

ASME STB-1-2020

Guideline on Big Data/Digital
Transformation Workflows and
Applications for the
Oil and Gas Industry



STB-1-2020

GUIDELINE ON BIG DATA/DIGITAL TRANSFORMATION WORKFLOWS AND APPLICATIONS FOR THE OIL AND GAS INDUSTRY

Prepared by:

Barbara Thompson, P.E.
TeamBS, LLC.



Date of Issuance: December 31, 2020

This publication was prepared by ASME Standards Technology, LLC (“ASME ST-LLC”) and sponsored by The American Society of Mechanical Engineers (“ASME”), Petroleum Division.

Neither ASME, ASME ST-LLC, the author, nor others involved in the preparation or review of this document, nor any of their respective employees, members or persons acting on their behalf, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe upon privately owned rights.

Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer or otherwise does not necessarily constitute or imply its endorsement, recommendation or favoring by ASME or others involved in the preparation or review of this document, or any agency thereof. The views and opinions of the authors, contributors and reviewers of the document expressed herein do not necessarily reflect those of ASME or others involved in the preparation or review of this document, or any agency thereof.

ASME does not “approve,” “rate”, or “endorse” any item, construction, proprietary device or activity.

ASME does not take any position with respect to the validity of any patent rights asserted in connection with any items mentioned in this document, and does not undertake to insure anyone utilizing a standard against liability for infringement of any applicable letters patent, nor assume any such liability. Users of a code or standard are expressly advised that determination of the validity of any such patent rights, and the risk of infringement of such rights, is entirely their own responsibility.

Participation by federal agency representative(s) or person(s) affiliated with industry is not to be interpreted as government or industry endorsement of this code or standard.

ASME is the registered trademark of The American Society of Mechanical Engineers.

No part of this document may be reproduced in any form,
in an electronic retrieval system or otherwise,
without the prior written permission of the publisher.

The American Society of Mechanical Engineers
Two Park Avenue, New York, NY 10016-5990
ISBN No. 978-0-7918-7399-1

Copyright © 2020
THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS
All Rights Reserved

TABLE OF CONTENTS

Foreword	vii
1 Purpose, Definitions and References	1
1.1 Scope	1
1.1.1 How to Use This Guideline	1
1.2 Definitions	2
1.3 References	10
2 Data Structure and Management.....	11
2.1 Introduction	11
2.2 Structured Data.....	11
2.2.1 Types and Usage	11
2.2.2 Databases.....	11
2.2.3 Examples	11
2.3 Unstructured Data.....	13
2.3.1 Types and Usage	13
2.3.2 Respective Databases	13
2.3.3 Examples	13
2.4 Security and Governance of Data.....	15
2.4.1 Responsibility of the Enterprise	15
2.4.2 Key Concepts of Information Security.....	15
2.4.3 Data Protection.....	15
2.4.4 Developing Software that is Secure	16
2.4.5 Facility Management Systems.....	16
3 Big Data in the Oil and Gas Industry.....	17
3.1 Introduction	17
3.1.1 Overview	17
3.1.2 Oil and Gas Facility Lifecycle Digital Requirements	17
3.1.3 Designing the Digital Facility	19
3.1.4 Understanding the Data in Oil and Gas Activities	19
3.2 Hydrocarbon Reservoirs, Drilling, Production, Transportation and Refining.....	19
3.2.1 Activities that Produce Data.....	19
3.2.2 Digital Facility Descriptions	25
3.3 Mechanical Equipment and Instrumentation.....	26
3.3.1 Pressure Control Equipment.....	26
3.3.2 Rotating Equipment.....	26
3.3.3 Electrical and Instrumentation.....	26
3.3.4 Process Control	27
3.3.5 Process Equipment	27
3.3.6 Civil/Structural	28
3.4 Pipelines/Storage.....	28
3.5 Operations	28
3.6 MetOcean	28
3.7 Health and Safety	29
3.8 Supply Chain	29
3.9 Special Note to this Chapter.....	30

