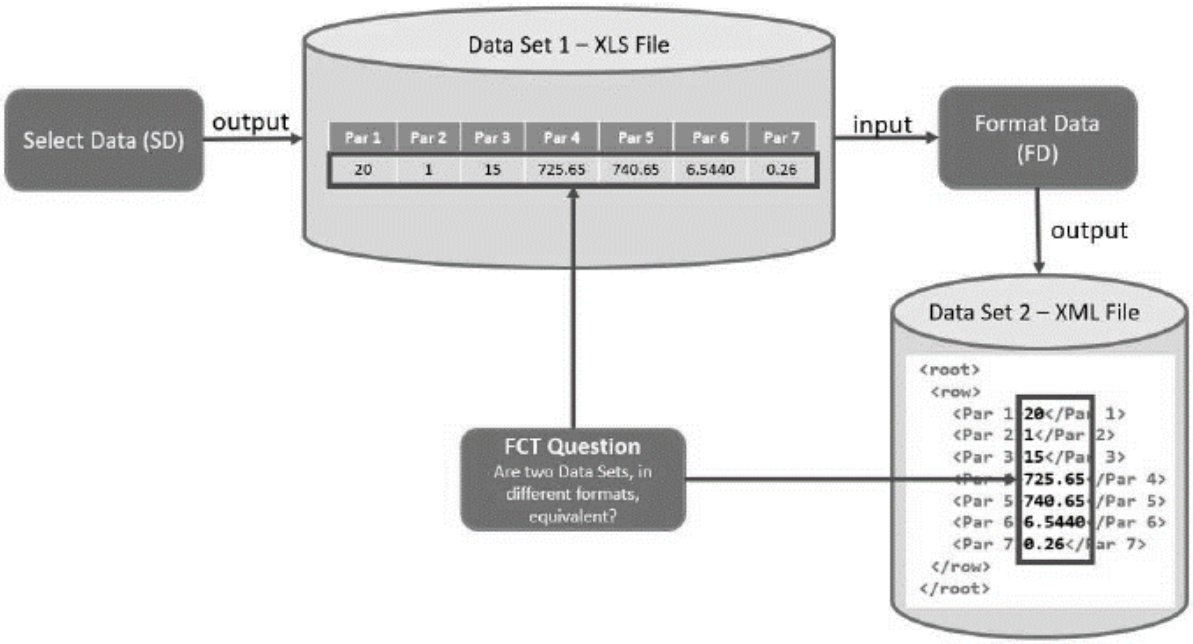

Fig. 4. An Example of the Feedback Check Technique (FCT)

Table 2 – Parameters, types, units, dependencies, and ranges

Parameter Name	Type	Unit	Resolution	Dependency	Range	Valid Configuration	
						Value	Range Check Results
Par 1	List	percentage	%0	none	0; 5; 10; 20 (4 data elements)	20	PASS
Par 2	Bool	none	%0	none	True (1) or False (0)	1	PASS
Par 3	Integer	deg	%0	none	-40 to 10	15	FAIL
Par 4	Floating point	kt	%0.00	none	438.60 to 1728.90 (25806 data elements)	725.65	PASS
Par 5	Floating point	kt	%0.00	Par 5 = Par 4 + Par 3	398.60 to 1738.90 (26806 data elements)	740.65	PASS
Par 6	Floating point	kt	%0.0000	Par 6 = Par 5 * Par 1	0.000 to 347.7200	69.5440	PASS
Par 7	Floating Point	kt	% 0.00	Par 7 = sin (Par 3)	-1 to 1	0.26	PASS

D. The Independent Redundancy Technique (IRT)

The Independent Redundancy Technique (IRT) involves processing the same data through two (or more) independent paths and comparing the data output. The idea is to achieve confidence that the Binary Data Set is correctly produced, using two different tools producing the same content. Fig. 5 presents this example.

