Vincent August Harahap

# Recommendations to Improve The Quality and The Automation of The OREDA Data Process

June 2019

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Mechanical and Industrial Engineering

Master's thesis

Master's thesis
2019
Vincent August Harahap

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

Norwegian University of Science and Technology

# Recommendations to Improve The Quality and The Automation of The OREDA Data Process

**Vincent August Harahap**

Reliability, Availability, Maintainability and Safety (RAMS)
Submission date: June 2019
Supervisor: Mary Ann Lundteigen
Co-supervisor: Siegfried Eisinger

Norwegian University of Science and Technology
Department of Mechanical and Industrial Engineering

**Abstract**

The OREDA database has been widely used to support the data-driven decision-making in the reliability analysis. Providing a good quality of database with the agility to supply recent data is crucial. To achieve such condition, there is a need for continuous improvement in the database quality and more automated process on the entire data process. Data quality analysis was performed by investigation on the timeliness and the value-added attributes of the reliability data quality framework and modification on the current quality assurance protocol. The attributes analysis gave the insight to determine the vital and the non-vital reliability data. Besides, unequal data quality from the data collectors of the company members has been indicated. Therefore, extra attention on the data collectors motivation is necessary. The automation process analysis was performed by investigation on the current quality assurance checklists and experience gained during the involvement with the phase 12 database. The quality assurance checklists analysis was comprised of the checklists instructions clarity, the timing to execute the checklists items, and the existence of the non-informative and the blank fields. The experience gained brought an understanding of the potential fields and checklists items to be automated with a calculation or further interpretation. List of potential fields and checklists to be automated along with their automation process description have been developed.

**Keywords:** OREDA, reliability database, timeliness, value-added, data automation process, quality assurance protocol, quality assurance checklist

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Background

OREDA (Onshore Offshore Reliability Data) is a project organization consists of several oil and gas companies who have continuously collect comprehensive reliability data for safety equipment (OREDA, 2015). These data are stored into a database and accessible by the members, enable them to share and exchange data. The OREDA database is commonly used worldwide as decision support throughout the whole life cycle of the system, from the design phase until operation. Currently, OREDA is about to release a new database, so-called the phase 12 database.

Data collection is taking a huge role within the whole process to establish proper and sufficient database. Collecting a large amount of vital reliability data is a necessity for data-driven decision-making. It can be questioned what kind of data elements are typically required for reliability analysis. The database should be able to provide all the necessary data, but it should not be overwhelmed by unnecessary and expensive data. The study of optimization in the current database is required. The study should improve the quality of data itself and reduce unnecessary work in data collection. Numerous studies have attempted to improve the quality of a reliability database. Cressent et al. (2013) studied the database for complex safety-critical systems development. They prepared a framework to show interactions between reliability expert and data collector. The framework forms a dysfunctional behavior database which allows companies to build and store their critical knowledge database. Duarte et al. (2013) proposed an architecture involving various variables that generate feedback information to the maintenance database. In the power industry section, Chen et al. (2017) carried out several investigations resulted in three main data quality issues: noise, incomplete and outlier data.

There is a broad recognition from the industries related to the importance of reliable data collection for optimizing the cost (Hahn et al., 2017). The issue of the reliability database has received considerable critical attention. There is a strong demand for making better use of operational experience to improve operation and maintenance by the high quality of reliability data. Excellent quality of data collection creates a countable database which brings cost benefit to the industries.

In phase 12, a new format with a better interface is being developed by OREDA. As the phase 12 data coming in, it needs to be combined with the old format. Such transformation of the database requires thorough manual work. It takes time to extract incomplete data, non-linked data and not readily readable data which not comply with the new standardized format (e.g., data sheets stored in pdf format). Manual work also means that the process is not sustainable. Haegemans et al. (2019) identified that manually acquired data has a significant problem; it is prone to errors. Even a single error in data collection may lead to immense impact which possibly goes to wrong decision making (Kozak et al., 2015). The extra effort requires to ensure data quality. Therefore, there is a considerable demand to improve the way the data is collected from manual to as automatic as possible.

In the data quality, OREDA has data collection assurance procedure and checklists to ensure that the collected data have reached the acceptable quality level. The quality assurances are performed during the spot check of received data and when the data being merged before the data is to be released. OREDA has an old quality assurance procedure with the last update in 2007. Therefore, OREDA requires to look into their quality assurance procedure in corresponding with the new database format.

Efficiency in data collection and quality analysis are therefore the theme of this master thesis in collaboration with SUBPRO, Equinor and DNV GL. The collaboration has been done through the conversion of the OREDA phase 12 database and supervisions.

## 1.2 Objectives

The main objective of this master thesis is to give recommendations to improve utilization of reliability data on the transition of phase 12 data to the new OREDA database. The recommendations include identification of data which are critical for reliability studies and investigation of data which can be automated, reducing the manual work and improve the efficiency (time, resources). Currently, OREDA has its own protocol and quality assurance procedures. Some improvement on those procedures are necessary to increase the quality of OREDA database. Some of the tasks have been covered during author's specialization projects. Remaining break down tasks to fulfill the objective have been listed as follow:

1. Establish an overview of data elements which are vital for reliability study and which are not.
   (a) Indicate the criteria for selecting necessary obtained data based on reliability data attributes.
   (b) Evaluate current procedure on change format activity.
   (c) Study the possibility for automation on data process.
2. Improve the quality check procedure of the received data based on reliability study.
   (a) Investigate and update the existing overall quality assurance procedure.

(b) Evaluate and verify current quality check procedure on the newly arrived data.

(c) Analyze the possibility for an automatic quality check process, knowing that the data collector should get immediate feedback when data are not enough or many empty fields are discovered.

3. Improve the quality check procedure of the merged database before the OREDA database is to be released.

   (a) Evaluate and verify current quality check procedure for the merged database.

   (b) Investigate the necessity to move checked item to the received data quality check, bringing the opportunity to detect any errors on the earlier stage.

   (c) Analyze the possibility for an automatic quality check process.

4. Investigate the possibility for an overall automatic data process.

   (a) Identify the problem area of data collection.

   (b) Identify the difficulties in transforming the raw database into the OREDA database.

   (c) Understand how the data collection process has been done manually.

   (d) Develop descriptions of automatic procedure on the data process.

5. Implement and verify the automatic classification algorithms on the new (phase 12) OREDA database.

   (a) Analyze the adequacy of automatic classification algorithms on several components of the new OREDA database.

   (b) Verify and update the automatic classification algorithms.

## 1.3 Scope & Limitations

The purpose of this master thesis is to do data analysis on the whole data processing process from the raw databases until the complete database. The process from the raw database into the complete OREDA database pass several transition steps. Raw databases come from several company members have to pass several quality checks and format change before they are into OREDA new database (phase 12). A schematic drawing is representing the scope of the master thesis illustrated in Figure 1.1.

Each company member collects desirable data and does its quality check before delivering the database to the data custodian. Data custodian has to do the quality check on the received database. At this point, there might be a case that data custodian returns the database to the data collector. This condition happens if the database does not reach the minimum quality level. Data custodian has to change the format of all checked database into the OREDA compatible format. The process continues with merging all database into one database. Before doing the

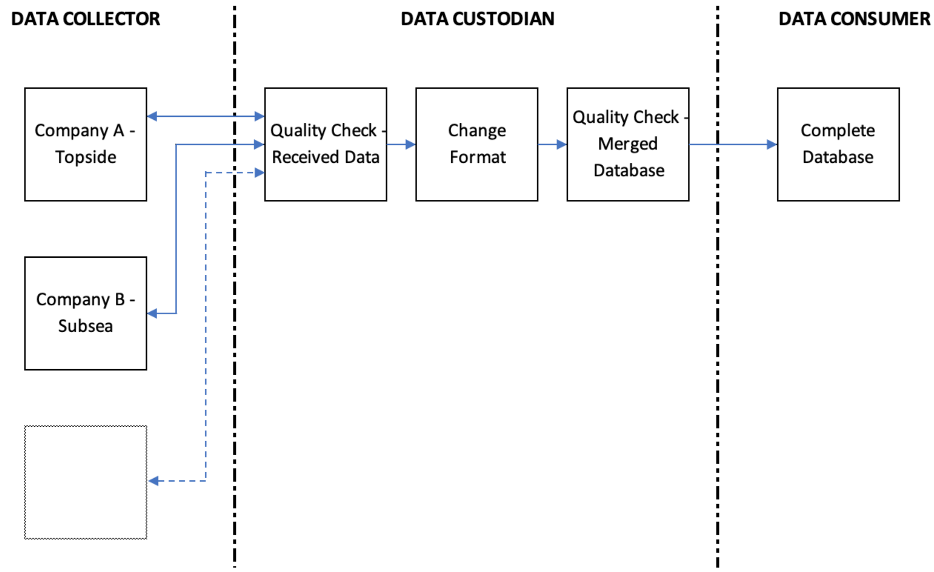



Figure 1.1: Scope of the master thesis

merging process, the data custodian has to do the quality check on the merged database. Once the process is done, the complete database is ready to be delivered to the data consumer. The data consumer may use the complete database for further reliability analysis.

In this master thesis, databases from the topside and the subsea applications are available from several companies. The scope of data custodian then focuses on both topside and subsea equipment database. On the other hand, the scope of data consumer focuses on subsea equipment only. The scope of analysis focuses on subsea application, due to the increasing number of subsea installation (Uyiomendo and Tore, 2015), the high cost of subsea intervention (Fanailoo et al., 2008) and the increasing novelty of subsea technology (Brissaud et al., 2010). Implementations of the automatic classification algorithms are limited to equipment in the subsea application.

Several pieces of software are used throughout the master thesis. They are consists of in-house software which is used to present the OREDA database and software used by the data collector. The OREDA database presented in OREDA@cloud application, while the data collector's software is BiCycle and Aveva. It is the data custodian task to change the format of the data collector's format into OREDA@cloud format. Even though the format changed from BiCycle and Aveva, the analysis and recommendations for automation process are performed for BiCycle software only. Aveva is considered as an old software which no longer used on data collection process in the future. Besides, Dbeaver software and Ms. Excel are used as an interface of that software. Knowledge of SQL database is necessary to be able to utilize the software.

The collaboration with DNV GL includes the task of changing the format and the quality check of the merged database. Once the merged database passes the quality check, the database is ready to publish. This master thesis relies on getting access to the published database for implementation and verification of automatic classification algorithms. Unfortunately, the author does not get a chance for accessing the database due to delay in the database publication.

## 1.4 Approaches

In general, the approaches conducted for this master thesis are by literature review and hands-on experience in the phase 12 OREDA database. The literature review has mainly been obtained through international standards, OREDA documents, and journal and conference papers. The primary international standard used on this master thesis is ISO14224. OREDA documents consist of OREDA guidelines and quality assurance procedures and checklists. Journal and conference papers are collected from a database such as oria.no, science direct, Elsevier, scholar.google.com., onepetro.org & engineering village. The journal and conference paper search began with using the terms "OREDA" and "reliability database." The following terms were also used in addition to narrow down the search results to get relevant results like "automatic data collection," "reliability database quality," "operational condition" and "quality assurance improvement." Hands-on experience in the phase 12 OREDA database includes the changing format activity and the quality check of the merged database. The author gained experience during the hands-on experience in the real companies databases. These experiences are based for analyzing current OREDA data process and investigating the possibility for automation.

Approaches in details are as follow :

1. Literature review

    - Review of ISO 14224. This review is necessary to get a broad understanding of the requirement, and taxonomy applies in the database. This review includes data collection rules and taxonomies for topside and subsea equipment.
    - Analyze current OREDA's guidelines and quality assurance procedures and checklists documents. The guidelines are the basis for a consistent approach of data collection, while the quality assurance procedures and checklist ensure that the database meets a certain quality level. Analyze those documents is essential to understand the concept, definitions, and rules of current reliability data collected by OREDA.
    - Studies of relevant journal and conference papers. These studies are from several fields such as reliability engineering and system safety, electronic and information engineering, offshore technology, information system, process industries, arctic engineering, etc. These studies are important to get information on the latest and recent studies.
    - Studies on the improvement of database quality from various industries. These industries consist of the transport system, the military system, renewable and sustainable energy, health care, and biomedicine. These studies bring insight into how to improve the database quality from different applications. Some of the improvements are implemented in this master thesis.

- Review of author's specialization project. The project is a preface of this master thesis, as the master thesis is the continuation of the specialization project. Parts of the specialization project are implemented again in the master thesis — notably, the implementation of the automatic classification algorithm on the new OREDA database.

2. Hands-on experience on the OREDA phase 12 database.

    - Review of topside and subsea reliability data set structure from several companies. Study on current installation, inventory, failure and maintenance database.
    - Compare and analyze company's database format and OREDA's database format. Identify the involved fields.
    - Convert the database format into OREDA@cloud format by fields mapping. Identify the equivalent reliability data fields between multiple databases.
    - Do and experience the quality check of the merged database and simulate the quality check of the received data.
    - Propose recommendations on both quality check of the received data and the merged database based on gained experience.
    - Evaluate the current quality assurance protocol and give suggestions for improvement.
    - Analyze the opportunity for automatic data process. Identify which fields can be classified automatically based on the consistency and distinct reliability data attributes.
    - Implement and demonstrate the automatic classification algorithms on the complete database. Update the algorithms for adequacy, if necessary.

The conclusions are formulated based on the study of international standard, OREDA guidelines and quality assurance procedures, author's specialization project report, journal and conference paper gathered throughout the master thesis. In addition, practical experience gained when the author involved on phase 12 database is beneficial for the analysis. The conclusions are presented to meet the aim of this master thesis. Furthermore, discussions on how to improve the reliability data collection process, opportunity for automated data collection process and additional reliability data to be collected are presented.

Throughout the master thesis, supervision has been conducted. Weekly meetings have been organized with the primary supervisor to gather reliable sources, to discuss doubts and evaluation. Periodically meetings have been conducted with second supervisor (DNV GL representative) to get a thorough explanation of current quality data assurance, explanation on the current data collection and analysis, and their expectation with this master thesis. The guidance was deliberate over the hands-on experience with the database. Joint meetings have been conducted at the beginning and the end of the project. The supervisions have been arranged to make sure that the project was on track.

## 1.5 Structure of the Report

A structure of the master thesis has been made according to the proposed task.The rest of the chapters structure and their contents are as follow:

1. Chapter 2 gives an overview of the reliability database. The chapter highlights the concept of relational reliability database and the quality of the reliability database in general. The way of the database divided, various methods of data collection, validation of data and quality assurance and assessment are presented. At last, the OREDA database structure is summarized.

2. Chapter 3 briefly describes the OREDA quality assurance and framework. This chapter presents the role of involving actors in OREDA data processes, discusses a suggestion for an improvement in the current quality assurance protocol, and points out the framework to be used for quality assessment of OREDA database. The analysis is performed with the focus on the timeliness and added-value attributes.

3. Chapter 4 explains the experience gained during involvement in the phase 12 database process. This chapter discusses some improvements during change the OREDA format activity and analyses on the current quality check procedures.

4. Chapter 5 presents several issues and challenges for further data collection. This chapter discusses the recent studies and the implications for OREDA.

5. Chapter 6 includes the summary and the conclusion, the discussion on how this study answers the thesis's objectives and suggestions for further research.

# Chapter 2

# Reliability Database Overview

This chapter describes and discusses the theory of reliability database. This chapter has been divided into three parts. The first part deals with the explanation of the reliability database as a relational database followed by the actor's role involved in the data process. The second part discusses the quality of reliability database. This part presents how the database is organized, what are the data collection methods, how the database is validated, and how the quality assurance and assessment are performed. The third part presents the OREDA database structure for the topside and the subsea application.

## 2.1 Relational Reliability Database

A database is a collection of data organized with a view to its utilization by programs corresponding to distinct applications and methods that facilitate the independent development of data (DNV, 1999). In reliability applications, a database is an organized filing system and continuously updated. It contains data that describe the development of the component's behavior of an installation as a function of time.

Engineers and manufacturer from various sectors have become more aware of the advantages of derived knowledge of their installations. Tapia et al. (2011) who put fusion material as their central focus of a study, found the importance of traceability in the reliability database. Linsday et al. (2008) pointed out the need for a reliability database in wind turbine technology as the market continued to expand and gained a significant share. Cressent et al. (2013) highlighted the need for a reliability database which described an interaction between components failure mode and dysfunctional behavior. Neto et al. (2018) provided valuable insight on how reliability database could be used as a decision-making tool, as well as formulated the design specification. Engineers start to collect data (such as components failures, maintenance data, and duration, etc.) to enable them to search such information on the historical data system. All data stored in a database.

A reliability database is constructed in a relational database format. A relational database is fast becoming one of the most widely used databases to store and manage data in organizations

(Nassiri et al., 2018). In a relational database, data are stored in tables which contain numbers of columns and rows. Each row represents a record, while each column represents a field.

The main benefit of using a relational database is the ability to retrieve data upon matching values of a shared field between a pair of tables which share relationship (Hernandez, 2013). These type of relationships facilitate users to store a massive amount of data and enable them to organize and to find the data more straightforwardly. No access path is defined beforehand, meaning that all manipulation of data in tabular form is possible. This ability achieved due to the construction of a relational database which requires a primary key. The primary key is a column that is used as a unique identification of rows in a table. This primary key act as a reference column. The primary key creates a link between tables, bridging the information, indicating a sort of relationship.

In a relational database, the rows in a table have no specific order, allowing users to organize the data in a different manner. User may set up rules that ensure the data remains consistent when adding, updating, and deleting data. Despite all the advantage sides, relation database has a constraint in updating the table. The database has to check whether the new data satisfies the relevant integrity each time a table is updated.

A relational database uses SQL as the standard language to access the data. SQL is a powerful query language with an easy programming interface. SQL consists of statements to insert, update, delete, and query data, used for managing and manipulating data. SQL main roles are for creating a database, querying the database to obtain the necessary information to answer the questions and controlling the security of the database (Wilton and Colby, 2005).

There are several actors role in the data processes of a reliability database. Strong et al. (1997) found that knowledge about data processes encompasses knowledge about the three critical main processes within a data production process. They are a collection of data, storage and maintenance of data, and user retrieval and manipulation of data. Three roles involve within the data production process are data collectors, data custodians, and data consumers (Lee and Strong, 2003). Data collectors are people who collect the information data from the sources. Data custodians are people who manage computing resources for storage and processing data. Data consumers are people who use the data.

The purpose of data production processes is to produce high-quality data for the data consumers (Wang and Strong, 1996). There are several definitions of data quality. Chen et al. (2017) pointed out that the data quality related to less noise, outliers, and incomplete of raw data. Saha and Srivastava (2014) mentioned in their study, that data quality referred to a recognition of the outlier data and elimination of the error data. Alizamini et al. (2010) described that data quality is a refinement process of a complex non-structural concept. Previous studies have demonstrated that there was no particular consensus on the definitions of the data quality. Researchers create their definition of data quality based on each application of study.

## 2.2 Reliability Database Quality

It is necessary to understand the purpose of the data and their quality requirement before creating a database. Baker (2000) and Walter et al. (2008) tried to formulate several objectives of reliability database:

- To provide generic data on the failure rates and repair rates of equipment.
- To improve equipment availability via modelling.
- To identify failure modes, failure causes and improve their intrinsic safety.
- To make the most appropriate technical choices for decision making.
- To define good maintenance strategy.