4	Methods of Analysis	31
4.1	General Information on How and When to Use These Methods	31
4.2	Descriptive Analytics and Data Mining	33
4.2.1	Importance and Objectives	33
4.2.2	General Statistical Descriptors	34
4.2.3	Descriptive Analytical Tools.....	34
4.3	Predictive Analytics	35
4.3.1	Importance and Objectives	35
4.3.2	Regression Problems and Solutions	36
4.3.3	Classification Problems and Solutions	36
4.3.4	Unstructured Data Problems and Solutions.....	38
4.3.5	Time Series.....	39
4.4	Prescriptive Analytics.....	39
4.4.1	Importance and Objectives	39
4.4.2	Optimization Problems and Solutions	39
4.4.3	Simulation Problems and Solutions	39
4.5	Application Program Interfaces.....	39
4.5.1	Importance and Objectives	39
4.5.2	Implementation.....	40
4.6	Visualization Tools.....	40
5	Data Analytics Project Workflows	41
5.1	Introduction	41
5.1.1	CRISP-DM	41
5.1.2	INFORMS and the Job Task Analysis	42
5.1.3	Structure, Roles and Responsibilities	42
5.1.4	Value to the Enterprise	42
5.2	Business Problem Framing.....	43
5.2.1	Description	43
5.2.2	Team Member Roles	44
5.2.3	Example Business Challenge – Permian Basin Production Forecasting.....	44
5.3	Analytics Problem Framing.....	45
5.3.1	Description	45
5.3.2	Team Member Roles	45
5.3.3	Example Business Challenge - Permian Basin Forecasting Model Continued	46
5.4	Data	46
5.4.1	Description	46
5.4.2	Team Member Roles	47
5.4.3	Example Business Challenge - Permian Basin Forecasting Model Continued	48
5.5	Methodology Approach and Selection	49
5.5.1	Description	49
5.5.2	Team Member Roles	50
5.5.3	Example Business Challenge - Permian Basin Forecasting Model Continued	50
5.6	Model Building and Testing.....	50
5.6.1	Description	50
5.6.2	Team Member Roles	51
5.6.3	Example Business Challenge - Permian Basin Forecasting Model Continued	51
5.7	Solution Deployment.....	52
5.7.1	Description	52

5.7.2	Team Member Roles	53
5.7.3	Permian Basin Forecasting Model Continued.....	53
5.8	Model Maintenance and Recycle	53
5.8.1	Description	53
5.8.2	Team Member Roles	54
5.8.3	Example Business Challenge - Permian Basin Forecasting Model Concluded	54
5.9	The Business Solution.....	55
5.9.1	The Continuing Challenge.....	55
5.9.2	The Important Role of the Engineer.....	55
	Mandatory Appendix I: Data Characterization Chart for Oil and Gas	56
	I-1 Digital Twin Representation Example	57
	Mandatory Appendix II.....	59
	II-1 Detailed Data Journey.....	60
	II-2 SIPOC Chart	61
	II-3 Job Function Descriptions.....	62
	II-4 RACI Chart	63
	Nonmandatory Appendix A: Case Study	64
	Nonmandatory Appendix B: Certifications Available	78
	Nonmandatory Appendix C: Glossary Definitions	80
	Copyright Declarations.....	94

LIST OF TABLES

Table 2-1: Data Description Table	12
Table 4-1: 5S Lean Approach to Data Mining	33
Table 5-1: RACI Chart for Business Framing.....	44
Table 5-2: Analytics Problem Framing RACI Chart.....	46
Table 5-3: Data RACI Chart.....	48
Table 5-4: Methodology Selection RACI Chart.....	50
Table 5-5: Model Building and Testing RACI Chart.....	51
Table 5-6: Solution Deployment RACI Chart.....	53
Table 5-7: Model Lifecycle RACI Chart.....	54

LIST OF FIGURES

Figure 2-1: Relational Database Example.....	12
Figure 2-2: Object-Oriented Database Example	12
Figure 2-3: Data Lake Example	14
Figure 2-4: NoSQL Data Types	14
Figure 2-5: Graph Database.....	14
Figure 3-1: Digital Facility Components.....	18
Figure 3-2: Onshore and Offshore Seismic Imaging Activities	20
Figure 3-3: Drilling Derrick and Casing Components	21
Figure 3-4: Example Well Log.....	22
Figure 3-5: Completed Well Example.....	23
Figure 3-6: Oil and Gas Production Facility	24
Figure 3-7: Gulf of Mexico Pipeline System.....	24
Figure 3-8: Oil Refining and LNG Processing	25
Figure 3-9: Process Control Graphical User Interface	27
Figure 3-10: Supply Chain Analytics Landscape.....	30

Figure 4-1: Data Analytics Journey Overview	32
Figure 4-2: Descriptive Analytic Tool Examples.....	35
Figure 4-3: Decision Tree.....	37
Figure 4-4: Neural Network	38
Figure 4-5: Data Journey from Source to API.....	40
Figure 5-1: CRISP-DM Business Process.....	41
Figure 5-2: Data Selection Decision Tree	47
Figure 5-3: Descriptive Analytics for Permian Basin Data Sets	49
Figure 5-4: Model Comparison Results	52

FOREWORD

Guideline Description

This guideline is the culmination of the efforts of ASME industry professionals in oil and gas to define Big Data and its useful applications to upstream, midstream and downstream businesses.