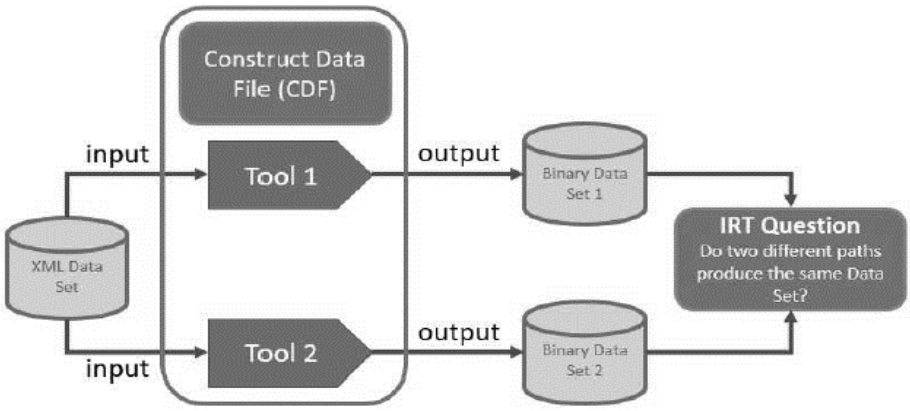

Fig. 5. An Example of the Independent Redundancy Technique (IRT)

IV. Considerations about Tools and Techniques

As presented before, when tools are used and confidence is assured by a tool, the qualification is mandatory, by using the RTCA DO-330. The benefits of tool qualification are described by Pothon (2013) [13]. For SCT, LCT, FCT, and IRT questions, we can use a tool to ensure:

- The comparison of parameters included in the Data Set to an expected value or range of values, thus automating the SCT Question;
- The comparison of two different Data Sets or elements and identifying inconsistencies between values based on operative dependency between parameters, thus automating the LCT Question;
- The comparison of two Data Sets in different formats to ensure they are equivalent, thus automating FCT Question; and
- The comparison of two Data Sets in the same format to ensure they are identical, thus automating the IRT Question.

The usage of Qualified Software Tools in Aeronautical Database development was presented by the authors in the 34th AIAA/IEEE Digital Avionics System Conference [15].

V. Case Study

The case study involves the release of five different Data Sets to be delivered to five different Flight Management System (FMS) suppliers. Data Sets contain different parameters and values. Figure 6 presents an overview of the case study.

Each Data Set was validated using the four techniques presented in this paper, and using some Qualified Tools internally developed by the companies involved. The tools and companies will not be identified in this paper to ensure the intellectual properties in their processes and developments. The results of the case study are presented in Table 3.

The usage of Qualified Tools will automatize the validation of the parameters to ensure that comparisons presented in Section IV are performed and errors are identified. Based on Table 4, we can see that Data Sets are very large and is almost unfeasible to perform a manual validation covering all parameters.

As an example, if one person uses 30 seconds to manual validate one parameter of Data Set A, he will need 4.75 years to finish all the validation alone, only for Data Set A. As said before, in some Navigation Databases for Flight Management Systems (FMS), an update is required each 28 days. Thus, it is unfeasible to have a manual process in place.

VI. Conclusion

This paper briefly describes four techniques: The Semantic Evaluation Technique (SET), the Feedback Check Technique (FCT), the Logical Consistency Technique (LCT), and the Independent Redundancy Technique (IRT). These techniques can be used to validate Data Sets included in Aeronautical Databases, ensuring their correctness.

This paper also covers an existing gap in the standards available, as no guidance and strategies are clearly defined in the literature.

The following four Research Questions (RQ) were identified in this work:

- RQ1. Are attributes for each parameter respected?
- RQ2. Are parameters with dependencies respecting the mathematical rules?
- RQ3. Are two Data Sets, in different formats, equivalent?
- RQ4. Do two different paths produce the same Data Set?

The main objective of the Semantic Evaluation Technique (SET) is to compare parameters included in the database to an expected value or range of values. This technique provides assurance that parameters defined contain the appropriate values. The SET fulfills and properly answers the RQ1. The Logical Consistency Technique (LCT) provides validation by comparing two different data sets or elements, identifying inconsistencies between values based upon operative dependency between parameters. The LCT fulfills and properly answers the RQ2.

The Feedback Check Technique (FCT) involves the comparison of a data sets. Basically, one Data Set 1 is used as an input of a Data Phase and a Data Set 2 is the output. The FCT fulfills and properly answers the RQ3.

