Processo de Especificação de Banco de Dados em Sistemas Críticos

Sarasuaty Yelisetty[0000−0001−7973−6982], Johnny Marques[0000−0002−1551−435X], and Luiz Alberto Vieira Dias[0000−0002−6544−7458]

Instituto Tecnológico de Aeronáutica, São José dos Campos, SP, Brasil
sara.mhy@gmail.com; {johnny,vdias}@ita.br

Resumo O software de sistemas críticos em segurança, incluindo banco de dados, requer processos robustos de especificação, desenvolvimento, validação, verificação, gerenciamento de configuração e garantia de qualidade. A pesquisa reportada neste trabalho, em seu estágio atual, propõe a criação de um processo de especificação e validação de requisitos para bancos de dados em sistemas críticos. A metodologia utilizada envolve a condução de uma Revisão Sistemática da Literatura em processos de especificação de bancos de dados. Adicionalmente, inclui-se um diagrama do processo de especificação e validação de requisitos para bancos de dados em sistemas críticos que inclui a criação dos cenários de validação de requisitos. Espera-se que o processo possa ser utilizado por empresas desenvolvedoras de sistemas críticos e apresentará rastreabilidade para as necessidades regulatórias de alguns domínios selecionados como o aeronáutico, médico e ferroviário. Adicionalmente, o processo também permitirá uma construção organizada de bancos de dados orientados por requisitos, visando a mitigação de erros inseridos em sua construção e mecanismos de garantias desde a especificação até a integração com o software de aplicação.

Keywords: banco de dados · sistemas críticos · requisito · processo · validação

Nível: Doutorado
Ano de Ingresso no Programa: 2019/2º Semestre.
Previsão de Conclusão: 2023/1º Semestre.

1 Introdução

Segundo Hernandes (2013) [6] os bancos de dados são conjuntos de dados que influenciam o comportamento do software de aplicação sem modificar códigos executáveis e podem ser gerenciados como itens separados.

De acordo com Gao et al. (2016) [5] e Woodall et al. (2015) [17], devido ao volume de dados gerados, à velocidade rápida dos dados que chegam e à variedade de dados heterogêneos, a qualidade dos dados está longe de ser perfeita. A baixa qualidade dos dados consumidos por um software de aplicação produz efeitos indesejáveis significativos nos sistemas críticos.

O desenvolvimento de sistemas críticos (*safety-critical*) geralmente faz parte de um ambiente regulamentado. Um erro de desenvolvimento de software pode causar diretamente perdas de vidas humanas [10]. Alguns exemplos incluem sistemas que controlam aeronaves, reatores nucleares e dispositivos médicos.

A sociedade tornou-se cada vez mais dependente de software e sistemas de banco de dados. Portanto, os esforços para garantir sistemas com software de aplicação e seus bancos de dados devem ser confiáveis e seguros. A indústria da aviação tem um bom histórico no desenvolvimento de sistemas críticos, mas à medida que a complexidade aumenta, a padronização e processualização tornam-se cada vez mais necessárias [14].

As agências reguladoras exigem que estes produtos atendam requisitos rigorosos de certificação, incluindo software e bancos de dados. Um erro no desenvolvimento de software pode causar perdas de vidas humanas ou ter outras consequências catastróficas. A correção desse software precisa ser demonstrada com alta garantia de segurança, consequentemente essa preocupação estende-se aos bancos de dados [11].

Sistemas Críticos como um computador de navegação aérea ou um respirador mecânico com interface digital, ilustrados na Figura 1 possuem forte uso de dados que precisam ser garantidos quanto à completude, correteza e integridade desde a sua especificação até a sua integração com o software de aplicação consumidor destes dados.

(a) Computador de Navegação Aérea (b) Respirador Mecânico com Interface Digital

Figura 1. Exemplos de Sistemas Críticos com Uso de Banco de Dados

O software de aplicação, assim como os bancos de dados de sistemas críticos, requer processos robustos de especificação, desenvolvimento, validação, verificação, gerenciamento de configuração e garantia de qualidade [1]. A especificação de requisitos para dados, e suas validações, é um passo importante para melhorar a qualidade dos dados. Quase todos os tipos de empresas começaram a prestar atenção à validação de grandes volumes de dados [18].