Understanding the purpose of the database becomes the base to define the organization of data to be collected. Having defined the data to be collected, the method for collecting and recording the data needs to be considered. After data collection, data are being processed and validated before entering it in the database.

### 2.2.1 Organization of Data

The organization of data represents the division on how the database being collected. This division is made based on the origin and the characteristic of data. Each database must consist of primary key fields as a reference to share the information. The collected data must have the intention to achieve the requirements and objectives. The database must contain the maximum possible information concerning the impact on reliability analysis, the difficulty to obtain the data, and the cost-effective analysis. The OREDA database consists of three related parts (SINTEF, 2015):

1. The Inventory part. This part is related to component identification where the information must allow the user to know all useful knowledge of the component. It compromises technical parameters of the component, design characteristic, operational data (e.g., operation duration, operation mode), environmental and performance data (e.g., internal and external environmental condition) and installation information (e.g., location).
2. The Failure part. Information presented in this part must describe the failure of the component. The description of failure usually is extracted from reports of events. It describes all failure events experienced during the period of surveillance. These data are related to the degree of failure (critical, degraded or incipient), failure mode, failure cause, etc.
3. The Maintenance part. Information concerning the maintenance policy, maintenance dates, duration of maintenance, actual maintenance being carried out.

### 2.2.2 Method of Data Collection

The choice of methods is a function of the data. The data can be either from an old equipment, under the real condition of operation, at the design stage or gathered from proven-known equipment. The methods are related to the characteristic of the data itself.

If the data is related to an old equipment, then the information most probably has come from published data handbook (Zio, 2009). The use of such reliability data source often appears to be difficult due to the lack of component parameters. Some missing information could be the central issue of using this method. The most common reasons are details of equipment that are not necessarily noted and sometimes it is hard to find the source of document with a specific area of interest.

If a good understanding of the performance under the real conditions of operation is obtained, then a database could be created by collecting operational data (DNV-GL, 2016a). It consists of observing their behavior under the real working conditions and recording all incident data. The advantage of this method is that it is specific to the equipment in operation. This method can be applied to a diverse range of equipment but should be limited to essential equipment only considering the economic constraints. The field data collection can be done in two different ways:

- Manual collecting process, which use the collection form as its main instrument. The form is a descriptive collection form that enables equipment to be identified. It may consist of failure collection form and operational collection form which provides operating details of the system. This collection is a fully manual operation. Therefore, this method is prone to high potential error.

- Data acquisition system, which consist of a transducer to collect raw data, a computer program for the processing of data and a storage in the database. Some of the advantages of the data acquisition system are the freedom from human error, the possibility of the collection at fixed intervals, automatic updating and less prone to data loss.

Data collection under real conditions of operation with these methods is the most common way to collect reliability data from industries. Current reliability databases are facing the challenge to leave the manual collecting process to the fully automated system.

If the equipment is at the design stage, the reliability test could be carried out to determine its reliability (Chen et al., 2019). As the equipment is at the design phase, the purpose is to derive parameters necessary for reliability studies and assessment. Two crucial test criterions to determine when the test should end are the duration and the number of failures. One of the disadvantages of the reliability test is that they seldom allow real operating conditions to be accurately reproduced.

If the equipment is known and only a relatively low degree of precision is required of the data, then evaluations can be made based on assessments by expert (Liu et al., 2018). This method is from the concept that experts confronted with the experience of systems and therefore, should be in a position to deduce some parameters of functional safety. One of the disadvantages lies in the fact that human memory is selective and eventually has more frequent failures. Another researcher, Barabadi (2014), proposed a model to forecast the reliability analysis in the Arctic area because the current reliability database is not covering the arctic environment condition. They used the existing reliability data as based to forecast reliability in the high-stress condition with accelerated life testing.

### 2.2.3 Process and Validation of Data

After data collection, mathematical processing is necessary to bring out those reliability parameters such as failure rates, repair rates and other reliability parameters. Furthermore, different mathematical methods (classical statistic, Bayesian statistic, etc) need to be performed for the determination of these parameters. Very often, more than one of the numerous types of statistic distribution apply to the collected data.

Numerous difficulties arise during data collection. Therefore, it is necessary to check data quality. The term quality implies a process that determines the accuracy and the relevance of the data collection for a clearly defined objective obtains validated high-quality data (DNV, 1999). The validation of data is the principal element that enables a reliable database to be obtained. The validation of data depends on the credibility and the exhaustiveness of the information.

Several methods to validate the reliability database quality are presented by DNV (1999), as follows:

- Consistency checking of data with simple or cross validation of fields of the data. The intention is to make sure that the value of each field of the collection is allowable value, in terms of quantitative variable and deviations which should be within an acceptable range of variations. Cross-validation between several fields also intended to check the consistency.

- Sampling validation among recorded data. The idea is to detect error from the counted sample. Then, the total number of errors can be modeled mathematically. This method needs the participation of an expert who has extensive knowledge of the equipment.

- Homogeneity check between different plants records. If there is a significant disparity among collection plants, it may conclude real differences in failure rate or it can also reflect the incomplete data collection. An audit may be performed to avoid significant disparity. The audit shall compromise the instruction for data collection procedure, ensure that the procedures are correctly applied, an only competent employee who is eligible to do the task and software check.

- Free text summary analysis. Validate the data with available information in the free text.

From the tasks mentioned above, it can be seen that doing the quality check process could be a tedious job, especially if the quality check process has to be done manually. An automated way of quality check process can make a countable quality check and reduce the resources.

Table 2.1: Conceptual framework for the data quality (Adapted from Wang and Strong (1996))

| Data quality category | Attributes | Explanation |
|---|---|---|
| Intrinsic | Believability | Real and credible |
| | Accuracy | Accurate, correct, reliable, errors can be easily identified |
| | Objectivity | Unbiased and objective |
| | Reputation | Reputation of the data source |
| Contextual | Relevancy | Applicable to the task at hand |
| | Timeliness | Age of the data |
| | Completeness | Breadth, depth and scope of information contained |
| | Appropriate amount of data | Quantity and volume of available data |
| | Value-added | Competitive edge and adds value |
| Representational | Interpretabiity | Appropriate language, unit and data definitions |
| | Ease of understanding | Easily understood and clear |
| | Representational consistency | Format consistency |
| | Concise representation | Concise, well organized and appropriate format |
| Accessibility | Accessibility | Accessible and retrievable |
| | Access security | Restricted access to the data |

### 2.2.4 Quality Assurance and Assessment of the Reliability Data

Quality assurance in the data collection is crucial to be able to provide meaningful data for reliability analysis. Lack of good quality of data may increase uncertainty and leads to inaccurate analysis (Barabadi et al., 2015). The quality assurance should start from the beginning of the course and keep continuing during all stages of reliability data process. In general, DNV (1999) have specified quality assurance procedure as follows:

- Audit the source of the data. The data sources should contain sufficient and adequate information. Even, it has to be realistic with the available budget.

- Establish a quality plan. The quality plan should clearly define the objectives and detail procedure to be followed in data collection and processing.

- Perform a quality audit. Periodic quality audit after completion of data collection exercise is necessary. The intention is to assess the degree of success achieved during implementation and having lesson learned experience for further improvement.

- Generate a report. The report should describe the project scope, state all assumptions made, discuss if there is any problem encountered, and recommendation.

In the framework of engineering asset management, Lin et al. (2008) has attempted to compare various definitions of data quality and found a representative definition said that data quality was data that fit-for-use by the data consumers. The question raised to assess whether the conducted quality assurance was good enough and has met the fit-for-use level. Fitness for use still involves multiple parameters. In the reliability study, Wang and Strong (1996) conducted several surveys and sorting studies to develop a hierarchical framework on how to assess the quality of the reliability data.

Table 2.1 presents the formulated data quality framework. The framework has been divided into four categories: intrinsic, contextual, representational, and accessibility. The intrinsic data quality denotes the essential aspect of data quality. It shows that the data have quality in their own right. The contextual data quality evaluated the data quality based on the relevant data consumer's work field. Data that is evaluated to be high-quality data in some work field area may be considered differently in another area. They depend on the context. The representational data quality includes the format and the meaning of data. The data must be concise, consistently represented, interpretable, and easy to understand. The accessibility of data quality recognizes the importance of information systems accessibility. Accessibility emphasizes that the data should be easily accessible yet secured.

## 2.3 OREDA Database Structure

Primarily, there is a completeness requirement to meet the quality of OREDA database. Some fields are marked as the mandatory field which requires 100% of the filled data. The mandatory field is a required field for further data analysis. Some fields are marked as the highly desirable field which requires 85% of the filled data. At last, some fields are marked as the desirable field which should be collected with reasonable effort and cost. There is no percentage target for the desirable data.

OREDA has split its database into the topside and the subsea database. This separation is made due to different taxonomy applies to both databases. Taxonomy is a systematic classification of items into groups based on common factors (SINTEF, 2015). This classification on hierarchy has been made following the ISO14224.

### 2.3.1 OREDA Topside Database Structure

Unlike the organization of data discussed in Chapter 2.2.1, OREDA topside database consist of four parts. The additional database is an installation database. Table 2.2 presents the tabulation of topside database structured. It may be observed that there are discrepancies on tables collected by data collectors and tables required by data custodians. These discrepancies are discussed further in Chapter 4.1.2. It also can be seen that the databases consist of eleven tables in total. Also, two out of those eleven tables were merged to construct a new table.

#### 2.3.1.1 Installation Database

Installation database (DC_INSTALLATION table) is a database to store the name and properties of installation where data are collected. Information related to installation category, operating category, and geographic location is recorded in this database. Installation category record information regarding the installation type, whether it is a fixed platform, floating ship, onshore, etc. Operating category record information whether the installation is manned, unmanned or

Table 2.2: OREDA topside database structure

| Database | Table | Data Collector's Table Name | Data Custodian's Table Name |
|---|---|---|---|
| Installation | Installation | DC_INSTALLATION | DAS_INSTALLATION |
| Inventory | Inventory | DC_INVENTORY | DAS_INVENTORY |
| | Inventory History | DC_INVENTORY_HISTORY | DAS_INVENTORY_HISTORY |
| | Inventory Subunit | DC_INV_SUB_UNIT | DAS_INVENTORY_SUBUNIT |
| | Inventory Specific | DC_INV_M02 | DAS_INV_M02 |
| | Inventory Instrumentation | DC_INV_INSTR | DAS_INVENTORY_INST |
| | Inventory Planned Maintenance Program | DC_PM_PROGRAM | DAS_PM_PROGRAM |
| Failure | Failure Event | DC_FAILURE_EVENT | DAS_FAILURE_EVENT |
| | Failure Item | DC_FAILURE_ITEM | DAS_FAILURE_ITEM |
| Maintenance | Maintenance Event | DC_MAINT_EVENT | DAS_PM_MAINT_EVENT |
| | Maintenance Items | DC_M_MAINT_ITEMS | DAS_MAINT_ITEMS |
| Combined | Failure Event | DC_FAILURE_EVENT | DAS_FEMAINT_EVENT |
| | Maintenance Event | DC_MAINT_EVENT | |

subsea. Geographic location relates to the local environmental conditions which may influence the reliability of the equipment.

#### 2.3.1.2 Inventory Database

Inventory database consist of several parts:

- General inventory data which are common to all equipment classes (DC_INVENTORY & DC_INVENTORY_HISTORY tables) DC_INVENTORY table consists of equipment name, the main system platform, equipment unit type, operational mode, the equipment unit manufacturer, equipment unit model, external environment, date of installation of unit and statement regarding the replacement of the unit at the end of surveillance period. External environment records the climatic condition that may affect the reliability of the equipment. External environment categorizes into heavily exposed, moderately exposed, little exposed or unknown exposure. Each inventory is monitored in many periods. These periods are captured in DC_INVENTORY_HISTORY table. This table consists of the start date of surveillance, the end date of surveillance, total active state time, calculation method and number of demands. The number of demands is recorded from several equipment classes for probability failure on demand calculation. The demands are including test activation and on-demand function.

- Equipment unit specific inventory data according to the inventory forms developed for each equipment class (DC_INVENTORY_M02 table). This table consists of instrument specific property, instrument category, measurement value and unit of measurement.

- Instrument data (DC_INV_INSTR table). This table records the number of instrument loops affected for each function trip, process alarm function and control function. Besides, it records which process parameter is measured such as flow, level, pressure or temperature.

- Subunit (DC_INV_SUB_UNIT table). This table records the breakdown of equipment unit into several subunits. This table specifies all subunit with the corresponding tag number and number of installed subunits.

- Preventive Maintenance program (DC_PM_PROGRAM table). This table records information related to preventive maintenance program. This table consists of data of involved subunit, periodicity of the preventive maintenance, maintenance activity, interval type (calendar time or operating time based), scheduled preventive maintenance interval (months if calendar time based or hours if operating time based) and scheduled manhours for preventive maintenance execution.

#### 2.3.1.3 Failure Database

The general definition of failure is "The termination or the degradation of the ability of an item to perform its required function(s)." While In OREDA, failures are when some repair or work order and some repair action from maintenance personnel are necessary. Therefore, the failure event shall be linked to a maintenance event showing the performance of corrective action. The failure shall be classified to the item that has to be in the lowest level of the system hierarchy where repair activity held. Failure database has 2 tables: DC_FAILURE_EVENT and DC_FAILURE_ITEM tables. These tables consist of information related to date when failure was detected, work order number, the effect by which a failure is observed (failure mode), failure effect on equipment unit level (severity class), observed cause of a failure (failure mechanism), initiating event or root cause (failure cause), the subunit where the failure occurred (maintainable item), activity by which the failure is discovered (detection method), and failure effect (failure consequence).

- Failure mode: describes the way an equipment unit may fail to deliver their intended function. The failure mode is a description of various possible abnormal conditions that may happen to an equipment unit. There are two types of failure mode: demanded to change and undesired change. Demand change comprises an event directly related to the operation of the function, while undesired change is related to the condition.

- Severity class: describes the failure effect on equipment unit level. A failure which causes immediate and complete loss of the capability of a system is recorded as critical. A failure that does not cease all function, but compromises that function is recorded as degraded. A failure which has no immediate effect on function is recorded as incipient. Moreover, there are several failure modes which are automatically considered as critical (see DNV-GL (2016a) )because it represents a complete fail function of an equipment unit during operation.

- Failure mechanism: describes the observed cause of the failure. It shall show more exact cause that can be easily deduced technically.

- Failure cause: identify the root cause in the sequence leading up to equipment failure.

- Maintainable item: describes the subunit where the repair is held.

- Detection method: describes the activity by which a failure is discovered. It classified into planned/scheduled activities, continuous observation of the system and unexpected occurrences by chance.

- Failure consequence: describes the failure effect on a higher level than equipment unit due to its configuration dependency.

#### 2.3.1.4 Maintenance Database

The maintenance activity can be initiated either by a failure event or by preventive maintenance. In the case of corrective maintenance, the maintenance event record is linked to the failure event record. This combined tables forming so-called DC_FEMAINT_EVENT table. In the case of preventive maintenance, the maintenance record is linked to the inventory records, forming so-called DC_PMMAINT_EVENT table. Maintenance database consists of 2 tables: DC_MAINT_EVENT and DC_MAINT_ITEM tables. These tables record the maintenance date, maintenance work order, code distinguishing between corrective or preventive maintenance (maintenance category), specification of the subunit on which maintenance was performed, description of the type of maintenance action (maintenance activity), maintenance interval, maintenance times, maintenance man-hours for various disciplines and information related with maintainable item involved.

- Maintenance date: records the date when the maintenance action started or when the work order issued. As one failure event may result in more than one maintenance actions, only one maintenance date shall be recorded with separate maintenance records (work order).

- Maintenance category: classifies corrective and preventive maintenance. Further, it provides a classification of preventive maintenance actions such as calendar time based or operational time-based maintenance.

- Maintenance activity: describes the type of restoration action being performed for corrective maintenance and describes the type of preventive action being performed for preventive maintenance.

- Maintenance interval: describes the scheduled interval of the maintenance activity. Calendar time based periodic maintenance shall be recorded in months, while operational time-based periodic maintenance shall be recorded in hours.

- Maintenance times: specifies active maintenance time and downtime. Active maintenance time is the calendar time used during which actions of maintenance are performed on an item, not including time for diagnosis the failure, administrative work, waiting on spares and tools, and startup time. While downtime is calendar time used during maintenance action which includes the mobilization of resources or spare and all the preparation required. Figure 2.1 shows the illustration of maintenance times.

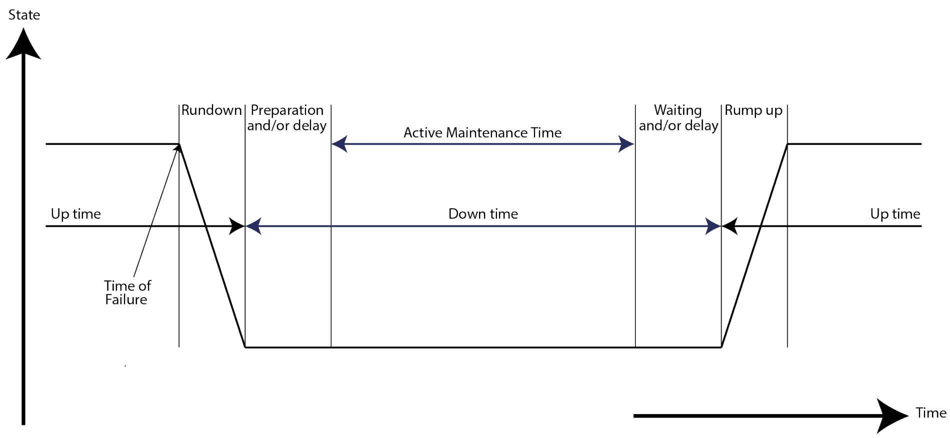



Figure 2.1: Maintenance Times (adapted from OREDA (2015))

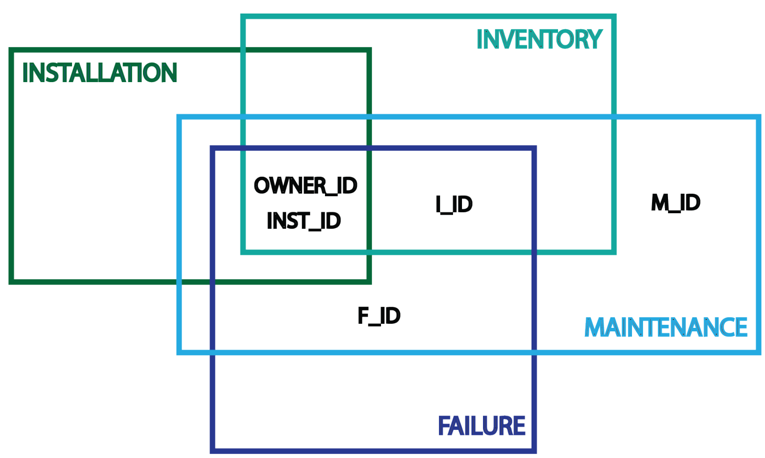



Figure 2.2: OREDA topside database primary key fields

Table 2.3: OREDA subsea database structure

| Database | Table | Data Collector's Table Name | Data Custodian's Table Name |
|---|---|---|---|
| Installation | Installation | DCS_INSTALLATION | SSDAS_INSTALLATION |
| | Installation History | DCS_INSTALLATION_HISTORY | SSDAS_INSTALLATION_HISTORY |
| | Installation Specific | DCS_INSTALLATION_M02 | SSDAS_INSTALLATION_M02 |
| Equipment Class | Inventory Equipment Class | DCS_INV_EC | SSDAS_INV_EC |
| | Inventory Equipment Class History | DCS_INV_EC_HISTORY | SSDAS_INV_EC_HISTORY |
| | Inventory Equipment Class Specific | DCS_INV_EC_M02 | SSDAS_INV_EC_M02 |
| Subunit | Inventory Subunit | DCS_INV_SU | SSDAS_INV_SU |
| | Inventory Subunit History | DCS_INV_SU_HISTORY | SSDAS_INV_SU_HISTORY |
| | Inventory Subunit Specific | DCS_INV_SU_M02 | SSDAS_INV_SU_M02 |
| Component | Inventory Component | DCS_INV_CO | SSDAS_INV_CO |
| | Inventory Component History | DCS_INV_CO_HISTORY | SSDAS_INV_CO_HISTORY |
| | Inventory Component Specific | DCS_INV_CO_M02 | SSDAS_INV_CO_M02 |
| Failure | Failure Event | DCS_FAILURE_EVENT | SSDAS_FAILURE_EVENT |
| Maintenance | Maintenance Event | DCS_MAINT_EVENT | |
| | Preventive Maintenance Events | DCS_PMMAINT_EVENT | SSDAS_PMMAINT_EVENT |
| Combined | Failure Event | DCS_FAILURE_EVENT | SSDAS_FEMAINT_EVENT |
| | Maintenance Event | DCS_MAINT_EVENT | |

OREDA topside databases have been structured following the relational database concept. Each table consists of several primary key fields which allow the table to be linked. Figure 2.2 illustrates the shared primary key fields for all parts. OREDA topside primary key fields are OWNER_ID, INST_ID, I_ID, F_ID M_ID. They are represented in a unique number which identifies the company who owns the database (OWNER_ID), an installation where the data was collected (INST_ID), inventory record (I_ID), failure record (F_ID) maintenance record (M_ID). Having these primary key fields create the possibility to find, for example, maintenance event associated with failure by link the failure event and the maintenance event tables on the three primary key fields: OWNER_ID, INST_ID, F_ID.