The Independent Redundancy Technique (IRT) involves processing the same data through two (or more) independent paths and comparing the data output. The IRT also fulfills and properly answers the RQ4.

In addition, we have also identified that tools can be used to automate the identified techniques. In such cases, when confidence is assured by a tool, the qualification is mandatory per the RTCA DO-330.

VII. Future Work

Other guidelines must be provided to ensure different aspects of aeronautical database development, so the authors are working in the following future work:

- Databases should be considered as a component of the embedded software. Their development should be assessed using the RTCA/DO-178C and/or the RTCA/DO200B. Assurance should be provided that the associated software is tested for all accepted database, ensuring compatibility and proper integration among different versions of database with software products that uses data for their processing; and
- Unless otherwise justified by the system safety assessment process, the detection mechanism for partial or corrupted database loaded in an embedded computer should be assigned to the same failure condition or software level associated with the function that uses the software loaded. However, guidelines on how to define a failure conditions associated to partial or corrupted database are needed.

Acknowledgment

The authors would like to thank: 1) The Aeronautics Institute of Technology (Instituto Tecnológico de Aeronautica - ITA); and 2) The RTCA (Radio Technical Commission for Aeronautics), specially members of Special Committee SC-217.

Table 3 – Final results for the case study

Data Sets	Number of Parameters	Number of Errors Detected			
		Attributes	Dependencies	Format Change	Different Paths
A	5012534	2	25	0	1
B	4963571	3	32	0	0
C	3250524	5	11	0	2
D	4480200	0	5	0	0