Assim, o objetivo deste artigo é descrever o estágio atual de uma pesquisa de doutorado em andamento no Programa de Pós-graduação em Engenharia Eletrônica e Computação, área Informática do Instituto Tecnológico de Aeronáutica (ITA).

2 Informações sobre a Pesquisa

Por se tratar do uso de banco de dados em sistemas críticos, torna-se necessário fornecer garantias da correta especificação e validação de requisitos desses bancos de dados, mitigando seus impactos em segurança. Assim foi identificada a Questão Primária (QP) que busca saber **como garantir que os bancos de dados em sistemas críticos são corretos, completos e íntegros?**, que se refere ao estudo principal e como parâmetro para as questões secundárias e seus objetivos. Primeiramente, procura-se identificar as necessidades e problemas decorrentes do uso de bancos de dados incompletos, incorretos e sem integridade esperada.

A partir da Questão Principal formuladada, outras duas questões secundárias são de interesse desta pesquisa. A Questão Secundária 1 (QS1) busca saber **como especificar os requisitos adequadamente para construção de bancos de dados?**. A QS1 envolve a identificação das práticas existentes na especificação de requisitos de bancos de dados. Já a Questão Secundária 2 (QS2) busca saber **qual a regulamentação existente nos domínios críticos que são aplicáveis aos bancos de dados?**. A QS2 envolve exatamente a identificação das possíveis necessidades regulatórias existentes. A Tabela 1 organiza a QP, QS 1 e QS 2, com as motivações para suas formulações.

Tabela 1. Questões de Pesquisa

ID	Questão de Pesquisa	Motivação
QP	Como garantir que os bancos de dados em sistemas críticos são corretos, completos e íntegros?	Identificar as necessidades e impactos do uso de bancos de dados incompletos, incorretos e sem integridade esperada.
QS 1	Como especificar os requisitos adequadamente para construção de bancos de dados?	Identificar as práticas existentes na especificação de requisitos para bancos de dados.
QS 2	Qual a regulamentação existente nos domínios críticos que são aplicáveis aos bancos de dados?	Identificar a regulamentação que deverá ser atendida pelo processo de especificação de requisitos e validação de bancos de dados.

As questões apresentadas nesta pesquisa relacionam-se com que, atualmente, não existe um processo definido e disponível na literatura sobre como especificar e validar requisitos para construção de bancos de dados em sistemas críticos. Assim, busca-se garantir que todos os valores, estrutura, atributos são conservados e consistentes durante todo o processo de desenvolvimento desde o levantamento das fontes de dados que alimentarão o banco até a sua integração com o software de aplicação.

A partir das questões anteriormente enunciadas, elaborou-se a hipótese a ser comprovada durante este trabalho de pesquisa: **"É possível criar um processo de especificação e validação de requisitos para bancos de dados**

em sistemas críticos que atenda diversas orientações regulatórias e buscando que estejam corretos, completos e íntegros".

2.1 Objetivo

O objetivo geral desta pesquisa consiste em **"criar um processo de especificação e validação de requisitos para bancos de dados em sistemas críticos"**.

Os objetivos específicos apresentam, de forma detalhada, as ações que se pretende alcançar e estabelecem estreita relação com as particularidades relativas ao objetivo geral já apresentado. Desta forma, os seguintes objetivos específicos foram identificados:

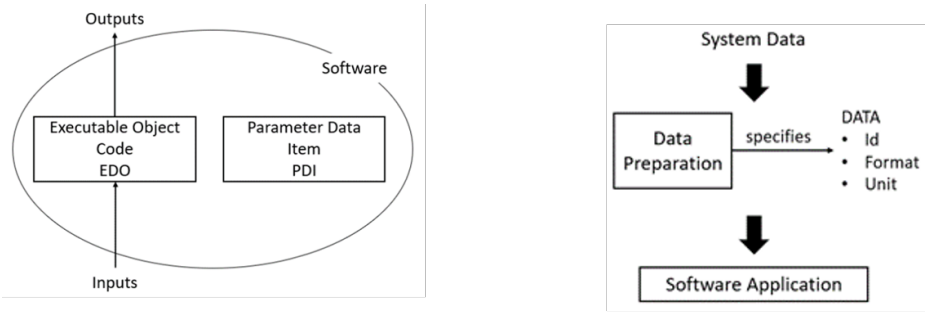
- Realizar um levantamento sobre a regulamentação existente em domínios críticos selecionados, a saber: aeronáutico, saúde e ferroviário;
- Realizar uma revisão sistemática da literatura que busque trabalhos correlatos;
- Apresentar rastreabilidade do processo proposto para a regulamentação existente; e
- Definir um conjunto de atividades, Requisitos de Qualidade dos Dados (RQDs) e Cenários de Validação (CVALs) que compõem o processo proposto.

3 Revisão de Normas de Sistemas Críticos

A RTCA DO-178C [15] é a norma existente para o meio aeronáutico e estabelece considerações para desenvolvedores, instaladores e usuários ao projetar um equipamento incorporado usando software e bancos de dados. Esta norma define cinco níveis de software. Cada nível de software foi definido em termos de objetivos que devem ser alcançados para aprovar o software e bancos de dados, como parte das certificações de aeronaves. Entre os cinco níveis de software (A, B, C, D e E), o nível A é o mais rigoroso. Esta norma define como Parameter Data Item (PDI) um componente de software que contém apenas dados e nenhum código executável, conforme apresentado na Figura 2(a). Esta norma prevê que um PDI seja especificado, revisado e verificado em sua estrutura, atributos e valores.

A norma IEC 62279 [8] é frequentemente utilizada no desenvolvimento de software e banco de dados para a área ferroviária [2]. Entre os quatro níveis (Software Integrity Level - SIL) de 1 até 4, o SIL 4 é o mais rigoroso. Segundo Boulanger (2015) [2], a partir dos dados do sistema, é possível especificar os bancos de dados com seus atributos e campos, que alimentarão uma aplicação genérica, permitindo uma determinada configuração, conforme apresentado na Figura 2(b). Na IEC 62279, os bancos de dados devem ser especificados, revisados e testados.

A norma IEC 62304 [7] é frequentemente utilizada no desenvolvimento de software e banco de dados para a área médica. A IEC 62304 requer que fabricantes atribuam uma classe de segurança para os sistemas com Software,



(a) RTCA DO-178C (b) IEC 62279

Figura 2. Relação entre Software e Banco de Dados na (a) RTCA DO-178C [15] e na (b) IEC 62279 [8]

incluindo seus bancos de dados. Esta classificação é baseada no potencial perigo que pode resultar em um prejuízo para o usuário ou o paciente, em caso de um comportamento anormal do sistema. Entre as três classes de software (A, B e C), a classe C é a mais rigorosa. Essa norma prevê que os bancos de dados sejam especificados, revisados e testados para construção dos bancos de dados envolvendo, principalmente, as características dos dados (por exemplo, numérico, alfanumérico, formato), intervalos válidos, limites e padrões.

4 Metodologia

Em consonância com o objetivo apresentado na seção 2.1, o processo proposto nesta pesquisa de doutorado será denominado como Processo de Especificação de Banco de Dados em Sistemas Críticos (PRE-BDC).

A metodologia definida para esta pesquisa e apresentada na Figura 3 consiste em 5 etapas, distribuídas em duas fases: (i) Desenvolvimento do Processo PRE-BDC; e (ii) Validação do Processo PRE-BDC. A etapa 1 consiste na condução de uma Revisão Sistemática da Literatura, seguindo o guia estabelecido por Felizardo et al. (2017) [3]. A etapa 2 envolve a concepção do diagrama do processo de especificação e validação de requisitos para bancos de dados em sistemas críticos. A etapa 3 envolve a criação dos cenários de validação de requisitos. A etapa 4 foca na execução de experimentos que validem o processo de especificação e validação de requisitos para bancos de dados em sistemas críticos. Finalmente, a etapa 5 avaliará a satisfação às normas previamente selecionadas como a RTCA DO-178C [15], IEC 62279 [8] e a IEC 62304 [7].