### 2.3.2 OREDA Subsea Database Structure

Similar to the OREDA topside database, OREDA subsea database also has an installation database. Besides, the inventory databases applied in subsea are divided based on a specific taxonomy level. In total, ORREDA subsea database consists of six parts with fifteen main data tables. Table 2.3 presents the tabulation of subsea database structured. It can be observed that there is no equivalent table for DCS_MAINT_EVENT table in data custodian's table. This table is combined with DCS_FAILURE_EVENT table forming SSDAS_FEMAINT_EVENT instead.

#### 2.3.2.1 Installation Database

Installation database (DCS_INSTALLATION table) stores one or more subsea installations comprises a set of hardware used for oil and gas production which controlled from the topside control system. Information related to the field name, installation name, installation layout, geographic location, maximum water depth, date when the installation is ready for production and intervention strategy are recorded in this database. Besides, installation history database (DCS_INSTALLATION_HISTORY) collects periods of time when the installation was under surveillance. This table consists of information related to surveillance start date, surveillance end date, total of demand count along the surveillance period and total calendar hours of surveillance.

#### 2.3.2.2 Equipment Class, Subunit and Component Database

These databases represent inventory table in various taxonomy levels. These databases have following tables: DCS_INVENTORY, DCS_INV_HISTORY and DCS_INV_M02 tables. Inventory database (DCS_INVENTORY) mainly related to the equipment unit level. DCS_INVENTORY consists of equipment identification tag, equipment name, the main design feature of an equipment unit, manufacturer of the equipment, model denomination, type of the control system, application of the equipment, control fluid type, and identification of multilateral well. Inventory history database (DCS_INV_HISTORY) creates the ability to divide smaller periods from total data collection time and enables to continue the data collection in subsequent periods. DCS_INV_HISTORY consists of the start date and end date of surveillance, specification of what start and end ate represents, number of items involved and number of demands occur on the specified period. Inventory specific database (DCS_INV_M02) specifies specific inventory data according to the inventory forms developed for each equipment class, subunit and components. This table consists of instrument specific property, instrument category, measurement value and unit of measurement.

#### 2.3.2.3 Failure Event Database

Failure event database (DCS_FAILURE_EVENT) records failure events that directly linked to components, meaning that failed components should be identified before failure event is being recorded. This table consists of summary line describing the failure event, date a failure was first observed, the effect by which a failure is observed (failure mode), failure effect (severity class), cause which has led to a failure (failure cause), the way a failure first observed (failure detection), operational phase during which a failure occurred, consequences due to a failure (failure consequences), amount of lost gas and oil, and underlying technical causes (failure mechanism). Some of the parameters recorded in this failure event have been explained on the topside section. The difference is that in the subsea application, failure mode and severity are identified in equipment, subunit, and component level.

#### 2.3.2.4 Maintenance Database

Maintenance database (DCS_MAINT_EVENT and DCS_PM_MAINT_EVENT) have two ways of recording maintenance data. Similar to the topside application, they are maintenance linked to the failure event for corrective maintenance and maintenance without preceding failure for preventive maintenance. These tables record date when maintenance action was carried out, maintenance work order, type of maintenance action performed (maintenance activity), consequence of maintenance action (maintenance consequence), duration of active maintenance time, total duration spent in unavailable state (downtime), type of main intervention vessel,

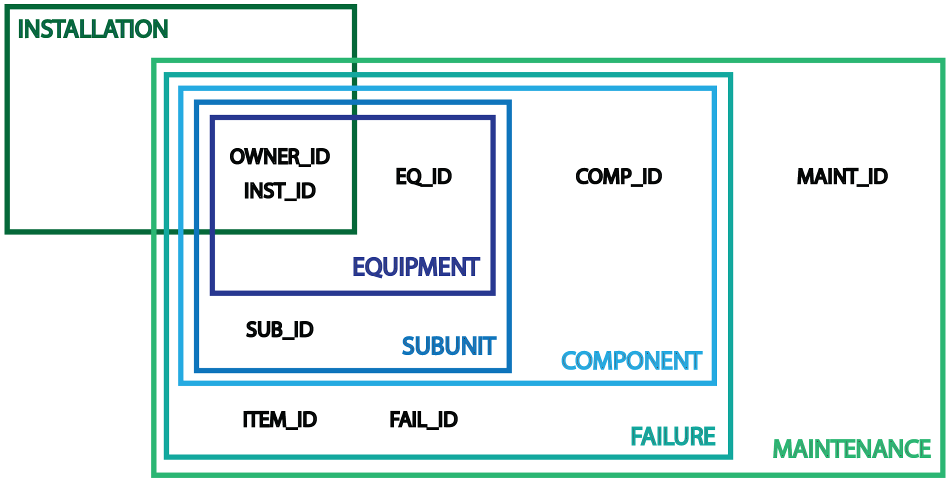



Figure 2.3: OREDA subsea database primary key fields

time spent of vessel on location for restoring particular failure, type of support intervention vessel, time spent of support vessel on location for restoring particular failure, main maintenance resources and time for mobilization of vessel in days. Some of the parameters recorded in this failure event have been explained on the topside section.

- Main intervention vessel: specifies the type of vessel being used (if any) for the intervention.
- Main vessel time: records the total time of the vessel spent on location for doing the intervention work, excluding sailing, waiting and idle time.
- Support intervention vessel: specifies the type of support vessel being used (if any) for the intervention.
- Support vessel time: records the total time of the support vessel spent on location for doing the intervention work, excluding sailing, waiting and idle time.
- Maintenance resources: specifies which maintenance resources being used. The resources could be either diver or ROV.
- Mobilization time: records number of days required to mobilize the intervention vessel until it is ready.

OREDA subsea databases have been structured following the relational database concept. Each table consists of several primary key fields which allow the table to be linked. Figure 2.3 illustrates the shared primary key fields for all parts. OREDA subsea primary key fields are OWNER_ID, INST_ID, EQ_ID, SUB_ID, COMP_ID, ITEM_ID, FAIL_ID and MAINT_ID. They are represented in a unique number which identify the company who own the database (OWNER_ID), installation where the data was collected (INST_ID), equipment inventory record (EQ_ID), subunit inventory record (SUB_ID), component inventory record (COMP_ID), failure item record

(ITEM_ID), failure event record (FAIL_ID) & maintenance record (MAINT_ID). Having these primary key fields create the possibility to find, for example, the failure on subunit by link the failure event and the inventory subunit tables on the four primary key fields: OWNER_ID, INST_ID, EQ_ID & SUB_ID

# Chapter 3

# Analysis of The OREDA Quality Assurance and Framework

This chapter analyzes the current OREDA quality assurance and assesses the current OREDA quality based on the reliability database framework. First, the roles of involved actors in the OREDA data processes are introduced. Second, current quality assurance protocol, which involves all actors, is analyzed. The analysis resulted in a formulation of the new extension of the OREDA quality assurance protocol. Finally, comprehensive explanations on the attributes of the reliability database framework, which are not explicitly mentioned as the OREDA's main quality focuses are discussed. These attributes are timeliness, value-added, and accessibility attribute. As results, the fundamental technological principal approach and list of the unnecessary & empty fields are presented.

## 3.1 OREDA Data Processes

There are several actors involved in the OREDA data processes. They are data collector, data custodian and data consumer. Each of them have their own dedicated tasks and roles. The summary of the OREDA actors involved is illustrated in Figure 3.1.

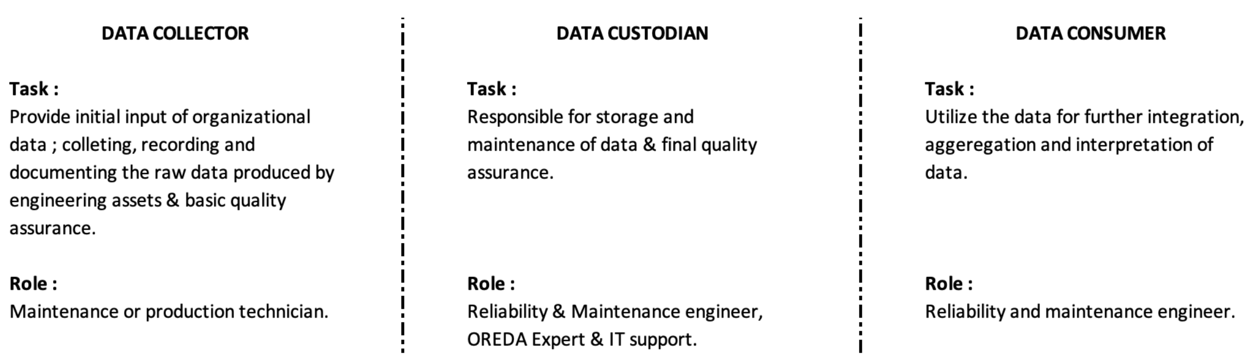



Figure 3.1: Actors in OREDA data processes

Data collectors are in charge of the data-collection process. Data collectors provide initial input of organizational data. Lee and Strong (2003) emphasized that knowledge held by the data collectors played a vital role in the data quality. He also mentioned that accuracy and completeness were the main attributes contributed by data collectors. Lee (1996) said that data collectors should have asked why did people need these data while collecting the data. This knowing-why was gained from experience, objective understanding and implementation of procedures during the collection execution (Lee and Strong, 2003). It is essential that the data collectors understand why they need to collect the data and how it is utilized. By knowing why, the data collectors may appreciate the process and contribute to produce better quality data. In addition, data collectors have a task to do necessary quality assurance before delivering the database to the data custodian (SINTEF, 2007).

Data custodians are in charge of data storage and maintenance. The data custodians have to ensure that all fields are completely filled, to complete the data processing on time and to make the data accessible. In other words, completeness, timeliness and accessible are the main attributes contributed by the data custodians. Lee (1996) said that data custodians should have asked what data should they be storing. This knowing-what determine the objective of the whole data processes. As the implication, data custodians determine the required fields and specify the completeness rules so that the valid range value that has to be filled is achieved. Data custodians have to do quality assurance on the received data and final quality assurance before release the database in the form of an application that further be used by the data consumers (SINTEF, 2007).

Data consumers are in charge of the data-utilization processes which may involve aggregation, integration and retrieval of data. The data consumers knowledge is correlated with the relevancy attribute. It is only the data consumers who know whether the data are relevant or not. Lee (1996) said that data consumers should have asked how to use the data. Data consumers with high knowing-how can assess and use the data for reliability analysis. Furthermore, feedback from data consumers, as the database users, to the data custodian may bring quality improvement (SINTEF, 2007).

It is crucial for collecting high quality of reliability and maintenance data that has a specification giving a clear description of boundaries, taxonomy, equipment hierarchy, data format, etc. OREDA understand that use of such data may lead to entirely wrong decisions if the data being collected are not comparable or they are different. Therefore, OREDA prepared a sort of procedure named data collection guideline, which currently used as a standard. The guideline specifies the technical requirements of reliability data, the data collection procedures and the quality control procedures.

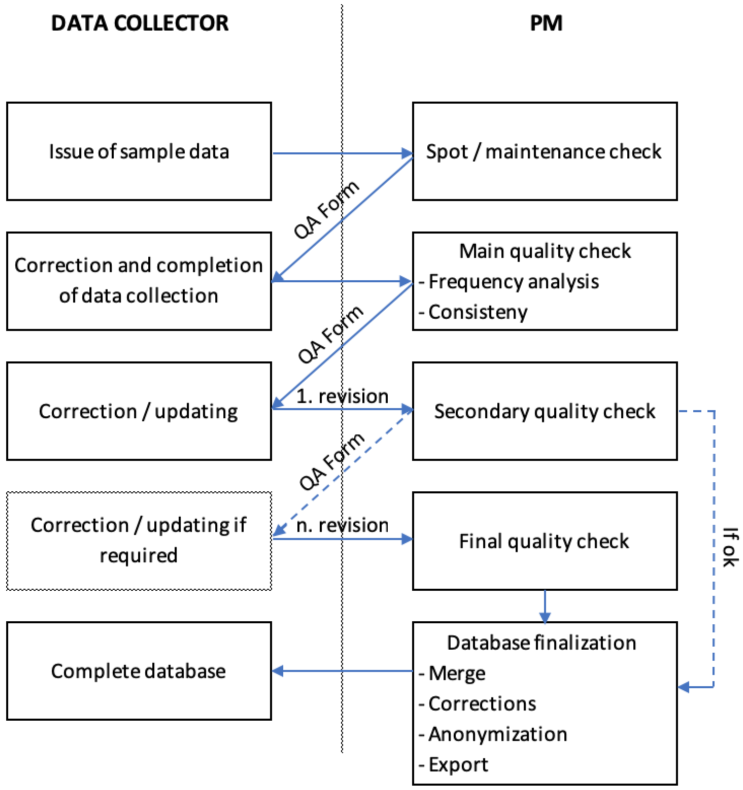



Figure 3.2: Existing OREDA QA/QC Protocol (adapted from SINTEF (2007))

## 3.2 OREDA Quality Assurance Protocol

### 3.2.1 Current OREDA Quality Assurance Protocol

In general, the objective of OREDA is to provide a basis for improving safety and cost-effectiveness through data collection and analysis by establish and exchange reliability database (OREDA, 2015). A quality assurance (QA) procedure has been generated to achieved the main objective. Vroeijenstijn (1995) in their previous studies have referred quality assurance to a systematic, structured, and continuous attention to quality in terms of quality maintenance and improvement. The OREDA quality assurance consists of three main activities (DNV-GL, 2016a):

- Specify the requirement as to data that is to be collected.
- Verify that these requirements are fulfilled during the data collection.
- Initiate proper corrective actions in case there is a deviation in data quality compared to desired goals.

OREDA has created a quality assurance protocol at such to ensure that the data processes resulted in a high-quality database. Figure 3.2 presents the current OREDA quality assurance protocol. The protocol consists of two actors: data collector and PM. PM stands for project manager which represent the data custodian. Although quality assurance protocol has been developed, it has not entirely described the main activities. Therefore a new extension of quality assurance protocol is proposed. Figure 3.3 shows the complete quality assurance protocol based on experience and discussion with the OREDA expert.

### 3.2.2 New Extension of the OREDA Quality Assurance Protocol

The quality assurance protocol starts with the data custodians set the objectives, requirements and planning measures of the data processes (ISO14224, 2016). Data custodian establish the guidelines and procedures as the basis for a consistent approach in the data collection. The guideline describes the OREDA concept and gives definitions, code and interpretation rules for the to-be-collected reliability data (SINTEF, 2015).

The data collectors start to collect data with the guideline as their primary reference. Data collectors may observe problems and deviations in data collection guideline. In that case, data collectors may report on changelog site to the data custodians. Data custodian, if necessary, to update the guideline. Then, data collectors have to verify the collected data and perform the automatic basic QA. At this stage, the deviations in the data itself are observed. Data collectors should do necessary action to eliminate the deviations before handed over the database to the data custodians.

Data custodians receive the collected data from the data collectors. Data custodians run the automatic verification tools and do the quality assurance check on the received data. It is expected that the data custodians should get an immediate response from these activities. If there are deviations, data custodians document the findings in the QA form and return the collected data. If there are no deviations, data custodians continue with manual data verification. OREDA provides list of the automatic verification tools and checklist for quality check on the received data in their quality assurance checklists document.

Any deviations, corrective actions, interpretation problems, change of proposal, problem related to interpretation of the guideline, data collection procedure and quality assurance shall be documented in writing (SINTEF, 2015). Feedback of information is essential part of the data collection and the validation exercise of the quality review. Every feedback is collected as accumulating experience with the OREDA.

In case of immediate action is required due to corrective action, a changelog can be utilized. If there is no recommendation received, project manager should do the evaluation and give recommendation of necessary corrective action for further action.

Data collectors do the correction actions respective to the QA form. All corrections actions then recorded in the same QA form. Then, data collectors have to perform another data verification and automatic basic QA before sends back the updated data to the data custodians.

