

飞行试验工程大数据治理思考

党怀义

(中国飞行试验研究院, 西安 710089)

摘要: 飞行试验工程大数据是典型的工业大数据, 是试飞工程规划、设计、执行、评估以及开展航空科学研究的最重要的基础; 文章分析了试飞工程大数据的质量特性及其影响因素, 学习借鉴国际标准化《数据治理白皮书》提出的数据治理思想体系与模型, 针对试飞工程大数据管理与应用特点, 结合多年来在试飞工程中的大数据管理应用的实践经验, 提出了以飞行试验大数据标准化体系为基础, 以涵盖试飞工程全过程和全业务流程的试飞数据质量监控系统 and 一体化的试飞大数据管理与应用系统为并行相互支持的大数据治理技术平台, 将试飞业务流程、业务策略、业务标准、业务逻辑以及组织管理有机有效地融入到数据管理与治理体系中, 形成能够不断自我完善、自我更新、自我规范、开放共享的试飞工程大数据治理体系, 对飞行试验工程以及航空科学研究步入“大数据科研范式”奠定数据基础。

关键词: 飞行试验工程; 大数据; 数据治理; 数据管理; 体系

Considering of Flight Test Engineering Big Data Governance

Dang Huaiyi

(Chinese Flight Test Establishment, Xi'an 710089, China)

Abstract: Flight test engineering big data are one of the typical industry big data, and are the most important data resources for the flight test engineering programming, designing, implement, assessment and aerial science research. It analyses the quality properties and their influencing factors, on the basis of using for reference of the data governance models and methodology system of the ISO 《White Book on Data governance》, pointing to the flight test engineering data management and application characters, relying on years experience about flight test data management and application in the flight test, puts forward one of the most effective data governance systems about flight test engineering big data; on the basis of the flight test big data standards, setting up a flight test big data governance platform which combines with the flight test data quality monitoring system and flight test data management and application integrating system, dissolving effectively and organically with flight test operation processes, operation strategy principles, operation logicalness, organization management and so on in the systems during the whole flight test program process period. The flight test engineering big data governance systems are more appropriate to flight test engineering and aerial scientific development, as the systems being improved on, updated continually, standardized roundly and shared widely without intermission, and are the most important basis of the aerial research paradigm of Big Data—Data—intensive scientific discovery.

Keywords: flight test engineering; big data; data governance; data management; system

0 引言

近年来, 随着信息技术的发展, 大数据成大数据已经成为世界各国持续发展、提升综合竞争能力的强力助推剂, 提升到国家战略层面大力进行产业化研究、推动和发展, 大数据范式已经成为科技发展的第四科研范式^[1]。1980年, 阿尔文·托夫勒等人预知大数据时代即将到来^[2], 与当时信息化方兴未艾所涉及的信息爆炸其内涵应该是相同的。正如所料, 人类已经由以IT(信息技术)为代表的信息技术时代转为以DT(数据技术)为关注焦点的数据信息技术时代, 大数据时代的核心是以大数据为基础的大数据分析、处理和预测^[3], 数据已经成为企业组织最大的价值来源和最重要资产, 数据模型是支撑全企业业务架构数字化平台与全管理架构信息化平台的重要基础, 该数据模型的组成

结构、数据流向、逻辑关系、数据内容、数据状态等, 形成了数字孪生企业, 以数字形式反映企业的业务全貌。因此, 数据质量直接关系到基于数据的企业管理、业务决策、创新发展等工作的正常、准确开展, 直接影响数据应用价值的发挥。航空飞行试验是航空产品研制过程中的重要一环, 而飞行试验数据的形成过程及组成结构极为复杂, 飞行试验数据治理是确保试验数据信息全面、完整、科学、准确的重要手段, 也是开展以试验大数据为基础的航空技术发展研究的重要基础, 本文针对飞行试验工程及航空技术发展应用需求, 提出了一套完整的飞行试验数据治理体系, 对于以试验大数据为基础的数字化试飞、试验大数据科研价值挖掘以及促使航空技术快速步入大数据科研范式的发展之路, 具有非常重要和深远的意义。