5 Processo PRE-DBC

O Processo de Especificação e Validação de Requisitos para Banco de Dados em Sistemas Críticos (PRE-BDC) envolve a sua estruturação em duas partes

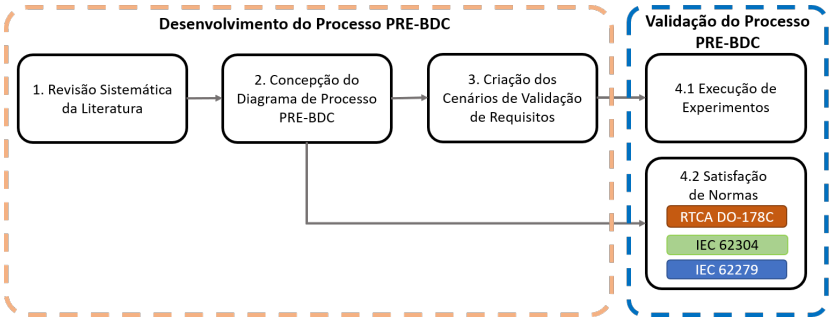


Figura 3. Etapas da Metodologia

principais: (i) os Requisitos de Qualidade dos Dados (RQDs); os Cenários de Validação (CVALs); e (iii) as atividades existentes desde a identificação dos dados até a integração com o software de aplicação.

Em um sistema de Banco de Dados, não há apenas os dados registrados, mas também toda a definição da estrutura de tabelas que compões o banco: (i) Nomes de tabelas; (ii) Definições de parâmetros; (iii) Formatos de armazenamento; (iv) Índices criados; e (v) Possíveis restrições em relação aos dados. Toda essa informação é definida na literatura como metadados [13]. No contexto deste trabalho de pesquisa, essas informações são expressas como Requisitos de Qualidade de Dados (RQDs).

Os Requisitos de Qualidade de Dados (RQDs) são definidos e inclusos em cada atividade do PRE-BDC. Espera-se que estes sejam definidos dependendo das necessidades existentes em cada atividade do projeto de banco de dados até a integração com o software de aplicação.

Cada atividade do processo PRE-BDC terá seu próprio conjunto de RQDs. Os RQDs envolvem: precisão; resolução; confiança de que os dados não foram corrompidos enquanto armazenados, processados ou transmitidos; capacidade de determinar a origem dos dados com rastreabilidade para suas fontes; nível de confiança de que os dados são aplicáveis ao período do uso pretendido; e formato. Adicionalmente, para cada conjunto de RQDs será criado um Cenário de Validação (CVAL) que garantirá que os dados gerados e/ou tratados em cada atividade do processo foram devidamente validados contra os RQDs definidos para a atividade. A Figura 4 apresenta o sequenciamento de 8 atividades previstas no processo PRE-DBC.

A atividade **Identificar as Fontes de Dados (IFD)** envolve a identificação e seleção de fontes de dados (manuais, regulamentos, outras bases de dados) que podem apoiar a especificação de um banco de dados como parte de um sistema crítico.

A atividade **Construir a Estrutura do Banco (CEB)** envolve a construção da estrutura em que os dados serão apresentados em forma de tabelas, campos, atributos e formatos. Nesta fase, a estrutura criada ainda não possui os dados, estes serão inseridos posteriormente.