Data custodians do the manual data verification and perform the final spot check as well. At this stage, the collected data is expected to be the final data. If there are still deviations detected, re-circulation of QA form with revision needs to be done. Once the collected data pass the manual data verification and QA, the collected data is ready to be merged with data from the other

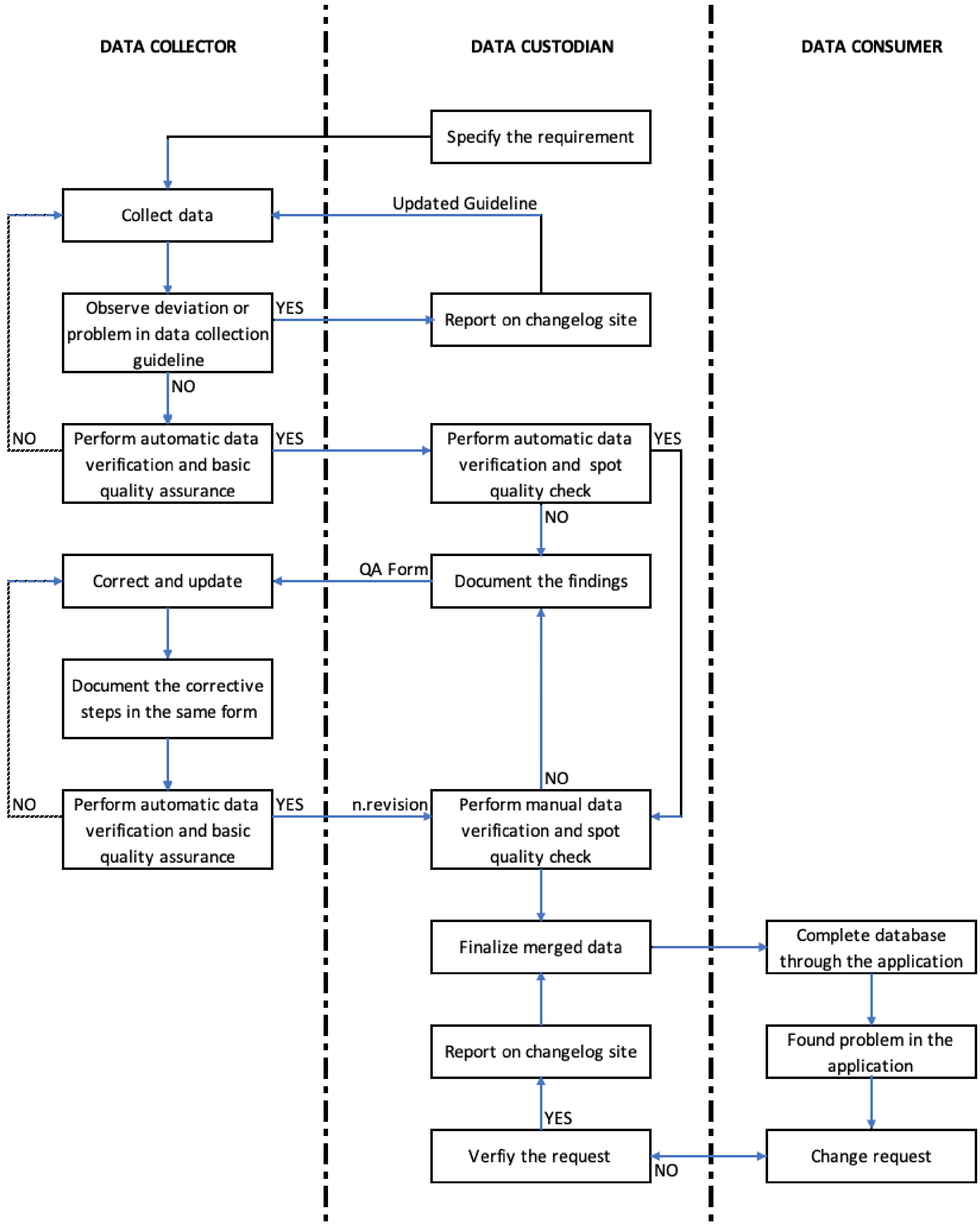



Figure 3.3: New extension of the quality assurance protocol

company's data collectors.

Data custodians have to merge all the collected data into one database. Data consumers only have access to the database via an application as an interface. the changelog is also featured to show a problem in the application. If the data consumers found that some of the data are wrong, the formula does not working, or they have something in addition. Data consumers may create a change request addressing their issues. If the data custodian assesses and found that it is a good point, then the data custodians make a changelog entry of that issue and update the application. If the data custodian found that it is a miss understanding, the change request goes back to the data consumers, and the data custodians explain the condition and no change in the application.

This proposed quality assurance protocol applies for experience data collectors who can collect all data before doing basic QA and handed over the data to the data custodians. It is suggested for those data collectors with a low level of experience, first to deliver initial or sample data. Perform often data verification and automatic basic QA on an ongoing basis. The intention is to get feedback on an error on the earliest phase.

## 3.3 OREDA Reliability Framework

### 3.3.1 OREDA Requirement Specifications

Any reliability analysis relies on the quality of the collected reliability data. A recent study by Ciliberti et al. (2019) shows that higher data quality enabling better data-driven decision-making. To assess the quality of reliability data, ISO14224 (2016) has defined the definition of data quality into several main elements as the followings :

- Completeness of data in relation to specification.
- Compliance with definitions of reliability parameters, data types and formats.
- accurate input, transfer, handling and storage of data (manually or electronic).
- Sufficient population and adequate surveillance period to give statistical confidence.
- Relevance of the data to the need of the users.

Those five elements mentioned above emphasize the necessity of completeness, representational consistency, accuracy, appropriate amount of data and relevancy attributes. These elements are in line with the data quality framework formulated by Wang and Strong (1996). Table 3.1 tabulated the match between the two.

SINTEF (2015) stated that to achieve a high quality of data, the data custodian should prepare the quality plan measures. A quality plan consists of listed tasks start from the organization, the documentation, the data collection, and the data processing. The main items for the quality plan are:

Table 3.1: OREDA data quality main elements fitted in the conceptual framework for the data quality (Adapted from Wang and Strong (1996))

| Data quality category | Attributes | OREDA main elements |
|---|---|---|
| Intrinsic | Believability | |
| | Accuracy | Accurate input, transfer, handling and storage of data (manually or electronic). |
| | Objectivity | |
| | Reputation | |
| Contextual | Relevancy | Relevance of the data to the need of the users. |
| | Timeliness | |
| | Completeness | Completeness of data in relation to specification. |
| | Appropriate amount of data | Sufficient population and adequate surveillance period to give statistical confidence. |
| | Value-added | |
| Representational | Interpretabiity | |
| | Ease of understanding | |
| | Representational consistency | Compliance with definitions of reliability parameters, data type and formats. |
| | Concise representation | |
| Accessibility | Accessibility | |
| | Access security | |

- The data collectors should have sufficient training in OREDA software and guideline. It must also be investigated whether the personnel with sufficient competence to undertake the data collection task are available. If no, a contractor with expertise in this area should be used.
- Sufficient equipment population and data source are available. The data sources should be listed and an evaluation of the cost-benefit of using alternative sources and data collection approaches should be evaluated. The data sources should contain sufficient and adequate information for fulfilling the OREDA requirement.
- Data collection plans are prepared. This including schedules, milestones, organization chart which define the responsibilities, list of objectives and list of items on which data are to be collected.
- Tools and procedure for checking and validating the data quality.
- Routines and procedure for reporting deviations. Develop information flow diagram and establish the route feedback for information.

### 3.3.2 OREDA Data Verification Check Items

The OREDA data verification are split into two steps: automatic and manual data verification. Automatic data verification is performed by the data collectors before send the collected data to the data custodians. The data custodians also run the automatic data verification tool when receive collected data as a cross check of data quality. The automatic data verification give immediate feedback of the deviations in the database. On the other hand, manual data verification is performed as a final check before the database is to be released.

Besides data verification, data custodians have to do the quality check on received data. This quality check consist of listed items which specifies which equipment and subject to be checked. This list is continuously updated along the data processes. ISO14224 (2016) has established set of check items that should be verified at a minimum as followings:

- The origin of the data is documented and traceable.
- The data originate from similar equipment type, technology and operating conditions.
- The equipment is relevant for the purpose.
- The data comply with definitions and interpretation rules.
- Recorded failures are within the defined equipment boundary and surveillance period.
- The information is consistent
- Data are registered in the correct format.
- Sufficient data are collected to give acceptable statistical confidence.
- Operating and maintenance personnel are consulted to validate the data.

These items represents more detail approaches of data quality main elements, as discussed in Chapter 3.3.1. These items are considered as the main focuses of OREDA.It is interesting to find out which attributes from the data quality framework are covered and which are not. Taking the same table to compare with, Table 3.2 tabulated the match between the two. The highlighted blue rows represented the match between data quality attributes framework and OREDA quality check elements. Pragmatically, it can be observed that OREDA put less attention on following attributes: Timeliness, Value-added Accessibility

#### 3.3.2.1 Timeliness Attribute

Several studies investigating the timeliness have been carried out. Zhang et al. (2017) argued that timeliness was related to a product life cycle. When an item has passed the maturity and enters recession, fewer customers paid attention to it. As an implication, the older the data, the fewer the data being utilized. Zanetti et al. (2015), Sağlam and Temizel (2014) & Larsen et al. (2009) who had done research in the medicine field of study concluded the necessity of information timeliness. In this field, the timeliness referred to information which was sufficiently up-to-date at the time of publication. As an implication, the sooner the data being published after the data collection process, the better.

According to Wang and Strong (1996), timeliness is the age of data that considered appropriate for the current task. Both definitions on the timeliness about old data and up-to-date

Table 3.2: OREDA data quality check fitted in the conceptual framework for data quality (Adapted from Wang and Strong (1996))

| Data quality category | Attributes | OREDA quality check elements |
|---|---|---|
| Intrinsic | Believability | The origin of the data is documented and traceable. |
| | Accuracy | Recorded failures are within the defined equipment boundary. |
| | Objectivity | Operating and maintenance personnel are consulted to validate the data. |
| | Reputation | The data originate from similar equipment type, technology and operating condition. |
| Contextual | Relevancy | The equipment is relevant for the purpose. |
| | Timeliness | |
| | Completeness | The various data fields comply with the OREDA requirements for completeness. |
| | Appropriate amount of data | Sufficient data are collected to give acceptable statistical confidence. |
| | Value-added | |
| Representational | Interpretabiity | The data comply with the definitions and interpretation rules (e.g. definition of failure) |
| | Ease of understanding | The data are registered with correct format |
| | Representational consistency | The information is consistent |
| | Concise representation | The data are registered with correct format |
| Accessibility | Accessibility | |
| | Access security | |

publication data are relevant. Although some research has been carried out on the timeless, the mechanism by which demonstrate the quantification on reliability database has not been established.

OREDA start the data collection in 1981 and first established data handbook in 1984 (OREDA, 2015). The new version of OREDA database is run in phases typically lasting for 2 - 3 years. Updating data in phases could be the solution for delivering up-to-date data. However, the entire database consists of up to 38 years of old data. All OREDA members have access to the database via an application which enables them to filter the data based on desired phases. They may exclude old data records. However, such feature is not available for OREDA handbook users.

It can be argued whether the existence of such old data is still relevant. The wear out period of each equipment can be varied following the bathtub curve. Nevertheless, technology keeps developing in the past 40 years. For example, Yokogawa (2019) shows an evolution of their pressure transmitters product. It can be seen that there is a change in the fundamental technology principal. In this case, it is the sensoring principal. The sensoring principal has been developed from an analog sensor to a digital sensor. There might be a case where the companies are no longer has the old technology installed in their installations due to the wear out or obsolete item.

An approach of timeliness analysis in OREDA database can be performed by 'manufact and model' analysis. OREDA collects manufacturer and model information of all items. The timeliness analysis is performed by comparing all stored items with the installed items. Items with 'manufact and model' information which not exist in the installed item are considered as out-of-date data. To do so, the completeness of 'manufact' and 'model' data are necessary. Currently, 'manufact' is considered a mandatory field, while 'model' is only considered a desirable field. It is proposed to include 'model' field to mandatory field for a better timeliness analysis.

This approach may be too narrow. Pragmatically exclude the out-of-date data may result in very few data. Data that are very few are not considered representative for reliability analysis. Understanding the changes in the fundamental technology principal become vital information. If the changes are minor, it is better to keep the data because the changes may not affect the failure behavior. It might be challenging to do the work for excluding some data. However, changes in the fundamental technology principal can be used as a basis for excluding out-of-date data. Change in the fundamental technology principal changes the failure behavior. Unfortunately, current OREDA database does not collect this information.

In the reliability study, failure is classified as a systematic and random failure. This classification is made based on the root cause. In the author's specialization project, Harahap (2018) found that the systematic failures dominated the total system failures. IEC61508 (2010) defines the systematic failure as deterministic failure, which can only be eliminated by design modification or operational procedures, or documentation or other relevant factors. One of the influence factors of the systematic failure is the complex system interaction. As the system complexity increases, then the probability to experience systematic failure also increases. Changes in the fundamental technology principal may increase the complexity of the system. Therefore, using the out-of-date data influences the reliability analysis.

#### 3.3.2.2 Value-added Attribute

Wang and Strong (1996) defined the value-added data as data that give value to the operation. Adams (2018) pointed out that there was a trade-off between the quality and the availability of the database. Most of the cases, it fell to the side of availability. There was a tendency to collect more and more data, even though those data might not contribute an added-value to the database users.

In the transportation field, Cats and Loutos (2016) evaluated the added-value of online bus arrival prediction scheme by conducting qualitative research on the passenger. The research aims were to find out the waiting time savings and capability to foresee downstream vehicle trajectories. Pelzer et al. (2014) analyzed the added-value of the planning support system from the user perspective. The user feedback was used as an input for the better utilization at present days. Both researchers show the importance to include the data consumers experience for the added-value study.

It is unfortunate that this study is only able to evaluate value-added from the data custodian perspective. This research does not have a chance to gather feedback from data consumers. As the data custodians receive collected data from the data collectors, the data custodian may be aware of empty fields in the database. These fields are good-to-have fields which are useful but not prioritized to be collected. The filling level of these fields is low; even none of the company

Table 3.3: OREDA topside unnecessary & empty fields

| Tables | Field |
|---|---|
| DC_FAILURE_EVENT | MERGED_I_ID<br>MERGED_EC_CODE |
| DC_MAINT_EVENT | M_RES_DRILL_RIG<br>M_RES_DIVING_VESSEL<br>M_RES_SERVICE_VESSEL<br>M_RES_DIVERS<br>M_RES_ROV_ROT<br>M_PATF_PERS<br>MERGED_I_ID<br>MERGED_EC_CODE |
| DC_INSTALLATION | I_SOURCE<br>M_SOURCE<br>F_SOURCE<br>INST_LAST_I_ID<br>INST_LAST_F_ID<br>INST_LAST_M_ID |
| | INST_SURV_START |
| | INST_SURV_END |
| DC_INVENTORY | I_SURV_START_DATE_BAK |
| | I_SURV_END_DATE_BAK |
| | I_OPER_TIME_BAK |
| | I_OPER_TIME_CODE_BAK |
| | I_NO_OF_STARTS |
| | I_SUBSEA_CS |
| | MERGED_I_ID |
| | MERGED_EC_CODE |
| DC_INVENTORY_HISTORY | IH_ID |
| DC_INV_M02 | INV_SPEC_NUMB2 |
| | INV_SPEC_DATE |

Table 3.4: OREDA subsea unnecessary & empty fields

| Tables | Field |
|---|---|
| DCS_MAINT_EVENT | VESSEL_MAIN_TIME<br>VESSEL_SUPP<br>VESSEL_SUPP_TIME |
| DCS_INSTALLATION | WATER_PROD<br>OIL_PROD<br>GAS_PROD<br>H2S<br>C02<br>CONTROL_FLUID<br>COND_PRODE |
| DCS_INSTALLATION_HISTORY | I_START_REASON<br>I_END_REASON<br>I_NO_OF_ITEMS<br>IH_ID |
| DCS_INV_CO_HISTORY | ICO_START_REASON<br>ICO_END_REASON |
| | ICO_NO_OF_ITEMS |
| | ICH_ID |
| DCS_INV_CO_M02 | INV_SPEC_NUMB2 |
| | INV_SPEC_DATE |
| DCS_INV_EC_HISTORY | IEC_START_REASON |
| | IEC_END_REASON |
| | IEC_NO_OF_ITEMS |
| | IEH_ID |
| DCS_INV_EC_M02 | INV_SPEC_NUMB2 |
| | INV_SPEC_DATE |
| DCS_INV_SU_HISTORY | ISU_START_REASON |
| | ISU_END_REASON |
| | ISU_NO_OF_ITEMS |
| | ISH_ID |
| DCS_INV_SU_M02 | INV_SPEC_NUMB2 |
| | INV_SPEC_DATE |

members collect it. Moreover, these fields are not included on the merged database. These fields potentially represent unnecessary fields which no longer used for data analysis. It can be concluded that these fields may not give any added-value for data analysis. It is recommended to reconsider the existence of these fields. Reducing the number of fields may reduce the workload of every data processes roles. Furthermore, reducing the number of fields means reducing the spend resources (time and expenses).

A thorough evaluation has been performed to both topside and subsea collected data from various company members. The potential unnecessary fields were justified when none of the companies filled those fields and the data from the fields were not stored into the final database. Table 3.3 shows the empty fields for the topside database with its respective tables. Table 3.4 shows the empty fields for the subsea database with its respective tables.

It can be seen from Table 3.3 several fields that are repeated on different tables, for examples 'MERGED_I_ID' and 'MERGED_EC_CODE'. These fields are made to accommodate merging identification number of installation and equipment class. It can be argued when was the last time these fields utilized or how often these fields are being utilized. In phase 12 database, for instance, these fields are empty fields.

For subsea application, Table 3.4 shows that 'START_REASON' and 'END_REASON' fields are repeated on the different tables. These fields may have various fields name based on the various taxonomy level. These fields indicate the reason for such item start or end. In terms of an item stop due to failure event, the reason for failures is covered by 'FM_CODE' which represent the failure mode and 'FCAUSE_NO' which represents the failure cause. Failure mode specifies all possible reason in which equipment may fail, while failure cause determines the cause of failure, whether systematic or random. In terms of operation, the information on the current state of operation is covered by 'OPR_MODE'.

Further investigation can be done to find out the root cause of these empty fields. It can be the case that the data collectors do not have proper knowledge of filling these fields. Feedback from the data consumers can become a consideration, whether they are using these fields for their reliability analysis or not.

#### 3.3.2.3 Accessibility Attribute

Accessibility in the database talks about how easily and quickly the available data is retrievable. Another dimension of accessibility also talks about the access security, in which extent the database can be kept secure by applying restricted access through it Wang and Strong (1996).

Many of the researchers in the information science conducted investigations to improve the information accessibility (Fidel and Green, 2004). There are various attributes to qualified the degree of the database accessibility. Fidel and Green (2004) used a survey to assess those attributes. As a result, the top three attributes have been identified: the knowledge of the sources, the time saved, and the physical position. Data consumers prefer to access the database in which they know how to find it, how to use it, and confidence that the database have the necessary information to satisfy their needs. They desire to have a data source in which can answer their needs quickly while the physical position refers to a condition where the data source located in the immediate work area of the data consumers, so it is easily obtained.

Gens (2008) claimed that the highest weight issue arising in the database are within the security. Fernández-Medina and Piattini (2005) explained how the organization increasingly depended on the correct security implementation and confidentiality of information. Thus, database protection is a strict requirement that must be carefully considered. As the information system and the database are getting more complex, sophisticated security is a necessity. In a recent study, Bailey et al. (2017) proposed several layers of security protection policy. First, the identity and access management control. This policy was utilized to grant remote access for all data consumers with their user name and password. Second, the multifactor authentication. This second layer authentication was introduced for access to the system critical components. Usually, it can be in the form of a digital token or a dongle. Next, the network traffic flow control by applying the network firewalls. At last, the implementation of encrypting stored data and data in transit.

OREDA has developed the cloud database application to be used by all the members. This cloud database improve the accessibility to the OREDA database. Even though OREDA do not explicitly mention the access security as their quality main element, it can be found that OREDA regulated the data handling process. SINTEF (2015) mentioned in the guiding rules that every data collectors shall ensure that collected data are stored within the confidentiality and security requirements. The security requirements, as minimum, consist of security code for computers with stored data, safe depot for all original hard copy or electronic storage media, regular backup of original data in separate work files and restricted authorized personnel for access to the data receipt, handling and shipment. The data collected shall be secured by password protection during transfer. All transferred data to OREDA project is considered as confidential.