1 飞行试验工程大数据及其质量

航空飞行试验是航空产品由成品走向产品的最后一个环节, 涉及到航空产品的设计、制造、试验及使用等多个组织、环节、专业的综合协同, 试验对象的研制本身是一

收稿日期: 2019-01-05; 修回日期: 2019-01-29。

作者简介: 党怀义(1967-), 男, 陕西合阳人, 研究员, 主要从事飞行试验数据处理与软件技术方向的研究。

个复杂的系统工程，而飞行试验更是一个基于复杂对象的复杂系统工程^[4]。如图1所示。飞行试验具有周期长、专业广、难度大、工程复杂、风险高、耗资巨等特点，因此成为很多国家望航空制造业而却步的最大障碍之一。在航空产品的规划研制初期，针对航空产品功能、性能以及系列使用指标的飞行试验即开始进行策划，并进行早期设计。在航空产品研制过程中，飞行试验是航空产品不断验证、完善、再验证的循环过程，即使航空产品服务于用户，飞行试验的使命仍然在继续，不仅为用户提供更好、更全面的产品使用服务，同时利用基于历史试验数据的验证试验，开始了下一代航空产品的研究发展。



图1 飞行试验过程组成结构示意图

航空产品飞行试验工程数据即围绕试验对象，涵盖飞行试验工程全过程所形成的数据。航空产品飞行试验工程数据是航空产品数据的不可分割的重要组成部分之一，不仅直接支撑航空产品的研制鉴定/定型，同时对产品的改进、模拟仿真、飞行员培训、新技术探索与预先研究、专业工程建模与优化等潜在巨大的科研价值。试飞工程本身的特点，决定了试验数据的本质特征，不但具有目前普遍认可的大数据特征之外，还具有精确性、复杂性、非结构化特征^[5]，以及本身所具有的科学性、系统性、时序性等显著特点。实际上，对于航空飞行试验来说，大数据问题一直存在，伴随着人类对航空技术的不断探索与解决问题的能力的不断提升而交替发展。如图2所示，试飞工程大数据的采集形成过程就是飞行试验系统工程的复杂运行过程^[6]。

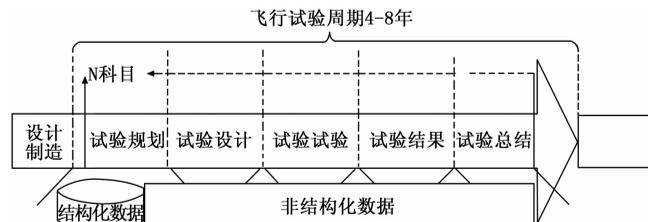


图2 航空产品试飞工程数据采集形成过程

飞行试验数据是科学大数据，具有典型工业大数据特征，工信部研究报告《2017工业大数据白皮书》指出，工业大数据不仅具有大数据的特征，还具有价值性、实时性、准确性和闭环性的典型特征，以及业界普遍认为的集成性、预测性、科学性等特征，分别从内涵、属性、用途、关系、来源等方面对工业大数据的特性进行了描述，工业大数据来源于工程实践，又是工程实施、改进、完善和发展的重

要基础，所以对工业大数据的质量要求非常高。航空飞行试验工程所形成的大数据更是如此。

多年的管理和应用研究表明，航空产品飞行试验数据的质量评估特征主要有：

①完整性。覆盖航空产品飞行试验工程全过程（包括由于试验过程中产品、试验方法等不断完善而反复进行试验所产生的相对应版本的数据），完整、全面地表现试验工程管理和试验过程信息。

②准确性。客观、真实地反映飞行试验的全过程，尽量减少和避免人工或系统产生的错误、误差等影响。

③规范性。存储或记录格式遵从规范。

④一致性。内涵或数据值无冲突。

⑤可追溯性。数据之间的关系明确，从试验过程的开始，到试验过程的结束，可由任何一个数据向前或向后进行工程关系的追溯。

而影响飞行试验工程数据的质量特性的因素主要有以下几个方面：

①试飞工程持续时间长，试验场分布多，试验数据形成时空跨度大。

②航空试验涉及到管理和众多工程专业，试验内涵深、外延广，数据形成过程及其组成复杂。

③试飞工程涉及的环节多、工作面广，数据来源既有众多的工程管理部门，还包括多个技术协同协作部门和人员，同时还有海量的工程测试数据，试验数据采集来源广。

④试飞工程试验数据组成复杂，既有结构化数据，还要非结构化数据，更多大量的科学时序工程数据，而且很多类型的数据属于专用类型，需要专业的数据处理软件进行解析，数据类型非常之多。