The concept of Big Data intimidates decision-makers and business leadership. While the introduction of new digital technologies is the cornerstone of the industry's digital transformation, use of Big Data requires the sharing of asset-specific or operational data of a size and scope that is difficult to grasp. Industry data may also be of poor quality, requiring significant effort and resources to cleanse and aggregate prior to its analysis. In addition, the gathering, aggregation, analysis, and data storage and maintenance is very expensive. The industry needs a standardized and efficient end-to-end workflow to validate the quality of the data, and subsequently, the accuracy of the results from the analysis.

The goal of this document is to alleviate that intimidation by providing a framework for understanding and a workflow for utilizing data analytical techniques to solve business problems; without requiring the reader to be a full-time statistician or data scientist professional.

Who Should Use this Guideline?

The design of this Guideline is intended to be universal in its application to Big Data challenges in the oil and gas industry. It is written not only for oil and gas professionals who are beginners to Big Data techniques but also for data professionals looking to contribute to unique oil and gas applications.

Specific users can be characterized as:

Citizen Data Scientist: a subject matter expert in engineering, operations, supply chain, planning, project management or operations that requires data insights.

Early Career Engineer: a young professional that is looking to improve his or her career by adding a data dimension to their problem solving.

Data Scientist Professional: a data science professional that is looking to apply his or her deep experience in data analytics by learning the unique sets of data and operational challenges of the oil and gas industry.

The author acknowledges, with deep appreciation, the activities of the ASME volunteers and staff who have provided valuable technical input, advice and assistance with review of, commenting on, and editing of, this document, particularly the activities of the Peer Review Group (PRG) consisting of Michael Wells, Thalia Kruger, Amit Kumar, Mete Mutlu, Brian Webster, Kathryn Hyam, and the activities of the ASME Petroleum Division Leaders and Petroleum Division Big Data Task Group consisting of Jim Kaculi, Ed Marotta, John O'Brien, and Jamie Hart.

Established in 1880, ASME is a professional not-for-profit organization with more than 100,000 members and volunteers promoting the art, science and practice of mechanical and multidisciplinary engineering and allied sciences. ASME develops codes and standards that enhance public safety, and provides lifelong learning and technical exchange opportunities benefiting the engineering and technology community. Visit <https://www.asme.org/> for more information.

ASME ST-LLC is a not-for-profit Limited Liability Company, with ASME as the sole member, formed in 2004 to carry out work related to new and developing technology. ASME ST-LLC's mission includes meeting the needs of industry and government by providing new standards-related products and services, which advance the application of emerging and newly commercialized science and technology, and providing the research and technology development needed to establish and maintain the technical relevance of codes and standards. Visit <http://asmestllc.org/> for more information.

1 PURPOSE, DEFINITIONS AND REFERENCES

1.1 Scope

This guideline explains the current use and application of data analytics and data science in the oil and gas industry. It is designed to provide guidance on how to utilize data analytics and machine learning/artificial intelligence (ML/AI) to address a given business need, resulting in value-creation. Within the guidelines will be descriptions of various data analytics techniques and the recommended tools for the respective techniques.

1.1.1 How to Use This Guideline

(a) Basics

This document is designed as both a “how-to” and a reference document. The chapters are arranged in a building block sequence that parallels the journey of a business data project. These building blocks should help provide a roadmap to data-driven projects.

Each chapter is also a standalone reference for its respective topic. The reader can reference these individual chapters as needed to fill gaps in his or her understanding of either data techniques or oil and gas.

(b) Chapters

In this first chapter, the user is introduced to definitions and acronyms common to both oil and gas and to the data analytics strategies highlighted in this guide.

Chapter 2 provides background to the various types of data in the oil and gas industry, including how they are curated, described, used, and safeguarded.

Chapter 3 describes the Digital Facility and the related oil and gas activities and operations that produce data. It also defines the types of data produced by each operation and potential applications for data-driven insights.

Chapter 4 explores the types of data analytics tools available to a data project leader and how they can be utilized. This chapter is where the user can find information about topics such as regression models, classification, machine learning, optimization, and data visualization.

Chapter 5 is a description of a full data project. The user will learn how to structure a project, assemble the appropriate team resources, frame the questions to be answered, model and execute the project, and usefully deploy the resulting model.

Two sets of Appendices are included for references and details that supplement the discussions of the chapters. The Mandatory Appendices are presented to help understand data projects and the team members required to execute them. Nonmandatory Appendices are included to provide the user with additional definitions, sample case studies as examples, and a description of relevant certifications available if the reader wishes to increase his or her knowledge and proficiency in data analytics.