# Chapter 4

# Analysis of The Phase 12 OREDA Database

The fourth chapter is concerned with the summary of the experience gained during the involvement on the phase 12 database process. Any discoveries throughout the work and potential improvement are discussed. The chapter starts with discussions on potential automated fields and potential deleted fields which are discovered while changing the OREDA format. List of formulas are presented for the potential automated fields. Then, a thorough analysis of the current OREDA quality assurance checklists is presented. As a result, new extended quality assurance checklists are proposed.

## 4.1 Change the OREDA Format Experience

The author had an opportunity to experience himself worked with the phase 12 database. The primary duty was to convert the existing data collector's database format into the new data custodian's database format. Then, the new databases from company members merged into one database. The task includes the quality assurance on the received database from the data collectors, change the database format of several company members involved in the phase 12 database and the quality assurance before merging the databases. During the work period, the author gained various experience that may be beneficial for the future of OREDA data processes.

One of the stage while converting the database was fields mapping. (Sun, 2004) denoted mapping as a process to identify the equivalent fields between multiple databases. He discussed the necessity of the automated mapping process because the manual process consumed time and resources. Oliva et al. (2019) proposed a predictive model for the mapping process based on the relevant pattern.

If the databases are using the same term, the mapping process can be done directly. If they are different, a thorough analysis of the filled data and the patterns are necessary to obtain the correct mapping. A demonstration of how the mapping process has been performed can be seen in Figure 4.1. The demonstration takes the installation history table from the subsea application as an example. There are Two columns. The first column is representing the available fields from the data collector side, while the second column is representing the required fields from

| DCS_INSTALLATION_HISTORY | |
|---|---|
| DATA COLLECTOR | DATA CUSTODIAN |
| OWNER_ID | OWNER_ID |
| INST_ID | INST_ID |
| I_START_DATE | I_SURV_START_DATE |
| I_END_DATE | I_SURV_END_DATE |
| I_NO_OF_DEMANDS | DEMAND_COUNT |
| | CALENDAR_HRS |
| I_START_REASON | |
| I_END_REASON | |
| I_NO_OF ITEMS | |
| I_OREDA_PHASE | |
| I_COMPANY_PHASE | |
| IH_ID | |

Figure 4.1: OREDA fields mapping process for SSDAS_INSTALLATION_HISTORY table

the data custodians side. These cells have been highlighted in colors for a more straightforward explanation.

The blue highlighted cells represent the match fields. Some of the match fields have the same name, while others are slightly different. The mapping on those fields with the same name can be done directly, while the slightly different name may create confusion and requires further interpretation. A new template with the same field name for both the data collector and the data custodian may reduce the difficulty and the probability of an error on the mapping work. Moreover, having the same field name may beneficial for the automatic mapping.

The yellow highlighted cells represent the new required field from the data custodian side, which is not equivalent to the data collector's table. The data collector's table is not able to supply direct information for the data custodian's needs. However, these fields may be generated from the other fields through a calculation or further interpretation. These kinds of fields are potential fields for the automation process.

The green highlighted cells represent the potential unnecessary fields to be collected. These fields are no longer needed on the final merged database. It can be argued whether to keep the fields or to remove them. Reducing numbers of fields to be collected could be a significant saving for the project.

The orange highlighted cells represent the utility fields. These fields might not be needed for the final database. However, these fields are used to identify the properties of the current database status. It is crucial to have the status information even though they are not used any further.

Table 4.1: OREDA topside automation opportunity

| Tables | Potential Automate Field | Reference Fields | |
|---|---|---|---|
| DAS_FAILUE_EVENT | AGE | F_DETECTED_DATE | |
| DAS_INVENTORY_HISTORY | CALENDAR_HRS | I_SURV_START_DATE | I_SURV_END_DATE |
| DAS_INV_M02 | UOM | PROPERTY | |
| DAS_PM_PROGRAM | INTERVAL_UNIT | PM_TYPE | |
| DAS_FEMAINT_EVENT | HAS_DONWTIME | M_DOWNTIME | |
| | HAS_ACTIVE_TIME | M_ACTIVE_TIME | |
| | HAS_MANHOUR | M_TOTAL_MANHOUR | |

### 4.1.1 New Fields - Potential Automated Fields

In OREDA, the fields mapping process has been done manually. Most of the fields from both databases are having the same terms. Some of them are slightly different. However, there are fields which are empty and have an opportunity for the automation process.

Table 4.1 shows a summary of potential automated fields on the topside application and Table 4.2 shows a summary of potential automated fields on the subsea application. Each potential automated field has reference fields to look into as a basis source to supply the required information. A calculation or a function are implemented to the reference fields to obtain the required fields.

#### 4.1.1.1 Automation Opportunity on The Topside Application

On the topside application, there are 7 potential automated fields from 5 topside tables. Calculations and functions have been developed to construct the automation process.

First, the 'AGE' field, which represents the age, has no specific information stated on it. However, since this field exists in the failure event tables, it is reasonable to consider it as the age of failure. the age of failure can become a reference whether to include or to exclude the data for the reliability data analysis. Equation 4.3 shows a formula to fill this field automatically.

$$AGE = F\_DETECTED\_DATE - CURRENT\_DATE \tag{4.1}$$

Second, the 'CALENDAR_HRS' field, which represents the calendar hours, is the interval of time between the start and the end of data surveillance (DNV-GL, 2016b). Calendar hours calculate calendar time in hours. Equation 4.2 shows calculation on how to obtain the information for this field automatically.

$$CALENDAR\_HRS = ((I\_SURV\_START\_DATE - I\_SURV\_END\_DATE) + 1) * 24HOURS \tag{4.2}$$

Next, the 'UOM' field, which represents the unit of measurement, has limited available information. The abbreviation has been revealed through investigation on examining the current table format and the filled data. The information about the unit of measurement was gathered by finding the value on corresponding rows (Ms.Excel vlookup function) of the 'topside properties' table. This process is performed by taking the 'PROPERTY' field as the reference field. The properties table for the topside application can be seen in Appendix A.

Next, the 'INTERVAL_UNIT' field, which describes itself, represents a timely unit between periodic maintenance. Interval unit is so related to the 'PM_TYPE' which represents the type of preventive maintenance, either the calendar time-based or the operating time-based. Interval unit is months for the calendar time-based preventive maintenance and hours for the operating time-based preventivr maintenance. This field can be filled automatically by applying a formula which regulates 'if PM_TYPE = C (calendar time), then INTERVAL_UNIT = M (months)' and 'if PM_TYPE = O (operating time), then INTERVAL_UNIT = H (hours).'

At last, the set of fields from 'DAS_FEMAINT_EVENT' table. They are 'HAS_DOWNTIME', 'HAS_ACTIVE_TIME', and 'HAS_MANHOUR'. These fields show the existence of each particular time, which are the downtime, the active maintenance time and the total manhour. The field filled with the binary value of either 1 or 0. 1 represents that the particular time exists, while 0 represents the nonexistence. The formula for automatic process can be developed as 'if M_DOWNTIME > 0, then HAS_DOWNTIME = 1, else HAS_DOWNTIME = 0', 'if M_ACTIVE_TIME > 0, then HAS_ACTIVE_TIME = 1, else HAS_ACTIVE_TIME = 0' and ' 'if M_TOTAL_MANHOUR > 0, then HAS_MANHOUR = 1, else HAS_MANHOUR = 0'. There are several fields of manhour recorded such as the electrical manhour, the mechanical manhour and the instrument manhour. However, the formula treats the total manhour as the whole representative.

#### 4.1.1.2 Automation Opportunity on The Subsea Application

On the subsea application, there are 11 potential automated fields from 9 subsea tables. Some subsea potential automated fields are the same with the topside application. However, they have different terms used as the reference fields. Some subsea potential automated fields also repetition fields from other tables. The repetition is due to the different taxonomy level: the equipment class, the subunit, and the component level.

First, the 'AGE' field. The age field on failure event has the same meaning with age field from the topside application. Thus:

$$AGE = F\_DATE - CURRENT\_DATE \tag{4.3}$$

Second, the set of fields from 'SSDAS_FEMAINT_EVENT' which consists of 'HAS_DOWNTIME', 'HAS_ACTIVE_TIME', and 'HAS_MAINT_RESRC_TIME'. These fields are fields to show the existence of each particular time, which are the downtime, the active maintenance time and the maintenance resource time. The field filled with the binary value of either 1 or 0. 1 represents that the particular time exists, while 0 represents the nonexistence. The formula for automatic process can be developed as 'if DOWNTIME > 0, then HAS_DOWNTIME = 1, else HAS_DOWNTIME = 0', 'if ACTIVE_TIME > 0, then HAS_ACTIVE_TIME = 1, else HAS_ACTIVE_TIME = 0' and ' 'if MAINT_RESRC_TIME > 0, then HAS_MAINT_RESRC_TIME = 1, else HAS_MAINT_RESRC_TIME = 0'.

Next, the calendar hours which represent the time between the start and the end of data surveillance in hours. On the subsea application, calendar hours calculated on the different taxonomy level and the installation history. The different taxonomy level is on the equipment

Table 4.2: OREDA subsea automation opportunity

| Tables | Potential Automate Field | Reference Fields | |
|---|---|---|---|
| SSDAS_FAILURE_EVENT | AGE | FDATE | |
| SSDAS_FEMAINT_EVENT | HAS_DOWNTIME | DOWNTIME | |
| | HAS_ACTIVE_TIME | ACTIVE_TIME | |
| | HAS_MAINT_RESRC_TIME | MAINT_RESRC | |
| SSDAS_INSTALLATION_HISTORY | CALENDAR_HRS | I_START_DATE | I_END_DATE |
| SSDAS_INV_CO_HISTORY | CALENDAR_HRS | ICO_START_DATE | ICO_END_DATE |
| SSDAS_INV_CO_M02 | UOM | PROPERTY | |
| SSDAS_INV_EC_HISTORY | CALENDAR_HRS | IEC_START_DATE | IEC_END_DATE |
| SSDAS_INV_EC_M02 | UOM | PROPERTY | |
| SSDAS_INV_SU_HISTORY | CALENDAR_HRS | ISU_START_DATE | ISU_END_DATE |
| SSDAS_INV_SU_M02 | UOM | PROPERTY | |

class, the subunit, and the component level. They are calculated in the same way but using different terms.

$$CALENDAR\_HRS_{installation} = ((I\_START\_DATE - I\_END\_DATE) + 1) * 24 HOURS \tag{4.4}$$

$$CALENDAR\_HRS_{component} = ((ICO\_START\_DATE - ICO\_END\_DATE) + 1) * 24 HOURS \tag{4.5}$$

$$CALENDAR\_HRS_{equipmentclass} = ((IEC\_START\_DATE - IEC\_END\_DATE) + 1) * 24 HOURS \tag{4.6}$$

$$CALENDAR\_HRS_{subunit} = ((ISU\_START\_DATE - ISU\_END\_DATE) + 1) * 24 HOURS \tag{4.7}$$

On the subsea application, the 'UOM' has the same definition as on the topside application. However, the 'UOM' on the subsea application is recorded on each different taxonomy level: the equipment class, the subunit and the component level. The information about the unit of measurement is gathered by finding the value on corresponding rows (Ms.Excel vlookup function) of 'subsea properties' table. This process is performed by taking the 'PROPERTY' field as the reference field. Each taxonomy level has its own properties table and 'property' field. The function should be applied on each properties table with the respective 'property' field. The properties table for the subsea application on various taxonomy level can be seen in Appendix B.

### 4.1.2 Potential Unnecessary Fields

The potential unnecessary fields are those fields which are not utilized by the data custodians. These fields are not stored in the final database. As discussed in Chapter 3.3.2.2, these fields are the potential deleted fields because they do not bring added-value on the final database. The tabulation of unnecessary fields which has the potential to be deleted are shown in Table 3.3 and Table 3.4.

Should be kept in mind that these lists of the unnecessary and empty fields are made based on the phase 12 database only. There is a chance that phase 12 data does not represent the whole database characteristic. There may be another reason why OREDA still ask their data collectors to fill these fields even though they are not further used. Otherwise, filling these fields is just a waste activity. Therefore, it is suggested to evaluate the unnecessary and empty fields. The fields are considered to be a strong candidate for deletion.

## 4.2 OREDA Quality Check Experience

OREDA has developed a document to ensure quality assurance and control the OREDA data (SINTEF, 2007). This document consists of guidelines and checklists applicable for the received data from the data collectors and for the merged database before data is to be released. The received database checklist document is split into topside and subsea application due to the difference in taxonomy, while the merged database checklist document is combined for both applications.

The current checklists consist of fields showing a list of checks equipment with the corresponding description on what to be checked. It has been experienced that the execution of the received database and merged database check activity consumed much time. The reason was due to the absent information of which table and fields to look into for each checklist item. A new extended version of the checklist has been made by introducing several new fields. They are ’Table’ field, ’Fields’ field and ’Automation Process Description’ field. The ’Table’ field shows the main table to be focused on, the ’fields’ field shows fields that are prone to the checking activity and the ’Automation Process Description’ field describes how the checklist item can be automatically executed. The extended checklists are shown in Appendix C.

The extended checklists consist of the quality assurance checklist on the received database - topside application, the quality assurance checklist on the received database - subsea application and the quality assurance checklist on the merged database. Each checklist has been featured with information of ’Comment Description’ and ’Automation Process Description’ field. The comment description presents the more comprehensive comment from existing checklist items, while the automation process description declares the suggestion on how to do the check automatically. The development of the automation process has been made with the standard functions used in Ms. Excel. Some of the functions are if, count, find & replace and conditional formatting function. The automation process descriptions have been suggested for all possible checklist items. The current OREDA data process has developed script software for automation on several checklist items. No automation process is proposed for such checklist items.

The main focuses have been put into some of the checklists items. They are identified in highlighted colors. The main focus is on the unclear checklists, the checklist items which applies on the other checklist and the checklist items which require extra attention with a quantitative indication. Investigation on the unclear checklists is necessary to give a complete understanding for the checklists user and to reduce resources on performing the quality check activity. Some of the checklists have been identified to be relocated. The reason is to get earlier feedback for erroneous and to shift some task which not requires the involvement of the data collectors. At last, The quantitative approach is essential for reliability analysis. The comment description of some checklist items has specified the degree of an acceptable level. However, none of those checklist items specify the numbers which considered as acceptable. In addition, there is an additional highlighted color to denote exceptions, for example, the missing checklist items on the received database - subsea application checklist and the irrelevant checklist item which proposed to be removed on the merged database checklist.

### 4.2.1 Quality Assurance of The Received Database - Topside Application

#### 4.2.1.1 The Unclear Checklists

The identification of the unclear checklists are made based on the clarity of the comment description. Some of the checklists require more explanation on how to conduct such check. Based on author experiences, following checklists are considered as not clear checklists: Checklist number 11, 12 & 24.

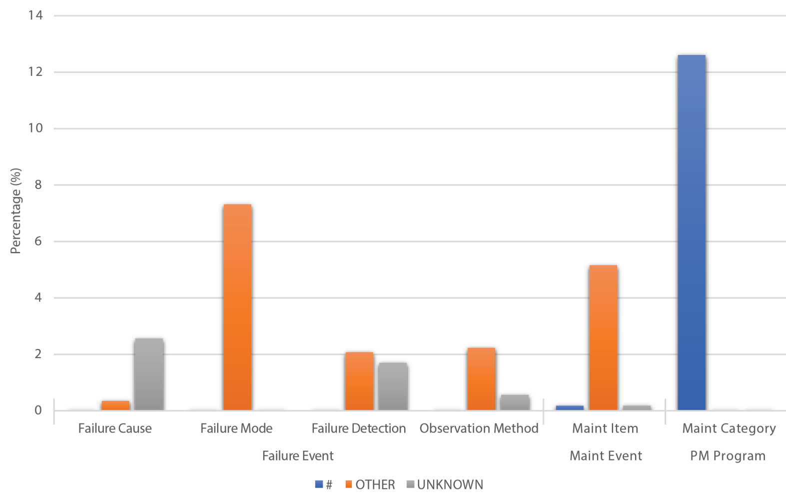
Checklist number 11 and 12 are applicable for compressors. Checklist number 11 asks to check if several compressors on one shaft driven by the same drivers, then it should be recorded as one equipment. Checklist number 12 requires checking on the number of casing correspond with the number of the installed item. Both the number of shaft and number of the casing are not collected in OREDA. There is no further explanation exist on how to do these checks.

Checklist number 24 asks to check the date of maintenance if the maintenance activity deferred. The information regarding to the maintenance date and the maintenance activity are collected as 'M_MAINT_DATE' and 'MAC_CODE.' However, the purpose of this particular check is not clear.

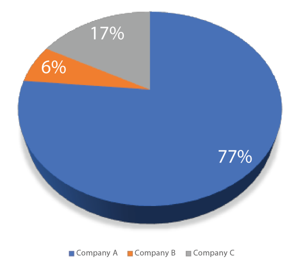
#### 4.2.1.2 The Checklist Which is Applicable on The Merged Database Checklist

One of the checklist on the quality assurance check asks for the anonymity check. The purpose of the anonymity check is to ensure that there is no company-specific information that is included. The main focus is on the 'additional information' field and the 'equipment name' field. This check can be performed during the quality assurance check on the merged database. An extra effort can be performed by reminding the data collectors to avoid the usage of any sensitive information to the company identity, for example, usage of the tag name which consist of plant or installation abbreviations on the 'additional information' field.

The anonymous records are also applicable for inventory tables on the installation, plant/unit and section level. This anonymity work is part of the data custodian duty. The anonymity is done by changing company's number with random number. Only the data custodian who know which number represent which company. It is better to not involve the data collectors at this stage so that they would not know their anonymity number. This anonymity check is not mentioned in the merged database checklist. Thus, author proposed to do this check on the merged database checklist.




(a) #, other and unknown fields distribution chart




(b) #, other and unknown fields contribution chart

Figure 4.2: OREDA topside: #, other and unknown fields

#### 4.2.1.3 The Checklists with Extra Attention on Quantitative Indication

Several researchers done their study based on failure data provided by OREDA. Spüntrup et al. (2018) in their recent study on improving the reliability of compressors based on asset data, put the OREDA database as their primary source of data. They discussed several observations experienced with OREDA database throughout their research. Their main concerns were the usage of 'unknown' as the most significant contributor to compressors failure and the existence of illogical combination fields.

The results of the reliability data analysis are sensitive to the documented failure event. The more 'unknown' data, the less accurate the reliability model. Meaning that there is no decision can be made based on reliability data analysis.

OREDA take into account this concern by having quality check on the non-informative codes (#, unknown, other) and the blank fields. These checks exist on topside QA checklist item number 30 & 31. Unfortunately, it is only stated that the percentage should be low without any quantification of how low is considered acceptable.

A review on the whole fields of topside data set has been conducted. The data combined from 3 different companies with more than 1000 failure events in total. This review aims to find the distribution of the non-informative code and the blank fields all over the fields. Those unnecessary fields listed in Table 3.3, which consist of blank fields that are excluded on this review. Besides, pie chart showing the contribution of deviations from each company is presented.