⑤试验数据量巨大，以大型客机C919为例，平均单架次各类试验数据总量将近1TB，并且与试验工程数据之间具有复杂的关联关系。

要实现飞行试验数据治理目标，需要针对飞行试验数据的质量特性及其影响因素，全方位围绕试飞工程业务，建立体系化的试验数据治理方案。

2 数据治理及其发展

数据是业务的核心，数据治理是伴随数据的产生而存在的。而进入DT时代，数据治理问题随着大数据发展而更为突出，越来越多的企业意识到数据治理的必要性、重要性和紧迫性，越来越希望通过进行数据治理来创造更多的业务价值。数据治理是确保数据的完整性、规范性、一致性的重要方法，确保以数据为核心的业务流程的准确执行和企业业务恰如其分的正确决策，解决困扰企业的信息孤岛问题、数据质量问题、应用效率问题、数据安全问题等一系列问题，为建立现代智慧企业提供数据保障。国外大型信息技术企业，从上世纪信息时代发展初期就意识到这一问题，持续投入巨大的人财物资源开展数据治理研究工作，IBM公司从1992年开始，形成了一整套完整的数据治

理技术体系，发展出一系列用于数据治理的数据高附加值产品。近年来，数据治理及相关标准体系的研究也成为国内各行各业研究的热点之一，我国数据治理技术发展取得了一定成绩，在国际标准化工作组 SC40/WG1（IT 治理工作组）提出了《数据治理白皮书》国际标准研究报告，并获得了认可，同意由中国国家委员会负责进行相关国际标准指南的编写工作^[7]。

《数据治理白皮书》给出了标准化的数据治理模型，如图 3 所示。

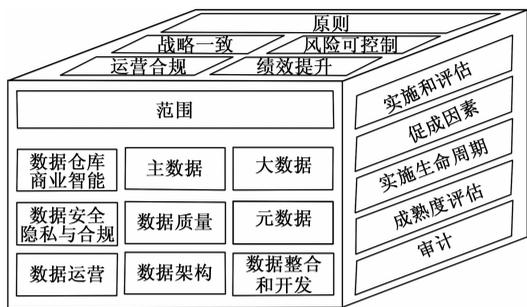


图 3 标准的数据治理模型

白皮书给出了标准的数据治理模型由 3 个框架组成：原则、范围和方法，即数据治理的三要素：数据治理的基本原则、数据治理的范围以及数据治理实施评估方法，全面概括了数据治理的要素，描绘出了数据治理的基本思路和技术蓝图。

业界的数据治理标准（ISO38505—2）给出的数据治理基本过程如图 4 所示。

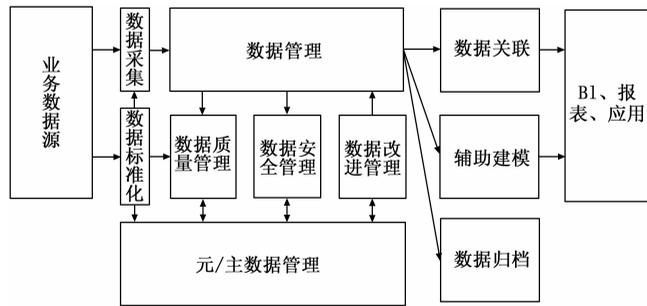


图 4 标准的数据治理过程

数据治理过程就是围绕涵盖业务开展相关全过程的数据源，以符合专业应用的元数据/主数据标准规范为基础，按照确定的业务流程和关系进行业务数据采集、管理和应用，覆盖业务的全过程，扩展到业务数据的生命周期。参考标准的数据治理过程，针对飞行试验工程特点，建立飞行试验大数据治理体系。