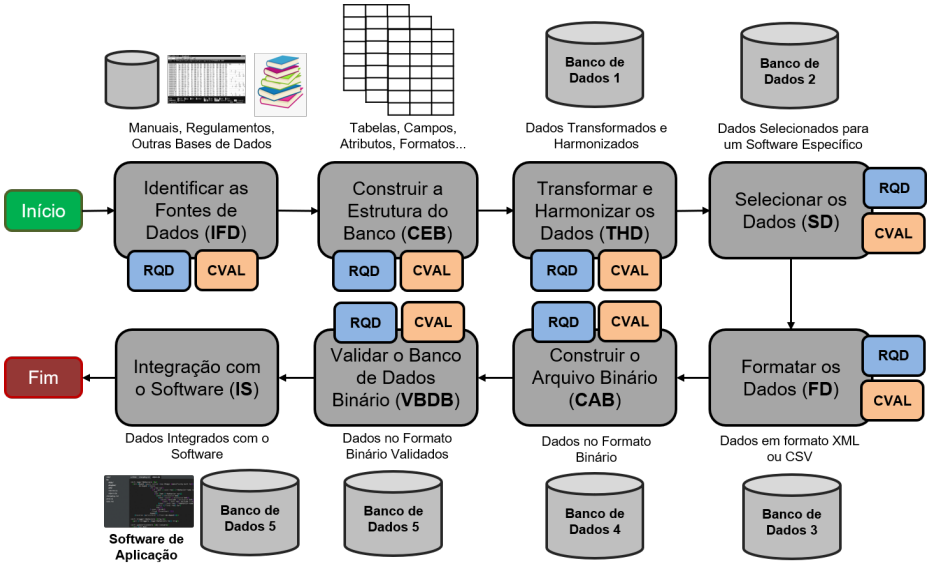


Figura 4. Processo de Especificação de Banco de Dados em Sistemas Críticos (PRE-BDC)

A atividade **Transformar e Harmonizar os Dados (THD)** envolve as mudanças em como as informações são expressas. Como os dados são provenientes de diversas fontes de dados, pode ser necessário transformá-los para que fiquem harmonizados. Nesta fase, os dados são inseridos nas estruturas definidas na atividade anterior (CEB), construindo a primeira versão do banco, denominada **Banco de Dados 1**.

A atividade **Selecionar os Dados (SD)** envolve a seleção de dados para uma configuração específica, e dependendo do projeto pode ser uma fase opcional. Dependendo do caso, pode ser necessário manter todos os dados dos bancos ou não. Caso essa atividade seja executada, o **Banco de Dados 2** será um subconjunto do **Banco de Dados 1**.

A atividade **Formatar os Dados (FD)** envolve a conversão do subconjunto de dados em um formato intermediário, como eXtensible Markup Language (XML) ou Comma Separated Values (CSV), em um formato aceitável para que essa representação seja utilizável em ferramentas que podem gerar o banco de dados num formato que seja entendido pelo software de aplicação, assim teremos o **Banco de Dados 3**.

A atividade **Construir o Arquivo Binário (CAB)** envolve a conversão de dados a partir do formato intermediário (**Banco de Dados 3**) em dados no formato binário visando a compatibilização com o software aplicativo, assim teremos o **Banco de Dados 4**.

A atividade **Validar o Banco de Dados Binário (VBDB)** tem por objetivo garantir que os atributos definidos na fase CEB foram respeitados, garantir

que dados gerados a partir de uma função matemática estão corretos nas diferentes tuplas (registros), garantir que dados em formatos diferentes são equivalentes e garantir que as possíveis ferramentas utilizadas neste processo, não introduzem erros ao produzir e/ou transformar dados. Portanto, o **Banco de Dados 5** possui os dados já validados a partir do **Banco de Dados 4**.

A atividade **Integração com o Software (IS)** envolve o recebimento do **Banco de Dados 5** para integração com o software. Esta fase não fará parte deste trabalho de pesquisa, no entanto, está representada na Figura 3 com o intuito de demonstrar que o banco de dados, especificado e validado pelo processo, está pronto para ser utilizado pelo software.