Figure 4.2a shows the overall percentage of the non-informative codes and the blank fields. In general, it can be observed that there is no existence of blank fields on the mandatory fields of the topside database. The absent of blank fields shows an indication of a good quality database. However, the existence of non-informative codes is variate. The highest percentage of other is in the 'failure mode' field with 7.3%, while the highest percentage on # is in the 'maint category' field with 12.6%. There is no justification for saying whether the quality of the topside database is adequate since the requirement only mentioned that the percentage should be low.

The non-informative contributions from each company can be seen in Figure 4.2b. From that figure, it reveals that company A dominates the contribution of non-informative code with 77%. This number of percentage is considerably high compared with the others company. This finding shows that in general, the ability of each company to collect data is not equal. This inequality may happen due to different maintenance personnel with different level of knowledge contribute to the OREDA data collection process. It is suggested to conduct a data collection training for company A for an improvement.

Spüntrup et al. (2018) found the combination of 'unknown' maintainable item and 'other' failure mode. This kind of combination is further called as the illogical combination. This combination is considered illogical because it is not possible to know the failure mode without knowing the failed item. There are various illogical combinations that may occur from the collected data.

OREDA take into account the concern of the illogical combinations, even though the current checklist does not accommodate all possible combinations. One of the illogical combination checks is stated on item number 20. The aim is to check the logic between 'severity class' and 'failure mode.' Check that for every 'failure mode' that by definition, are critical are coded with critical on 'severity class.' The justification for critical failure mode is regulated by DNV-GL (2016a) on failure mode table. Every type 1 failure mode is considered a critical failure mode. Type 1 failure mode is a failure mode that causes unobtained desired function. A review has been conducted on the topside database. As a result, no combinations at such found.

### 4.2.2 Quality Assurance of The Received Database - Subsea Application

#### 4.2.2.1 The Unclear Checklists

Similar to the topside application, the unclear checklists are identified based on the clarity of the check description. Based on author experience, following checklists are considered as not clear checklists: Checklist number 16, 23, 24, 25, 27.

Checklist number 16 examines that all the failure registered are within the boundary. There is no further explanation on the term boundary. This boundary may be interpreted as a the system boundary or the surveillance period boundary. In case of system boundary, if the item recorded is outside the system boundary than it should be excluded. in case of surveillance period boundary, the item recorded should be within the surveillance period. More detail information on the term 'boundary' should be provided.

A specific check on the control system is regulated in checklist number 23 & 24. It is mentioned that there should not be combination of 'Equipment Class = CS (Control System' with 'Component = SUTU (Subsea Umbilical Termination Unit)' nor 'Component = SC (Subsea Cabling)' in the data set, while in new taxonomy version, this combination exists. They are conflicted. As the checklists need to be updated on an ongoing basis. There might be a need to update the checklist according to the new taxonomy. It is also stated that 'Component = PR_TE_SENS (Combined Pressure and Temperature Sensors)' should be recorded as one because the temperature is used to compensate the pressure. As they are both combined and introduced as one field, there is no possibility to get multiple values. There might be another interpretation on this particular check. A better explanation should be provided.

A check on the 'right' number of couplers/connector and valve for the control system and the x-mas tree is mentioned on checklist number 25. There is no explanation on how to justify the 'right' number. More detail instruction also required for checklist number 27, which check the hierarchy structure according to the guideline. This description is too general.

In addition to the unclear checklists, there are some missing checklists on this checklist. The checklists items have been spotted. They are checklist number 4, 5, 8 & 9. There is no explanation whether these numbers has been intentionally removed or not.

#### 4.2.2.2 The Checklists Which are Applicable on The Merged Checklist

The anonymity requirement is also applicable on the subsea application. There is a specific check which regulates this requirement. Similar to the topside application, the anonymity for various taxonomy level in inventory tables is suggested to be implemented on the merged database checklist.

Besides, there is an examination of the data collection status on the final quality assurance, which should be on 'completed' status. This examination is considered too early to be implemented in the received database stage. The final quality assurance is held during the merged database check. Thus, this check is suggested to be moved to the merged database checklist.

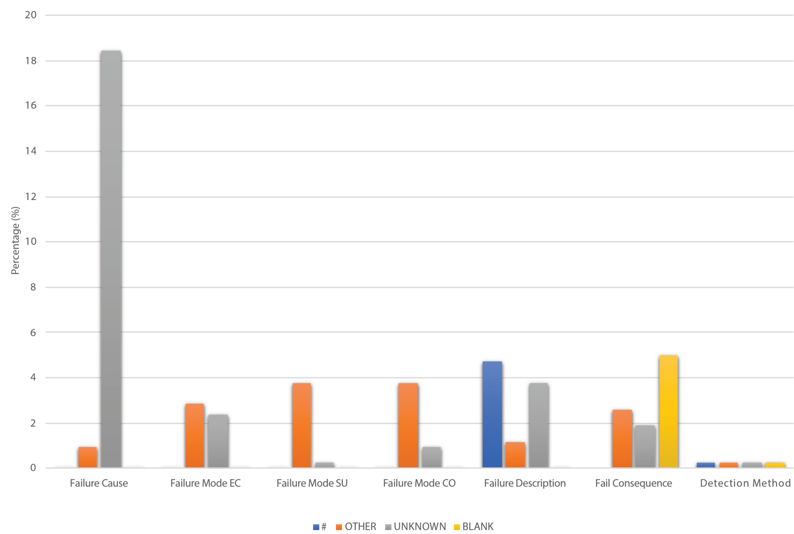
#### 4.2.2.3 The Checklists with Extra Attention on Quantitative Indication

The quality check on the non-informative code and the blank fields also being implemented on the whole fields of the subsea application. These checks exist on subsea QA checklist item number 11 & 20.

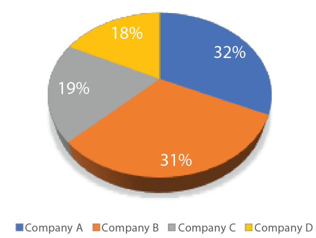
A review on the subsea data set has been conducted. The data combined from 4 different companies with almost 1000 failure events in total. This review aims to find the distribution of the non-informative code and the blank fields all over the fields. Those unnecessary fields listed in Table 3.4, which consist of blank fields that are excluded on this review. Besides, pie charts showing the contribution of deviations from each company are presented.

Unlike the topside application, the result in the subsea application can not be fully comparable. The reason is due to the existence of two companies which are not deliver the full complete tables. It is identified that DCS_INV_CO_HISTORY and DCS_INV_SU_HISTORY tables are missing for company A and company D. To make meaningful statistic analysis, the data is split into each database: failure data, inventory data and maintenance data.

In general, the completeness of data on the subsea application is lower than the topside application. Besides the missing tables, the subsea application also has empty tables. They are




(a) #, other and unknown fields distribution chart




(b) #, other and unknown fields contribution chart

Figure 4.3: OREDA subsea failure event: #, other and unknown fields

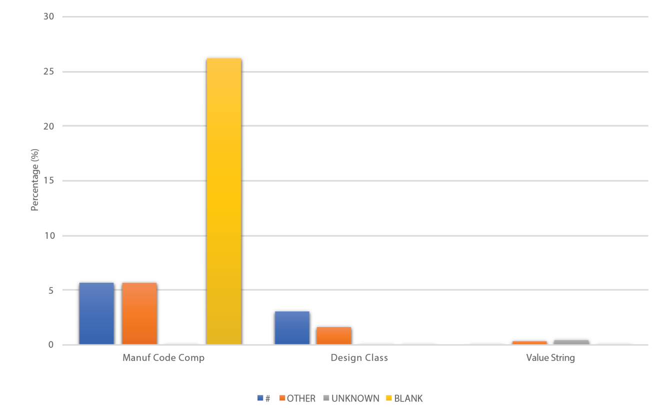
DCS_INSTALLATION_M02 and DCS_PMMAINT_EVENT tables. There is no explanation of why these tables kept blank. None of the companies filled any information in this table. Reevaluation on the function of these tables is proposed. Perhaps the tables can be skipped in the future.

Figure 4.3 shows the distribution of the non-informative and the blank data on the subsea failure event. The highest peak belongs to the unknown in 'failure cause' field with 18.5 % and the highest percentage on the blank cell is in the 'fail consequence' field with 5%, while the others are below 5%. The contribution from each company for the non-informative and the blank data on the failure event can be seen from Figure 4.3a. It can be observed that the distribution is somewhat equal from all companies.

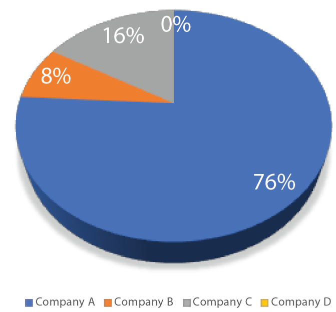
Figure 4.4 shows the distribution and contribution chart of the non-informative and the blank data on the subsea inventory data. The subsea inventory is split into equipment class, subunit, and component. This separation is made based on the subsea taxonomy. The missing tables are part of the equipment class and the component inventory database. However, the distribution and contribution chart is developed based on the remaining tables. Therefore, the equipment class and component chart are not able to adequately represent their database.

Based on the OREDA completeness requirement discussed in Chapter 2.3.2.4, the 'manufact' field in inventory component is observed violating the requirement. This field is marked as a highly desirable field. In actual, it has 99% of blank cells. This deviation is bizarre. There might be a possibility that the information of the completeness requirement has not transferred well to the data collectors. Some of the others blank cells are observed to have a high percentage. However, they can be neglected because they only have desirable completeness requirement.

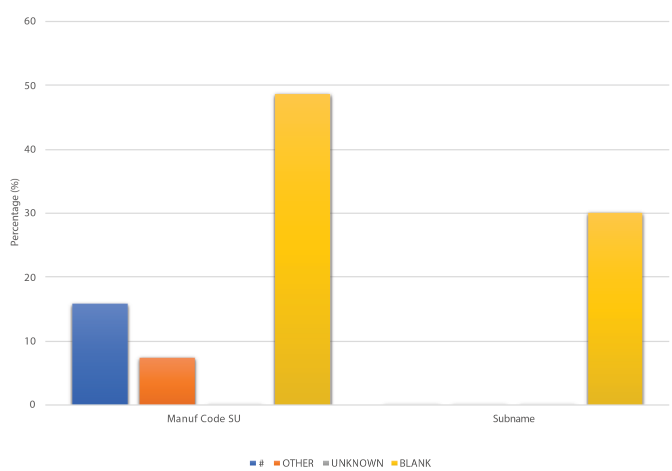
Figure 4.4b, Figure 4.4d, and Figure 4.4f present that there is significant contribution from company A on inventory equipment class with 76% and company C on subunit with 46%. How-



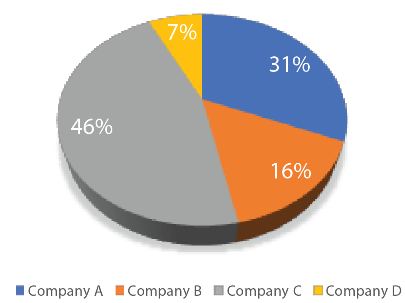

(a) inventory equipment class: #, other and unknown fields distribution chart



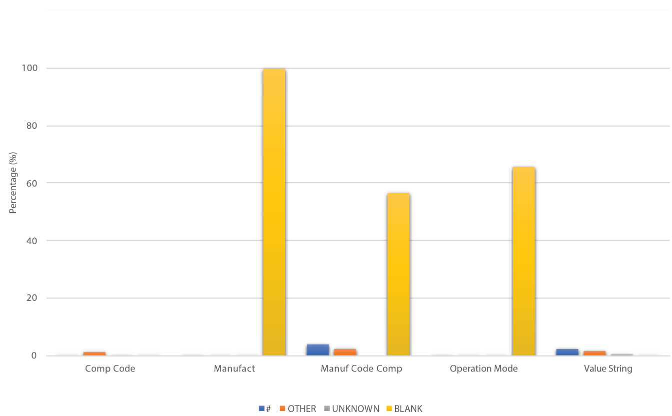

(b) inventory equipment class: #, other and unknown fields contribution chart



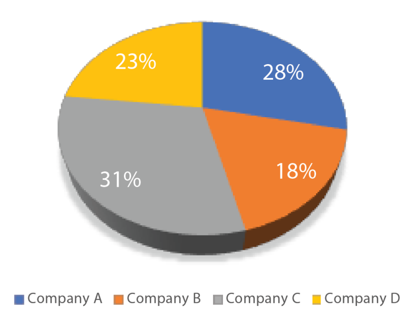

(c) inventory subunit:#, other and unknown fields distribution chart




(d) inventory subunit: #, other and unknown fields contribution chart

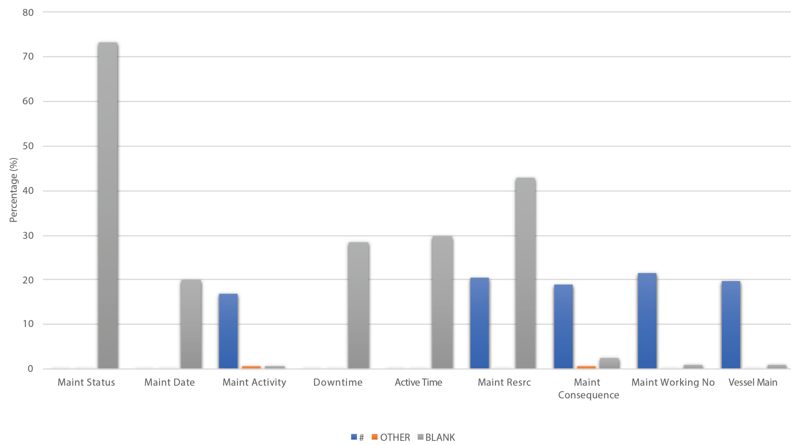



(e) inventory component: #, other and unknown fields distribution chart

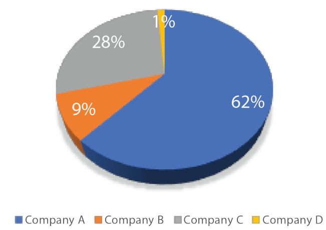



(f) inventory component: #, other and unknown fields contribution chart

Figure 4.4: OREDA subsea inventory: #, other and unknown fields




(a) #, other and unknown fields distribution chart




(b) #, other and unknown fields contribution chart

Figure 4.5: OREDA subsea maintenance event: #, other and unknown fields

ever, investigation and follow up are proposed for company A and company D, which have 2 missing tables. Overall, both companies indicate low performance in the collection of inventory data.

Distribution and contribution of the non-informative and the blank data on the maintenance event can be seen in Figure 4.5. The distribution chart shows shallow usage of 'other,' moderate usage of '#' for some of the tables with a percentage around 20% and relatively high usage of 'blank' with the highest percentage on 73%. The 'blank' percentage is acceptable because the respective fields are considered as a desirable field only. On the other hand, the 'other' percentage can be questioned since (again) there are no acceptable quantitative indication for the non-informative code. The contribution chart shows a significant contribution from company A with 62%. Necessary action is considerably required to motivate company A collect higher quality of data.

### 4.2.3 Quality Assurance of The Merged Database

#### 4.2.3.1 The Unclear Checklists

In the merged database checklist, there are several checks which have an unclear description and irrelevant for the future check. They are checklist number 4,5,8,10,11 & 16.6.

It is mentioned that checklist number 4 and 8 are relevant only for phase IV data. Furthermore, checklist number 8 stated the name of company members on the description. This statement should be avoided and keep the anonymity of company members. Checklist number 5 asks to manually update and fill the inventory data if the data collected before specific software revision. There was no information about the current software revision. It can be argued whether this check is still applicable in the current OREDA phase as the checklist was prepared in 2007. The OREDA phase 9 was established reliability data between 2006 and 2008, so this checklist might be applicable for OREDA phase 9 and below. These checklists are irrelevant and not applicable anymore. it is suggested to remove such checklist items from the overall checklist.

The description on checklist number 10 is not complete and checklist number 11 asks to insert 'figure 1' in specific condition without any explanation on the figure. Although the check descriptions are unclear, the execution of these checklists is automatic. Both checklists are executed with the help of specific database scripts developed by OREDA. However, a clear understanding of the purpose and expected result is necessary.

Checklist number 16.6 asks to check whether the operational time is equal for driver and driven unit when the surveillance period is the same. In the collected data, there is no information on the driver and the driven unit. More explanation is needed on how to proceed with this check.

Overall, the checklists have to be evaluated and updated frequently. The term updated is not only related to add more item to be checked, but also to delete the irrelevant item. Delete the irrelevant item reduce time and effort in the future quality assurance check. To check the relevancy of the checklist items, a new column can be introduced to indicate when the checklist item is made or represent affected phases. This column can be use to determine whether the checklist item is still relevant for the on going or the future check.

#### 4.2.3.2 The Checklists Which are Applicable on The Received Database Checklist

The role of the merged database checklist is a final checklist of the database before it is released. It consists of checklist items which catch all the discrepancies and improve the quality of the database. However, some of the check items are proposed to be implemented on the received database as well. The idea is to be able to detect the discrepancy as early as possible. By doing this, the discrepancies could be detected on an early stage and immediate action can be done. The checklist items to be implemented on the received database are those checklist items which has relation to the data collectors activity. It takes much time and works to involve data collectors in the the merged database checklist stage. These followings checklist items are proposed to be implemented on the received database check: Checklist number 3, 13, 14, 15, 16.01, 16.02, 16.03, 16.04, 16.05.

Checklist number 3 asks to change the unit of measurement of particular equipment, and checklist number 15 asks to check the incorrectly spelled codes. Both checks can be done on the received database check. It is better to capture the deviations on the earliest stage so that the data collectors are able to ask the data collectors to fix it.

Checklist number 13 examines the calculation of total manhours and checklist number 14 examines the spelling consistency of the 'manufacturer' and the 'model' fields. OREDA has developed a script to execute both checks automatically. If there is some discrepancy revealed during the check, it is better to have a clarification from data collectors who collect those data.

Checklist 16.01 - 16.05 concerns on the data records outside allowed dates. They are consist of several checks related to surveillance period and operational period. The intention is to ensure that there is no invalid combination event outside those periods. Some of these checks also mentioned on the received database check already. Please refer to Appendix C.0.1 and Appendix C.0.2. It is better to have some repetition on the manually conducted check. The repetition activity on the manual check is intended to verify the database.

# Chapter 5

# Challenges on OREDA's Data Collection Process

This chapter discusses the challenges which emerged from findings of the principal issues during the literature review. One of the most significant current discussion in reliability database is how to improve the data quality. In the recent years, there has been increasing interest in discovering an automation opportunity to improve the data quality. Another issue that has dominated the field for many years concerns in environmental condition data as data to be collected. These findings have a number of important implications and insights for the future research.

## 5.1 Increase The OREDA Data Quality

The collected reliability data quality playing a significant role in the reliability analysis for decision making. Langseth et al. (1998) analyzed the OREDA data for the maintenance optimization on the gas turbine. They utilized the collected reliability data for improving the existing facility operation. Besides, they mentioned the importance of the OREDA for development of the new oil fields, safer operation and increased production availability. Koronios et al. (2005) studied the critical impact of the data quality for asset management. The collected data was used to develop a picture of asset health and performance. This picture gave information for the decision making about asset replacement and repairs.