E	5120254	1	19	0	1

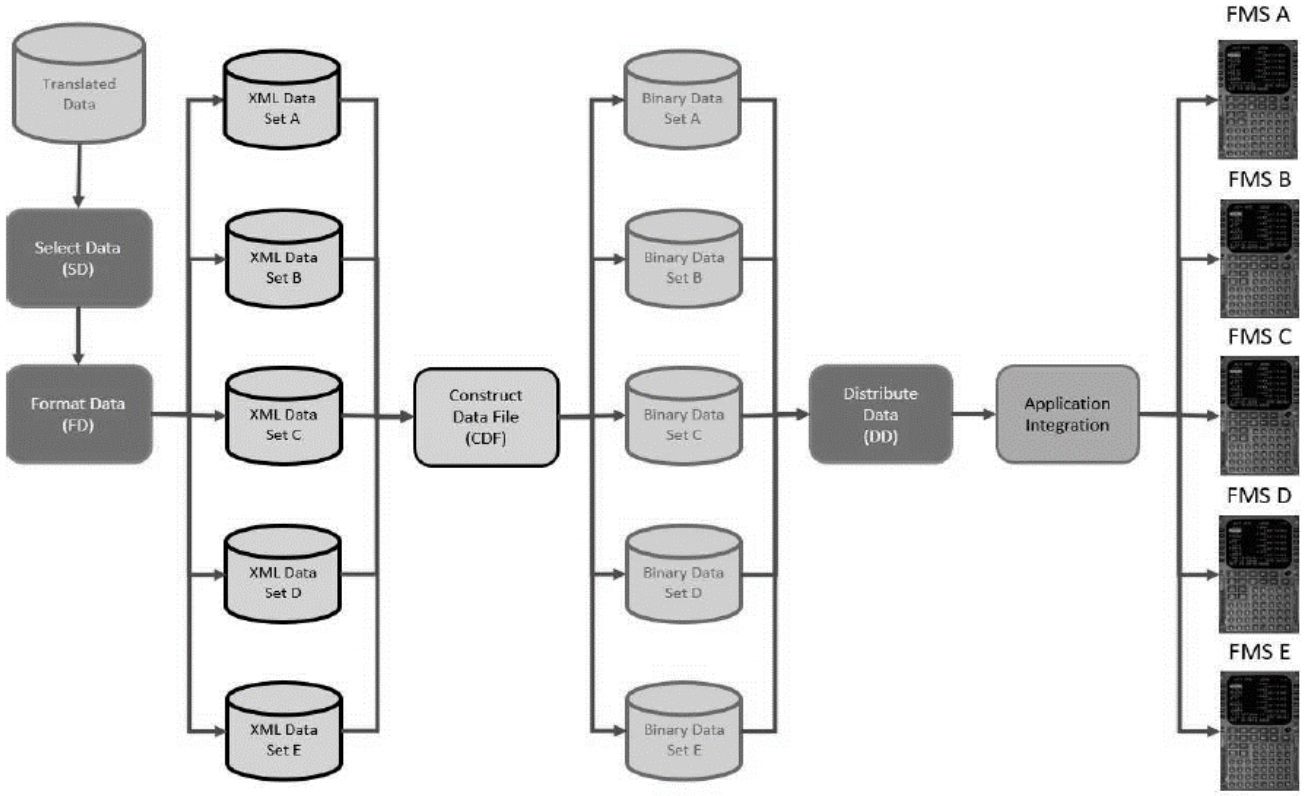

Fig. 6. Case Study

REFERENCES

[1] J. Gao, C. Xie, and C. Tao. "Big Data Validation and Quality Assurance--Issuses, Challenges, and Needs". In Proc. of 2016 IEEE Symposium on Service-Oriented System Engineering (SOSE), Oxford, UK, 2016, pp.433-441.

[2] P.Woodall P, J.Gao, A. Parlikad A, and A. Koronios. Classifying Data Quality Problems in Asset Management. Springer International Publishing, 2015.

[3] J. Marques, A. M. Cunha, Use of RTCA DO-330 in Aeronautical Databases, Prague, Czech Republic: 34th IEEE/AIAA Digital Avionics Systems Conference, 2015.

[4] L. Rierson, Developing Safety-Critical Software: A Practical Guide for Aviation Software and DO-178C Compliance CRC Press, 2013.

[5] M. Hernandes, Database Design for Mere Mortals: A Hands-On Guide to Relational Database Design Addison-Wesley Professional, 2013.

[6] C. Xie, J. Gao, C. Tao, Big Data Validation Case Study, United States, San Francisco: 3rd IEEE International Conference on Big Data Computing Service and Applications (BigDataService), 2017.

[7] Federal Aviation Administration, Order 8110-49 Software Approval Guidelines Change 2 United States, 2017.

[8] RTCA, DO-178C Software Considerations in Airborne Systems and Equipment Certification, Washington, United States, 2011.

[9] RTCA, DO-200B Standards for Processing Aeronautical Data, Washington, United States, 2017.

[10] Federal Aviation Administration, 20-153B - Acceptance of Aeronautical Data Processes and Associated Databases, Washington, United States, 2016.

[11] J. Marques, A. M. Cunha, Verification Scenarios of Onboard Databases under the RTCA DO-178C and the RTCA DO-200B, St. Pettersburg, United States: 36th IEEE/AIAA Digital Avionics Systems Conference, 2017.

[12] RTCA, *DO-330 Software Tool Qualification Considerations*, Washington, United States: RTCA, 2011.

[13] F. Pothon, *DO-330/ED-215 Benefits of the New Tool Qualification Document* United States, 2013.

Johnny Marques was born in Toronto, Canada, in 1977, but has been living in Brazil since 1986. He received the B.Sc. in Computer Engineering from University of the State of Rio de Janeiro (UERJ), the M.Sc. (in Aeronautical Engineering) and a PhD. (in Electronic and Computer Engineering) both from Aeronautics Institute of Technology (ITA). He is a current full professor in the Aeronautics Institute of Technology (ITA). Additionally, he worked at EMBRAER in software processes definition for 15 years and has recognized experience in standards used for airborne systems and software such as DO-178C, DO-254, ARP-4754 and DO-200B. He is also part of several committees in IEEE Standards Association.

Sarasuaty Yelisetty was born in Sao Jose dos Campos, Sao Paulo, Brazil. She received the B.Sc. in Computer Engineering from University of Vale do Paraiba (UNIVAP). In 2015, she finished M.Sc. program in Computer and Electronic Engineering in Aeronautics Institute of Technology (ITA). Additionally, she has been working at EMBRAER in software processes definition during the last 10 years and has recognized experience in standards used for airborne system and software such as DO-200B, DO-178C and DO-254.