6 Contribuições Esperadas

Conforme apresentado na seção 2.1, este trabalho tem como objetivo principal a criação de um processo de especificação de requisitos para bancos de dados em sistemas críticos. Espera-se que o processo possa ser utilizado por empresas desenvolvedoras de sistemas críticos em diversos domínios. O processo apresentará rastreabilidade para as necessidades regulatórias de alguns domínios selecionados como o aeronáutico, médico e ferroviário. Adicionalmente, o processo também permitirá uma construção organizada de bancos de dados orientados por requisitos, visando a mitigação de erros inseridos em sua construção e mecanismos de garantias desde a especificação até a integração com o software de aplicação.

7 Comparação com Trabalhos Correlatos

Fernandez et al. (2008) [4] apresentaram algumas idéias sobre como garantir que o banco de dados tenha o mesmo nível de segurança que o software de aplicação. Os autores apresentaram duas maneiras possíveis de olhar para o problema de desenvolvimento de software com bancos de dados: (i) construir um banco de dados independente do software de aplicação; ou (ii) construir um banco de dados que seja parte do software de aplicação. Nas duas maneiras, os autores fizeram considerações sobre a necessidade de garantia que os dados populados no banco estejam corretamente especificados e validados. No entanto, o processo PRE-BDC irá utilizar apenas uma das maneiras que é a construção do banco de dados independente do software de aplicação.

Irfan & Zhu (2011) [9] reforçaram que os sistemas de banco de dados têm muitas restrições, e que não é possível controlar um banco de dados em tempo de execução. No entanto, os autores informaram que é possível coletar requisitos orientados para os valores que alimentam os bancos de dados. Assim, eles reforçaram a necessidade de extração de requisitos de qualidade das partes interessadas, criando uma abordagem orientada aos valores especificados e o como que esses valores são considerados corretos. Já o processo PRE-BDC identificou a necessidade de especificação de Requisitos de Qualidade de Dados (RQDs) definidos em cada atividade do processo PRE-BDC.

Souza et al. (2019) [16] discutiram aspectos relacionados à aplicação do conceito de flexibilidade a uma classe específica de sistemas aviônicos: sistemas baseados na arquitetura *Integrated Modular Avionics (IMA)* fornecida por um dos principais fornecedores de sistemas aviônicos, tipicamente usados em aeronaves regionais e executivas. Os tópicos cobertos incluem uma estrutura para a implementação da abordagem de flexibilidade, baseado em *Parameter Data Items* pela RTCA DO-178C e as principais preocupações de certificação. No processo PRE-BDC, haverá uma maior cobertura regulatória com uma análise de atendimento às principais normas de ambientes regulados já apresentadas na seção 3.

Mousa et al. (2020) [12] apontaram em seu trabalho que os perigos que um de banco de dados pode estar exposto refere-se a um item, indivíduo ou outra entidade que representa um risco de uso indevido ou manipulação de dados confidenciais. Os autores reforçam que a maioria dos recursos de segurança do banco de dados deve ser desenvolvida para proteger o ambiente do banco de dados. O trabalho dos autores focou nos desafios relacionados à integridade e segurança das informações existentes em um banco de dados. O processo PRE-BDC incorpora os desafios apontados pelos autores.

8 Considerações Finais e Estado Atual da Pesquisa

Este artigo descreveu os primeiros resultados de uma pesquisa de doutorado em andamento no Programa de Pós-graduação em Engenharia Eletrônica e Computação, área Informática do Instituto Tecnológico de Aeronáutica (ITA).

As três primeiras etapas da metodologia apresentada na seção 4 já foram executadas e finalizadas. A revisão sistemática da literatura propiciou a visualização sobre o estado da arte no processo de desenvolvimento de banco de dados e identificou uma carência de pesquisas que foquem na especificação e validação de requisitos para bancos de dados de forma geral. Embora a pesquisa de doutorado em andamento foque em sistemas críticos, foi possível perceber a baixa disponibilidade de trabalhos sobre o processos para desenvolvimento de bancos de dados de uma forma geral.