3 飞行试验大数据治理体系

围绕飞行试验大数据治理特性，针对影响飞行试验大数据质量特性的诸多因素，利用先进的数据治理技术、理论、方法等体系思想，贯穿飞行试验系统工程的全过程，建立飞行试验大数据治理技术管理与应用体系，为飞行试

验工程技术的发展，乃至航空工业技术的发展提供数据基础。飞行试验大数据治理体系，通过将试飞工程业务流程、业务策略、业务标准、业务逻辑以及试飞工程组织管理体系进行有机、有效地设计、管理和实施，在试飞工程信息化业务系统中全面实现和执行，提升试验大数据的质量特性。

3.1 飞行试验大数据标准体系

飞行试验大数据标准体系是开展飞行试验大数据治理的重要技术保障，是支撑以试验大数据为基础的试飞工程业务及其数字化应用发展的重要基础。飞行试验的大数据标准体系组成结构如图 5 所示。

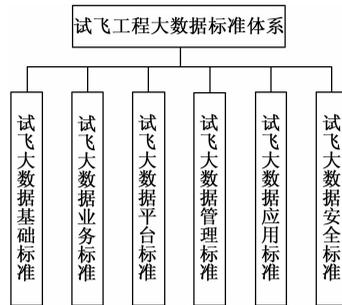


图 5 飞行试验大数据标准体系

飞行试验的大数据标准体系由试飞大数据基础标准、试飞大数据业务标准、试飞大数据平台标准、试飞大数据管理标准、试飞大数据应用标准和试飞大数据安全标准等 6 部分组成，标准的范围和内容涵盖了飞行试验大数据基本定义、产生、管理、存储、安全、应用等生命周期全过程。

①试飞大数据基础标准主要描述试飞大数据治理的基本要求、基本概念、基本术语、元数据等基础标准，规范和统一试飞工程大数据的标准体系。

②试飞大数据业务标准围绕试飞业务流程，按照大数据基础标准的规范要求，针对试飞大数据的采集、预处理、二次处理/综合处理、分析评估等业务数据进行规范。试飞大数据业务标准是试飞大数据治理标准体系的重要组成部分，以试验对象数据为核心，按照飞行试验鉴定/定型/适航技术要求，包括试验任务的规划、设计、仿真、验证、执行与监控、分析与评估、报告与结论等全过程。

③试飞大数据平台标准针对试飞大数据的技术产品和应用平台进行规范。包括结构化/非结构化数据库、智能分析工具、分析处理平台、数据应用等方面。数据平台是管理、维护大数据试飞业务关联关系准确性、可靠性和一致性的基础。主要包括平台架构、存取方式、分析、处理等标准规范，还包括试飞业务内部及设计、制造和维修等外部之间进行数据交换的各类数据的接口规范。

④试飞大数据管理标准主要描述试飞大数据产生和形成的全过程管理要求，包括大数据的采集、存储、传输、备份、再利用以及销毁等全生命周期的规范要求。

⑤试飞大数据应用标准针对试飞业务内、外提供试飞大数据的应用和服务进行规范，包括技术、功能、开发、维

护、管理等方面。也就是对于航空产品设计、制造、试验和维修等不同用户的数据集规范、应用范围、管理要求等。

⑥试飞大数据安全标准是试飞大数据治理标准体系的支撑，贯穿于试飞大数据生命周期，主要包括试飞大数据安全和保密两类。

3.2 涵盖试飞业务流程的数据质量监控

针对影响飞行试验工程数据的质量特性的诸多因素，涵盖试飞业务流程全过程，从组织管理、业务流程、技术管理等多方面建立飞行试验大数据的质量管控体系，达到飞行试验大数据的完整性、准确性、规范性、一致性和可追溯性的质量管理目标。

试飞工程是航空产品研制系统工程的重要组成，试飞工程本身是一个复杂的系统工程，飞行试验大数据治理则按照系统工程管理和运行的思想方法实施，与航空产品设计制造数据、运行维护数据形成完整的航空产品数据。图 6 给出了试飞系统工程与试飞大数据治理业务流程。