To achieve higher data quality, OREDA has developed the quality assurance procedure and checklist. In addition to that, several improvement strategies may be implemented. Researchers from various industries have implemented these strategies.

Haegemans et al. (2018) evaluated the initiatives to improve manually acquired data. Most organizations did the data cleaning intending to eliminate the deficiencies that reside in the database. The common data cleaning activity consists of identifying incorrect data and transform the database to eliminate the error. Bunea et al. (2008) in military system data, found that cleaning data was a time-consuming activity. Their experience showed that 2/3 of the time in data analysis was spent on extracting relevant information in the reliability database. However, data cleaning strategy may not be sufficient for an extended period because there is still a possibility that new erroneous data are inputted into the database. Data cleaning does not eliminate the root cause. Haegemans et al. (2018) discovered that one of the root cause was unmotivated data collectors who do not enter the data correctly. In the newest publication Haegemans et al. (2019) formulated a framework to find the causes of errors in the manually acquired data. They concluded that the main cause was the weak intention of the data collectors and low degree of understanding between the task and the technology.

### 5.1.1 Studies on The Data Collectors

Much attention has to be put on the data collectors side. Molina et al. (2013) discussed the external factors that might be influencing the data collector's performance. They conducted surveys to analyze the contribution of the managerial pressure, technological control and the intrinsic motivation to the data collector's performance. The supervisor pressure increased the performance of data collector with high intrinsic motivation but decreased the performance of the data collector with low intrinsic motivation, similarly with the technology control. Most of the data collectors felt the usage of input control system implementation to ensure the data quality, while there was still data collectors who think that using the input control system took more extended time than writing information down. Another researcher, Murphy (2009), tried to apply the theory of the planned behavior to the data quality. In this research, they evaluated the effect of the human factor, the time pressure, and the data collector's feedback. It was concluded that the consideration of the data collector's feedback was necessary. Nevertheless, the reasonable time pressures to achieve the compliance procedure should be evaluated. In general, researchers found that these factors have a different impact on different data collectors with various motivations.

From the psychological approach, Unsworth et al. (2011) mentioned that improving motivation can be done by setting a different level of goals (short-term goals, long-term goals). Lee and Strong (2003) stated that improving the knowledge of the data collectors on knowing-what, knowing-how and knowing-why was critical. Knowing what is understanding of all activities involved on the data collection process. Knowing-how understands the procedure on how to handle the data. Knowing-why is the ability to analyze based on the understanding of underlying principles and discover the solutions. Knowing-why is gained from experience, understanding the objective and learning from the execution of the procedure. In other words, it is important that the data collectors understand why they need to collect the data and how it will be utilized. By knowing why, what and how their data is used for, the data collectors may appreciate the process and have high motivation to collect a good quality of data.

### 5.1.2 The Implications for OREDA

As an implication, OREDA has to put extra attention on their data collectors. The attention must be focused on the influence factors and the basic knowledge on knowing-why at the minimum.

It is objectively stated in the ISO14224:2016 that the parties involved with the collection and exchange of the reliability data in the oil & gas industry must : "Train, motivate and organize the data collection personnel, e.g. interpretation of sources, equipment know-how, software tools, involvement of operating personnel and equipment experts, understanding/experience in analysis application of reliability data. Ensure that they have an in-depth understanding of the equipment, its operating conditions, this International Standard and the requirements given for data quality." Even though the regulation mentioned on the qualification of the data collectors, improving the data quality still become a challenge for OREDA.

## 5.2 Automation Opportunity in Data Collection

Automation can be one of the solutions to reduce the erroneous occurs during the data collection process. Automation can be employed for several reasons: reducing workload, replacement of human limitation work, and reducing cost (Wickens et al., 2015). Sharma (2016) emphasized the benefit of automation for consistency, quality, and cost-effectiveness. However, all those benefits may be achieved if the automation system is in good design, and the human operator has adequate knowledge on how to utilize it (Parasuraman and Manzey, 2010).

### 5.2.1 Studies on The Drop-down Boxes

For reliability variables, Hodkiewicz et al. (2006) suggested the development of a computerized maintenance management system (CMMS) that use codify maintenance and operational data to store it into the database. In the system, list and drop-down boxes with predetermined fields are necessary to be developed. The drop-down list suggests the data collectors to fill the field based on the available list. The downside of drop-down boxes is that short description is presented to the data collector. This condition is resulting in an increased incidence of inputting wrong code or excessive use of 'other' field. It is recommended that the data collector supplied with a rich description of available options.

A suggested way of developing the list and drop-down boxes is direct involvement between all roles within the data processes. The collaboration of all parties is necessary to avoid different interpretations of the same object or event. If the lists are too detailed, the data collectors may get frustrated trying to find the correct item that represents the actual object or event. Conversely, if the lists are too generic, the data consumers may get frustrated that the item selected does not adequately reflect their need. It also means that none of the items in the list represent how the work is done or the item fails. It is necessary that all parties have the same understanding of the wording to describe the object and the event. The system should be made informative yet user-friendly which make efficient and easy input of the data, even for those with the limited computer experience.

In their comprehensive survey on the data collectors, Molina et al. (2013) found that there was still some cynicism and resistance to the implementation of automation. These group of data collectors felt that the improvement in technology was restrictive and slowing them down. Adequate training on the guideline and knowing-why knowledge are essential for such issue.

### 5.2.2 The Implications for OREDA

OREDA has developed a semi-automatic system on their data collection. A drop-down list has been developed with the indication of completeness requirement and description of each item on the guideline. It can be suggested to include the item description on the drop-down list to help data collectors with a decent understanding of the available options. A 'question marked' button can also be developed as an additional feature for the drop-down boxes. This button may help the data collectors with some guidance explanation. Furthermore, control of the usage of 'other,' 'unknown' and blank is necessary to be able to provide meaningful data for the reliability analysis.

OREDA requires many fields to fill. Data collectors do not necessarily fill some of the fields because the value can be gathered from the others collected data. For instance, the 'surveillance period' field, which can be calculated based on information of 'start surveillance date' field and 'end surveillance date' field. Thus, the work of the data collectors is reduced and bring the opportunity for an automation process. Data custodians have to develop a script or a formula to accommodate these automation process. The idea to formulate the automation from others collected data has been discussed in Chapter 4.1.1.

## 5.3 Proposed Collected Data: Environmental Condition Data

Over the past decade,Barabadi and Markeset (2011) have seen the need for sufficient data in information about the operational and the environmental condition. They pointed out the different challenges in the Arctic region from the production performance point of view. Arctic operational conditions can change the reliability of the equipment dramatically. Currently, available data on oil & gas facilities only concern on the normal-climate region. These data cannot be used in the Arctic area, as they are not represent the effect of the severe weather condition. They introduced the dependability concept, which refers to a collective term that describes the availability performance and influencing factor. Influencing factor defined as all operational condition in which the failure of the system has occurred, and the repair process has been performed.

Moreover, Barabadi (2014) proposed a new model for the reliability analysis in the Arctic environment. They introduced an accelerated life testing model where the collected data of high stress (co-variate) was used to predict the behavior of the system in the normal condition. The main idea was to use the available data and extrapolate them. Two similar equipment with the same age can have different performance if they have been operated in the different load stress level and environmental conditions. Data available in OREDA were treated as a normal condition, while Arctic condition considered as a high-stress condition. The aim was to predict the reliability of the item under the high-stress condition. Meaning that the model was used inversely compare with the way it was normally applied. Okaro and Tao (2016) also implement the accelerated life testing to forecast the behavior of the subsea system. They revealed that subsea system components failed faster than what has been specified by OREDA. That was happened due to the accumulative marines stress. They implemented the model by taking the environment condition and the loading as their influence factors. Another researcher, Naseri et al. (2016), introduced a weather-dependent multiplicative model which were developed from the expert judgment. They tried to forecast the system availability by taking into account the harsh operation condition and the maintenance duration. All of these researchers tried to model their reliability assessment due to the absence of the operation conditions.

In another publication, Barabadi et al. (2015) emphasized the urgency of the data which reflect the actual conditions of what equipment has experienced during its operation time. Lack of sufficient data and information about the operational and environmental condition may result in inaccurate analysis and increase the uncertainty. They criticized the lack of attention for collecting the effect of influence factors. They emphasized the benefit of collecting such influence factors for the design and operation of the system in different types of the operational environment. They demonstrated how the influence factors on the time between failures and the time to repair affected the analysis. They proposed to collect influences factors such as the surrounding environment(e.g., environmental temperature, wave, humidity, icing), human aspects (e.g., the skill of operator and maintenance crew), and history of repair (e.g., number of repair and equipment condition after repair). The Arctic operational conditions represent the harsh environmental condition, have a significant effect on the RAMS characteristic. Hence, these data are suggested to be collected to improve the RAMS analysis.

Current OREDA installation database has collected information on geographic location. This information may be used to forecast the environment condition of such installation. However, the exact condition and related influencing factors are not collected. Further analysis on the trade-off between the difficulty to collect these data, the impact on reliability analysis and the economic aspect is required.

# Chapter 6

# Summary and Recommendation for Further Work

The final chapter draws upon the entire thesis, tying up the analysis of the OREDA quality assurance and OREDA database in order to improve the OREDA database quality and to improve the automation process. The purpose of this chapter is to present a summary of the project. A recommendations related to current OREDA quality assurance and checklists are proposed. A discussion regarding the author's experience in dealing with phase 12 database is presented. Finally, areas for further research are identified.

## 6.1 Summary and Conclusion

An investigation of current data collection has been conducted throughout this master thesis. The investigation has been performed during the execution of the phase 12 OREDA database with the supplement of supporting documents such as guidelines, checklists and papers. As the results, recommendations to improve the database quality and the automation on data process are proposed.

### 6.1.1 Recommendations to Improve The Database Quality

In respect to the reliability data quality framework, the current OREDA database indicates minor attention on the timeliness and value-added attributes. Rise the awareness of the timeliness attributes can be done by removing the out-of-date data based on the fundamental technology principal approach. The development of the fundamental technology principal changes the failure behavior. Nevertheless, the dominance of the systematic failure from the overall failure, supporting this approach. On the value-added attribute, the list of the potential unnecessary and empty fields from both the topside and the subsea applications have been presented. These lists are subject to be removed from the data collectors fields since they are not populated with the adequate number of reliability data and no longer used on the data custodians fields. The results of value-added attribute analysis are shown in Table 3.3 and Table 3.4.

In respect to the quality assurance protocol, modification on existing quality assurance protocol is performed. The modified protocol draws all possible transitions from the involved actors. The protocol describes detail activities in accordance with the quality assurance on the OREDA guideline. The modified quality assurance protocol can be seen in Figure 3.3.

More concentration has to be put for the data collectors to collect a good quality of data. There are external factors such as managerial pressure, technology and intrinsic motivation which influence their performance. Among those factors, motivations play a significant role. Motivate means to educate them with the ability to understand the reason why, what and how the data will be utilized for the reliability analysis. Data collectors at such give useful feedback and involvement to the whole data process.

One of the results of this study indicates relatively unequal competency of the data collectors from various company members. This result is concluded based on the evaluation of the non-informative and blank fields of the phase 12 database. It is observed on the subsea application that some company members did not deliver complete required tables. Further more, some tables were kept empty. The discussion of the results are presented in Chapter 4.2.1.3 and Chapter 4.2.2.3. It is recommended to conduct refreshment training for the data collectors to refresh their knowledge. In addition, analysis on the illogical combination between failure mode and severity class has been performed. It is concluded that the OREDA phase 12 database is free from such illogical combination.

### 6.1.2 Recommendations to Improve The Automation on Data Process

In the data collection process, OREDA has implemented the codified data and the drop-down list. It is necessary to ensure that all involved actors have the same understanding of the coded data. The system shall be user-friendly and informative. For the drop-down list, it is recommended to provide a rich description of available options and some guidance explanation from 'help' or 'question marked' button.

One of the tasks with the phase 12 database is converting the data collector's tables into the data custodian's tables so that they are ready to be merged. During the conversion activity, it is found that the fields of their tables are different. Converting the database requires field mapping. Since they are different, it needs to be done manually. It is recommended to have the same fields template for both tables to avoid the miss-match mapping and bring opportunity to do it automatically.

The missing fields which are the requirement field from the data custodian side may be generated from other fields through a calculation or further interpretation. These fields are the potential fields for the automation process. The results of these potential fields are presented in Table 4.1 and Table 4.2. The explanations on how to make these fields automated are explained in Chapter 4.1.1.1 and Chapter 4.1.1.2.

The automation process is also expected on the quality assurance checklists. The result of this study is an extended quality assurance checklists, which has the information of specific table and fields to refer to, and automation process description to demonstrate how to do checklist items automatically. The extended quality assurance checklists are shown in Appendix C. Unfortunately, there are some issues in the current checklists related to the unclear checklists and the checklists which suppose to be moved to the other checklist. These issues are discussed in Chapter 4.2.1.1 , 4.2.1.2 , 4.2.2.1 & 4.2.2.2. It is recommended to clarify these issues so that the analysis of the automation opportunity for quality assurance checklists can be performed completely.

Overall, the study of the automation process has the intention to increase the quality of the OREDA reliability database. Automation also brings the opportunity to reduce the consumed time and resources.

## 6.2 Discussion

Initial objectives of the master thesis are to establish data elements which are vital for reliability study and which are not (objective 1), to improve the quality check procedure of the received data (objective 2), to improve the quality check procedure of the merged database (objective 3), to investigate the possibility for automatic data process (objective 4) and to verify the automatic classification algorithms on the phase 12 database (objective 5). Please refer to Chapter 1.2 for more details.

The first objective of this master thesis has been met. The non-vital reliability data have been identified based on evaluation of value-added attributes and experience gained during the database format conversion. This identification resulted in the list of potential unnecessary fields was presented. The vital reliability data have been identified based on the evaluation of timeliness attributes and study on related papers. This evaluation brought fundamental technology principal and environmental condition data as vital reliability that proposed to be collected.

The second objective and the third objective were related to the improvement of quality assurance. In general, the improvement of the quality assurance protocol has been fulfilled by modifying the current quality assurance protocol. The modification has been developed to make the quality assurance protocol in line with the OREDA guideline and ISO14224. In respect to the quality assurance on the received database and the merged database, a comprehensive analysis has been performed. The analysis investigated the clarity of the checklists instruction, the timing to distinguish whether to check the checked item on the received database or the merged database, and the existence of non-informative and blank fields.

The fourth objective has been achieved through analysis of the current data collection, experience gained during involvement in the phase 12 database and investigation on the existing quality assurance checklists. Analysis of the current data collection process with the support of the respective papers brought the result in the application of the operator's guidance. The experience gained during involvement in the phase 12 database gave the insight to manipulate the available fields to fill the missing fields. This manipulation consisted of calculation and interpretation. The investigation on the existing quality assurance checklist has been performed by implementation on the phase 12 database. The implementation identified which checklists were possible to be automated. A list of checklists with description on how to make execute it automatically has been generated as a result.

Unfortunately, this study has failed to meet the fifth objective. This master thesis was limited with the access to the new published database. There was a delay in the new database publication. This objective was intended to implement and verify the automatic classification algorithms on the new database. These algorithms were proposed by the author on his specialization project. Thus, current algorithms are still considered relevant.

## 6.3 Further Researches

There are still space for further research on the improvement of the reliability data quality and the automation process. In accordance with the master thesis's objectives, the results and the limitations, following further works have been formulated:

- The potential unnecessary fields which can be deleted was developed from the phase 12 database only. It can be argued whether this database can be fully representative of the entire OREDA database. An investigation of the existence of these fields on the complete database is required to confirm the vitality of these fields.

- This study proposed additional vital reliability data to be collected, which are the fundamental technology principal and the environmental condition data. These data are essential for the reliability analysis. However, a further feasibility study has to be conducted to evaluate the cost-benefit and the cost-effectiveness to collect such data.

- In the timeliness attribute analysis, it is proposed to delete out-of-date data based on their fundamental technology principal. The deletion of data may result in very few data which influence further reliability analysis. An implementation and verification of this proposal in the complete OREDA database is required.

- In the value-added attribute analysis, it is better to involve the data consumer's perspective. The user's feedback and experience are essential to determine the utilization of the collected data for their further data analysis. A study from the data consumers side is proposed.

- This study only did the analysis on the illogical combination between failure mode and severity class. However, there are various illogical combinations that could exist. To find out the other possible combinations, it is suggested to involve the perspective of the data consumers who use the data for reliability analysis. A further investigation on the other illogical combinations is recommended.

- Further analysis to collect the environmental condition data. Analyze the trade-off between the difficulty to collect the data, the impact on the reliability analysis and the economic aspect.

- Implement the classification algorithms to the other component on the new complete database. Analyze whether the proposed algorithms are suitable. During specialization project, the algorithms have been implemented to one component only. It can be questioned whether the algorithms are sufficient for the other component.