Atualmente, o foco está nas etapas 4 e 5. Na etapa 4, encontra-se em planejamento a execução de um experimento que prevê o uso do processo PRE-BDC na construção de um banco de dados de um sistema de gerenciamento de voo (*Flight Management System*-FMS) que inclui um software de aplicação e um banco de dados com parâmetros de pouso e decolagem de aeronave (*Take-off and Landing Database*-TOLD). Já na etapa 5, será executada uma análise de satisfação das normas, com uma matriz de rastreabilidade do processo PRE-BDC para as normas indicadas.

Referências

1. Barros, L., Hirata, C., Marques, J., Ambrosio, A.M.: Generating test cases to evaluate and improve processes of safety-critical systems development. In: 2020 IEEE In-

ternational Symposium on Software Reliability Engineering Workshops (ISSREW). pp. 311–318 (2020). https://doi.org/10.1109/ISSREW51248.2020.00090
2. Boulanger, J.L.: CENELEC 50128 and IEC 62279 Standards. Wiley (2015)
3. Felizardo, K.R., Nakagawa, E.Y., camargo Pinto Ferraz Fabbri, S., Ferrari, F.C.: Revisão Sistemática da Literatura em Engenharia de Software: Teoria e Prática. GEN LTC (2017)
4. Fernandez, E.B., Jurjens, J., Yoshioka, N., Washizaki, H.: Incorporating database systems into a secure software development methodology. In: 2008 19th International Workshop on Database and Expert Systems Applications. pp. 310–314 (Sep 2008). https://doi.org/10.1109/DEXA.2008.100
5. Gao, J., Xie, C., Tao, C.: Big data validation and quality assurance– issues, challenges, and needs. In: 2016 IEEE Symposium on Service-Oriented System Engineering (SOSE). pp. 433–441 (2016)
6. Hernandes, M.: Database Design for Mere Mortals: A Hands-On Guide to Relational Database Design. Addison-Wesley Professional (2013)
7. IEC: 62304:2015 amd 1 medical device software - software life-cycle processes – amendment 1. Tech. rep., International Electrotechnical Commission (2015)
8. IEC: Iec 62279:2015 railway applications - communication, signaling and processing systems - software for railway control and protection systems. Tech. rep., International Electrotechnical Commission, Washington (2015)
9. Irfan, M., Zhu, H.: Key role of value-oriented requirements to develop real-time database systems. In: 2011 IEEE 2nd International Conference on Computing, Control and Industrial Engineering. vol. 1, pp. 405–408 (Aug 2011). https://doi.org/10.1109/CCIENG.2011.6008044
10. Marques, J.C., Cunha, A.M.: Ares: An agile requirements specification process for regulated environment. International Journal of Software Engineering and Knowledge Engineering **29**(10), 1403–1438 (2019)
11. Marques, J., da Cunha, A.M.: Verification scenarios of onboard databases under the rtca do-178c and the rtca do-200b. In: 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC) (2017)
12. Mousa, A., Karabatak, M., Mustafa, T.: Database security threats and challenges. In: 2020 8th International Symposium on Digital Forensics and Security (ISDFS). pp. 1–5 (June 2020). https://doi.org/10.1109/ISDFS49300.2020.9116436
13. Osnu, M.T., Valduriez, P.: Principles of Distributed Database Systems. Springer (2020)
14. Rierson, L.: Developing Safety-Critical Software: A Pratical Guide for Aviation Software and DO-178C Compliance. CRC Press (2013)
15. RTCA: Do-178c software considerations in airborne systems and equipment certification. Tech. rep., Radio Technical Commission for Aeronautics, Washington (2011)
16. Souza, T., Souza, M., Rodrigues, P.: A flexibility framework for oem certification of software modifications in aircraft systems. In: 2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC). pp. 1–5 (Sep 2019). https://doi.org/10.1109/DASC43569.2019.9081761
17. Woodall, P., Parlikad, A., A. Koronios, A.: Classifying Data Quality Problems in Asset Management. Springer International Publishing (2015)
18. Xie, C., Gao, J., Tao, C.: Big data validation case study. In: 3rd IEEE International Conference on Big Data Computing Service and Applications (2017)