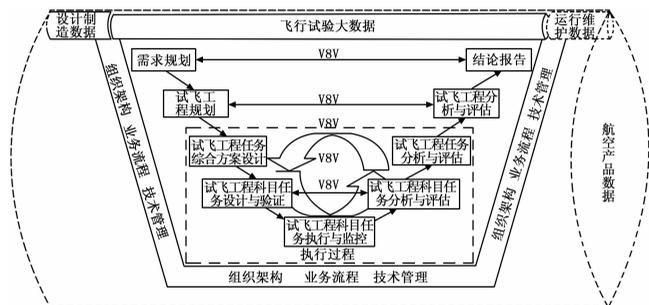


图 6 试飞系统工程与飞行试验大数据治理

在航空产品研制初期到产品样机制造完成，逐步掌握、分解、深化产品功能、性能及其综合指标需求，完善形成工程规划，以此为基础进行工程任务综合方案设计。依照综合方案，遵循产品试验工程技术原理，逐步进行科目综合试飞设计、仿真验证、执行与监控、分析与评估，渐进扩展和不断迭代循环、综合评估，最终验证和确认达到所有的试验需求，形成结论报告。在试飞业务执行与管理过程中，数据信息是试飞系统工程运行过程中的核心，以完善的组织执行架构，严谨的业务管理流程，结合先进的技术与管理平台，实现试飞工程数据治理的目标。飞行试验大数据质量监控平台如图 7 所示。

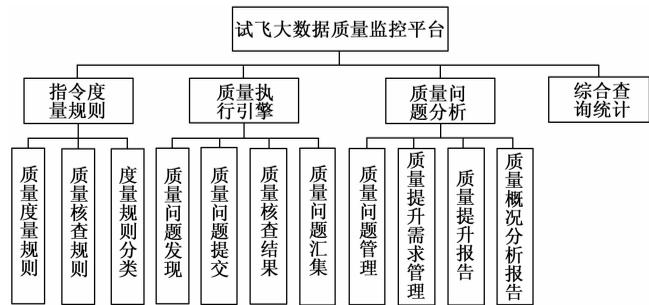


图 7 飞行试验大数据质量监控平台

型为基础，建立数据质量度量规则及其执行引擎，对数据对象进行质量检测与分析评估。

①数据质量度量规则能够根据试飞数据质量特性分类定义相应的度量规则，数据质量分析提供依据。试飞数据质量度量规则根据试飞数据管理和技术要求进行定义。

②质量执行引擎执行质量度量规则进行数据质量检查和分析，汇总并形成质量问题报告，为质量提升提供基础。

③质量问题分析根据质量提升需求，根据质量核查中发现的问题进行分析，形成数据治理分析报告和质量提升报告，为数据质量改进提供基础。

④综合查询统计可以实现对质量监控平台下的所有信息进行综合查询和分析。

3.3 涵盖试飞工程全过程的一体化试飞大数据管理与应用平台

基于单一数据源的一体化试飞大数据管理与应用平台是试飞大数据治理的技术保障，围绕试验对象数据，涵盖试飞工程需求、规划、设计、执行、评估、总结等全过程，建立包括飞行试验工程管理数据与专业数据的一体化试验大数据结构模型，充分利用先进的大数据技术，针对飞行试验大数据结构特点及其应用需求，建立基于统一数据模型的试验大数据全生命周期高效全面管理、智慧应用与维护体系和应用平台，为航空产品设计、制造、试飞、运维和航空科学技术发展研究等所有用户，提供可追溯、可查询、可分析的完整、准确、可靠以及多粒度的试验大数据高效服务。如图 8 所示。



图 8 试飞工程一体化数据管理平台

①试飞工程一体化试验数据结构模型 (FTDB) 是试飞工程大数据治理的数据基础，包括试飞工程管理数据结构模型，以试验对象数据为中心的试验规划、设计、测试改装、工程结果、评估结论等数据结构模型，建立和维护了完整、准确、可靠的试飞工程数据关联关系，能够实现试验数据信息的追溯和延展，真实地反映试飞工程管理和实施的数据态势。

②面向应用的服务中间件 (SOA) 是面向试飞工程应用，以试验数据结构模型为基础，实现试飞工程管理、规划与设计、分析与评估等应用功能，按照标准化的中间件接口定义规范，可以实现功能组件扩充，满足航空设计、制造、维护以及科研应用需要。