# Appendix A

# Unit of Measurement Property - Topside Application

| PROPERTY | UOM |
|---|---|
| PRIM_VOLT | kV |
| SEC_VOLT | kV |
| VOLT_WIND | kW |
| POWER | kW |
| TEMPR_RISE | oC |
| LEV_INSUL | kV |
| POWFACT | cos |
| FREQUENC | Hz |
| RATED_VOLTAGE | kV |
| RATED_CURRENT | A |
| WITHSTAND_CURREN | kA |
| RATED_DURATION | s |
| UTILCAP | % |
| FREQUENCY | Herz |
| VOLTAGE | V |
| POWERGEN | kVA |
| CAPACITY_UTIL | % |
| OP_FREQ_RANGE | Hertz |
| RATED_VOLTAGE | kV |
| VOLT_MAX | kV |
| CURR_RAT | Amper |
| FREQUENCY | Herz |
| SCC_COND | kA |
| SCC_SHEETS | kA |
| LEN_STATIC | m |
| LEN_DYNA | m |
| WATERDEPTH | m |
| SIZEHEIGHT | m |
| SIZEWIDTH | m |
| NUMBCYL | # |
| POWER | kW |
| POWERO | kW |
| SPEED | rpm |
| DISPRESD | barg |
| DISPRESO | barg |
| DISTEMPD | oC |
| DISTEMPO | oC |
| FLOWRD | Std. m3/h |
| FLOWRO | Std. m3/h |
| GASHANDLE | g/mole |
| NUMBCAS | # |
| NUMBSTAGE | # |
| POLYHEAD | barg |
| POWER | kW |
| SPEED | rpm |
| SUCPRESD | barg |
| SUCPRESO | barg |
| UTIL | % |
| AFRAHEIGHT | m |
| BOOMWEIGHT | mt |

| PROPERTY | UOM |
|---|---|
| HOISTSPEEDMA | m/s |
| HOISTSPEEDMI | m/s |
| HYDCONTRPR | barg |
| MAXANGLE | deg |
| MAXLOAD | mt |
| MAXRADIUS | deg |
| MBOOMLENGTH | m |
| MINAGLE | deg |
| OVHEIGHT | m |
| PEDWEIGHT | mt |
| SLEWSPEEDMA | m/s |
| SLEWSPEEDMI | m/s |
| UNITWEIGHT | mt |
| CTRLFLHP | psi |
| CTRLFLLP | psi |
| CTRLFLVHP | psi |
| CTRLFUNC | # |
| CTRLPOD | # |
| ELPODCONC | # |
| ELPODCONI | # |
| HYDRLP | psi |
| NUMBSIGN | # |
| UMBDIAM | inches |
| UMBLEN | m |
| WATERDEPTH | m |
| DESIGNWAVE | m |
| NUMANCHOR | # |
| NUMRISER | # |
| NUMUMBILICAL | # |
| VESSELDEPLA | mt |
| FREQU | Hz |
| POWERGEN | kVa |
| POWERFACT | cos |
| SYNCSPEED | rpm |
| TEMPRISER | oC |
| TEMPRISES | oC |
| VOLTAGE | V |
| POWER | kW |
| POWERO | kW |
| SPEED | rpm |
| VOLTAGE | V |
| NUMBSHAFT | # |
| POWER | kW |
| POWERO | kW |
| SPEED | rpm |
| INTEMP | oC |
| NOBURN | # |
| NOTUBES | # |
| OUTTEMP | oC |
| RATEHTREL | Kw |

| PROPERTY | UOM |
|---|---|
| SIZEDIAM | m |
| SIZELEN | m |
| HEAT_AREA | m2 |
| NUMBTUBE | # |
| PRESC | barg |
| PRESH | barg |
| RATEHEATT | kW |
| SIZEDIAM | m |
| SIZELEN | m |
| TEMPDROPH | oC |
| TEMPRISEC | oC |
| UTIL | % |
| ARM_DIAM | inches |
| MAX_DES_TEMP | oC |
| MIN_DES_TEMP | oC |
| PRESDPSI | psi |
| CONNSIZE | inches |
| DISTEMPO | oC |
| FLOWRATE | l/min |
| FLOWPR_SUBS | psi |
| FLUIDTEMP | oC |
| NOZLENGTH | mm |
| NOZWIDTH | mm |
| SPRAYANGLE | deg |
| FLOWRO | Std. m3/h |
| LINELEN | km |
| MAXWPRES | psi |
| PIPEDIAM | inches |
| WALLTHICKNESS | mm |
| WATERDEPTH | mm |
| FLOWRO | Std. m3/h |
| PRESDPSI | psi |
| PRESOPSI | psi |
| TEMPDROPH | oC |
| RISEDIAM | inches |
| TEMPO | oC |
| WATERDEPTH | m |
| PRESO | barg |
| TEMPO | oC |
| DESIGNWAVE | m |
| NUMANCHOR | # |
| NUMRISER | # |
| NUMUMBILICAL | # |
| VESSELDEPLA | mt |
| DISPRESD | barg |
| NUMBSTAGE | # |
| POWER | kW |
| SPEED | rpm |
| SUCPRESD | barg |
| UTILCAP | % |

| PROPERTY | UOM |
|---|---|
| NUMBSHAFT | # |
| POWER | kW |
| POWERO | kW |
| SPEED | rpm |
| D_TEMP | oC |
| ELPOW | kW |
| NOPATHS | # |
| PRESDPSI | psi |
| SAND_PROD | g/m3 |
| VOLSIGN | V |
| VOLTPOW | V |
| GASHANDLE | g/mole |
| INLFLOW | kg/h |
| INLPRES | barg |
| INLTEMP | oC |
| NUMBSTAGE | # |
| POWER | kW |
| POWERO | kW |
| SPEED | rpm |
| DIMLWH | m |
| FLOWPR_SUBS | psi |
| NUMB_MAV | # |
| NUMB_SL | # |
| NUMBLINEP | # |
| PRESDPSI | psi |
| PROD_GAS | scm/day |
| PROD_LIQ | scm/day |
| AMBTEMP | oC |
| WATERDEPTH | m |
| BATTBACKUPT | min |
| FREQIN | Hz |
| FREQOUT | Hz |
| FREQVARI | % |
| RECHARGE | min |
| VOLTAGEIN | V |
| VOLTAGEOUT | V |
| VOLTVARI | % |
| FLOWPRES | barg |
| FLUIDTEMP | oC |
| SHUTTOFFP | barg |
| SIZEVALVE | inches |
| NUMBBRAN | # |
| PRESD | barg |
| PRESO | barg |
| SIZEDIAM | m |
| SIZELEN | m |
| TEMPD | oC |
| TEMPO | oC |
| N_AMV_MO | # |
| N_AMV_RO | # |

| PROPERTY | UOM |
|---|---|
| N_ASWAB_NO | # |
| N_ASWAB_RO | # |
| N_AWV_MO | # |
| N_AWV_RO | # |
| N_CHOK_MO | # |
| N_CHOK_RO | # |
| N_COV_MO | # |
| N_COV_RO | # |
| N_PMV_MO | # |
| N_PMV_RO | # |
| N_PSWAB_MO | # |
| N_PSWAB_RO | # |
| N_PWV_MO | # |
| N_PWV_RO | # |
| PROD_GAS | scm/day |
| PROD_LIQ | scm/day |
| TUBHB | inches |
| WATERDEPTH | m |
| WHFLOWPR | psi |
| WHFLOWTE | oC |
| DRUMCAP | m |
| DRUMDIA | m |
| DRUMNO | # |
| LINEDIA | mm |
| MAXCAP | mT |
| MAXPOWER | kW |
| SPEED | rpm |
| TEMPRISES | oC |
| TEMPRISER | oC |
| POWER | kW |
| POWERO | kW |
| SPEED | rpm |
| NUMBSHAFT | # |
| SWL | mt |
| MAX_SWING | deg |
| MAX_MOMENT | ton-met |
| NUMRISER | # |
| NUMUMBILICAL | # |
| NUMANCHOR | # |
| DESIGNWAVE | m |
| VESSELDEPLA | mt |
| RATE_OUT | kVA/cos |
| SIZEVALVE | inches |
| FLUIDTEMP | oC |
| FLOWPRES | barg |
| SHUTOFFP | barg |

# Appendix B

# Unit of Measurement Property - Subsea Application

**SSDAS_INV_CO_M02**

| PROPERTY | UOM |
|---|---|
| INPUT_VOLTAGE | kV |
| JMPR_LENGTH | m |
| LENGTH | km |
| MAX_OUT_VOLTAGE | kV |
| MIN_OUT_VOLTAGE | kV |
| NO_PINS | # |
| NO_SHELLS | # |
| OPERATIONAL_VOLT | kV |
| PIN_SIZE | inch |
| POWER | kVA |
| POWER_RATING | kV |
| SIZE | inch |
| SUBSEA_CA_LENGTH | m |
| VOLTAGE | kV |
| VOLTAGE_RATE_UM | kV |
| VOLTAGE_RATE_UO | kV |
| VOLTAGE_RATING | kV |
| VOLTAGE_TARE_U | kV |
| CAPACITY | kVA |
| CONN_BORES | # |
| CONN-BORES | # |
| CONN_PRESS_RATE | psi |
| CONN_SIZE | inch |
| CORES | # |
| DESIGN_PRESSURE | psi |
| DIST_RISER_BASE | m |

**SSDAS_INV_EC_M02**

| PROPERTY | UOM |
|---|---|
| FLOW_TEMP | oC |
| FLUID_PRES | psi |
| MF_DIMENSION | m (LxWxH) |
| MF_SLOTS | # |
| MF_WEIGTH | ton |
| NO_CONS | # |
| PIPE_DIAMETER | inch |
| PU-STAGES | # |
| PU_DISCH_PRESS | psi |
| PU_POWER | kW |
| PU_SPEED | rpm |
| PU_SUCT_PRESS | psi |
| RETENTION_TIME | min |
| RISER_LENGTH | m |
| SHUT_OFF_PRESS | psi |
| SIZE_DIAMETER | m |
| SIZE_LENGTH | m |
| TRANS_DIST | m |
| TRANS_POWER | kVA |
| TRANS_VOLTAGE | kV |
| WALL_THICKNESS | mm |
| WATER_DEPTH | m |
| WORK_PRESSURE | psi |
| XT_ANNBORE_SIZE | inch |
| XT_PRODBORE_SIZE | inch |
| NUMBER_CONN | number |
| X-TREE-WEIGHT | ton |
| X-TREE-HEIGHT | m |
| DESIGN_PRESS_VES | barg |
| DESIGN_PRESSURE | psi |
| DESIGN_TEMP | oC |
| DESIGN_THROUGH | Sm3/d |
| DIMENSION L_W_H | mm (LxWxH) |
| DRY WEIGHT | kg |
| FLOW_LENGTH | km |
| FLOW_PRESS | barg |
| FLUID_TEMP | oC |

**SSDAS_INV_SU_M02**

| PROPERTY | UOM |
|---|---|
| NO_SEM | # |
| SCM_OUTPUT | # |
| UMB_DIAMETER | inch |
| UMB_HIGH_PRESS | psi |
| UMB_HIGH_VOLTAGE | V |
| UMB_LENGTH | m |
| UMB_LOW_PRESS | psi |
| UMB_LOW_VOLTAGE | V |
| WH_DES_PRESS | psi |
| WH_DES_TEMP | oC |
| WH_SIZE | inch |
| XT_ANNBORE_SIZE | inch |
| XT_DES_PRESS | psi |
| XT_DES_TEMP | oC |
| XT_PRODBORE_SIZE | inch |
| CABLE_DIAMETER | inch |
| CABLE_LENGTH | m |
| DESIGN_PRESSURE | psi |

# Appendix C

# Extended Quality Assurance Checklists

### C.0.1 Extended Quality Assurance Checklists on The Received Database - Topside Application

| NO | Equipment | Table | Fields | | | Comment Description | Automation Process Description |
|---|---|---|---|---|---|---|---|
| 1 | ALL | ALL TABLES | | | | Check completeness of data. Mandartoy = 100% , Desirable = 85%. | Use the COUNTBLANK function with the range as specified on the comment description. |
| 2 | ALL | DC_INVENTORY | EC_CODE | SYS_CODE | | Compare match combination of EC_CODE & SYS_CODE with DC_EC_SYSTEM from the taxonomy. | Use the conditional formatting function : create new rules for duplicate values. |
| 3 | ALL | DC_PM_PROGRAM | INTERVAL_UNIT | PM_TYPE | | Check if PM_TYPE = C (calendar time based), then INTERVAL_UNIT = M (months); if PM_TYPE = O (operational time based), then INTERVAL_UNIT = H (hours). | Use the IF function as stated on the comment description. |
| 4 | ALL | DC_MAINT_EVENT | M_TOTAL_MANHOUR | M_ACTIVE_MAINT | | Check if M_TOTAL_MANHOUR >= M_ACTIVE_MAINT. | Use the IF function as stated on the comment description. |
| 5 | VA | DC_INVENTORY | OP_CODE | | | Only 'OPEN' & 'CLOSED' are allowed. | Use the Find command as stated on the comment description. |
| 6 | VE, HE | DC_FAILURE_EVENT | SU_CODE | | | There is no inlet and outlet valves recorded. | N/A |
| 7 | CO,PU | DC_FAILURE_EVENT | SU_CODE | | | There is no suction and discharge valve recorded. | N/A |
| 8 | EG | DC_FAILURE_EVENT | SU_CODE | | | There is no circuit breaker recorded. | N/A |
| 9 | ALL | ALL TABLES | INST_ID | | | Check if I_ID <= 99,999. | Use the IF function as stated on the comment description. |
| 10 | ALL | ALL TABLES | OWNER_ID | INST_ID | | Change OWNER_ID & INST_ID to anonymous. | N/A |
| 11 | CO | | | | | NOT CLEAR. | N/A |
| 12 | CO | | | | | NOT CLEAR. | N/A |
| 13 | ALL | ALL TABLES | | | | Check no single quotes (') are applied. | Use the Find command as stated on the comment description. |
| 14 | ALL | DC_INVENTORY | I_MANUFACT | I_MODEL | | Check that the spelling is consistent. | N/A |
| 15 | ALL | DC_MAINT_EVENT | F_ID | M_ID | | Check that every F_ID has correspond M_ID. | Use the COUNTBLANK function, ensure no blank M_ID cells for every F_ID. |
| 16 | ALL | DC_MAINT_EVENT | MC_CODE | | | NOT CLEAR. | N/A |
| 17 | ALL | DC_INVENTORY<br>DC_INVENTORY_HISTORY | I_INSTALLED_DATE<br>I_SURV_START_DATE | | | Check if I_SURV_START_DATE >= I_INSTALLED_DATE . | Use the IF function as stated on the comment description. |
| 18 | ALL | DC_INVENTORY_HISTORY | I_SURV_START_DATE | I_SURV_END_DATE | | Check if I_SURV_END_DATE -I_SURV_START_DATE >= 2 years. | Use the IF function as stated on the comment description. |
| 19 | ALL | DC_INV_SUB_UNIT | ISU_NO_INSTALLED | | | Check that ISU_NO_INSTALLED are filled . | Use the COUNTBLANK function with the range as specified on the comment description. |
| 20 | ALL | DC_FAILURE_EVENT | SC_CODE | FM_CODE | | Check If Type of FM_CODE = 1, then SC_CODE = C (critical). | Use the IF function as stated on the comment description. |
| 21 | ALL | DC_INVENTORY<br>DC_FAILURE_EVENT<br>DC_MAINT_EVENT | ELC_STATUS_ID<br>ELC_STATUS_ID<br>ELC_STATUS_ID | | | Check If ELC_STATUS_ID = 3 (new code proposed), then write a change request. | Use the IF function as stated on the comment description with an indication asking to create a change report. |
| 22 | ALL | DC_MAINT_EVENT | M_DOWNTIME | M_ACTIVE_MAINT | | Check if M_DOWNTIME >= M_ACTIVE_MAINT. | Use the IF function as stated on the comment description. |
| 23 | ALL | DC_INVENTORY_HISTORY | I_SURV_START_DATE | I_SURV_END_DATE | I_OPER_TIME | Check if I_SURV_END_DATE -I_SURV_START_DATE >= I_OPER_TIME. | Use the IF function as stated on the comment description. |
| 24 | ALL | DC_MAINT_EVENT<br>DC_PM_PROGRAM | MAC_CODE<br>MAC_CODE | M_MAINT_DATE<br>M_MAINT_DATE | | Check if MAC_CODE = DEFFER, then check M_MAINT_DATE. | N/A |
| 25 | ALL | DC_INVENTORY_HISTORY | I_SURV_START_DATE | I_SURV_END_DATE | I_OPER_TIME | Check if I_SURV_END_DATE , I_SURV_START_DATE & I_OPER_TIME are continuation from previous collected data. | Use the IF function as stated on the comment description. |
| 26 | ALL | ALL TABLES | | | | Check that English is used in free text. | N/A |
| 27 | ALL | ALL TABLES | | | | Check that (,) used for high number and (.) use for decimal number. | N/A |
| 28 | ALL | DC_INVENTORY<br>DC_FAILURE_EVENT<br>DC_MAINT_EVENT | ELC_STATUS_ID<br>ELC_STATUS_ID<br>ELC_STATUS_ID | | | Check if ELC_STATUS_ID is not equal to 6 before QA process completed. | Use the IF function as stated on the comment description. |
| 29 | ALL | DC_FAILURE_EVENT<br>DC_MAINT_EVENT | SC_CODE<br>SC_CODE | | | Check that C, D, I distribution not so skewed compare to previous collected data. | N/A |
| 30 | ALL | ALL TABLES | | | | Percentage of code UNK and OTH should be low. | Use the COUNT function. The total count should not exceeding the acceptance limit. |
| 31 | ALL | ALL TABLES | | | | Check that there is no BLANK fields for drop-down code list. | Use the COUNTBLANK function with for selected fields. |
| 32 | ALL | ALL TABLES | | | | Check that there is no dummy records. | N/A |
| 33 | ALL | ALL TABLES | | | | Check that data collector have corrected their database based on the QA check form. | N/A |

= Unclear checklists.
= Applicable on the merged checklist.
= Extra attention on quantitative indication.

### C.0.2 Extended Quality Assurance Checklists on The Received Database - Subsea Application

| NO | Equipment | Table | Fields | | | Comment Description | Automation Process Description |
|---|---|---|---|---|---|---|---|
| 1 | ALL | DCS_INV_CO | EC_CODE | SUB_CODE | COMP_CODE | Compare match combination of EC_CODE, SUB_CODE & COMP_CODE with DCS_EC_SUB_COMP from the taxonomy. | Use the conditional formatting function : create new rules for duplicate values. |
| 2 | ALL | ALL TABLES | INST_ID | | | Check if INST_ID <= 99,999. | Use the IF function as stated on the comment description. |
| 3 | ALL | ALL TABLES | OWNER_ID | INST_ID | | Change OWNER_ID & INST_ID to anonymous. | N/A |
| 4 | MISSING | | | | | | |
| 5 | MISSING | | | | | | |
| 6 | ALL | ALL TABLES | | | | Check no single quotes (') are applied. | Use the Find command as stated on the comment description. |
| 7 | ALL | DCS_INV_EC<br>DCS_INV_SU<br>DCS_INV_CO | MANUFACT<br>MANUFACT<br>MANUFACT | MANUF_CODE_EC<br>MANUF_CODE_SU<br>MANUF_CODE_CO | MODEL | Check that the spelling is consistent. | N/A |
| 8 | MISSING | | | | | | |
| 9 | MISSING | | | | | | |
| 10 | ALL | DCS_INSTALLATION<br>DCS_INSTALLATION_HISTORY | PROD_STARTUP<br>I_START_DATE | | | Check if I_START_DATE >= PROD_STARTUP. | Use the IF function as stated on the comment description. |
| 11 | ALL | DCS_FAILURE_EVENT<br>DCS_MAINT_EVENT | FAILURE_MODE_SU<br>SEVERITY_SU | FAILURE_MODE_EC<br>SEVERITY_EC | FAILURE_MODE_CO<br>SEVERITY_CO | Check if Type of FAILURE_MODE_SU = 1, then SEVERITY_SU = C;<br>Check if Type of FAILURE_MODE_EC = 1, then SEVERITY_EC = C;<br>Check if Type of FAILURE_MODE_CO = 1, then SEVERITY_CO = C. | Use the IF function as stated on the comment description. |
| 12 | ALL | DCS_FAILURE_EVENT<br>DCS_MAINT_EVENT | STATUS<br>STATUS | | | Check if STATUS = 6 (completed). | Use the IF function as stated on the comment description. |
| 13 | ALL | DCS_INSTALLATION_HISTORY | I_START_DATE | I_END_DATE | | Check if I_END_DATE - I_START_DATE <= the OREDA phase period. | Use the IF function as stated on the comment description. |
| 14 | ALL | DCS_FAILURE_EVENT<br>DCS_MAINT_EVENT | STATUS<br>STATUS | | | Check If STATUS = 3 (new code proposed), then write a change request. | Use the IF function as stated on the comment description with an indication |
| 15 | ALL | ALL TABLES | | | | Check completeness of data. Mandatory = 100% , Desirable = 85%. | Use the COUNTBLANK function with the range as specified on the comment description. |
| 16 | ALL | DCS_FAILURE_EVENT<br>DCS_INSTALLATION_HISTORY<br>DCS_FAILURE_EVENT | FDATE<br>I_START_DATE<br>OP_PHASE | | | Check if I_START_DATE <FDATE <I_END_DATE. | Use the IF function as stated on the comment description. |
| 18 | ALL | DCS_INV_EC<br>DCS_INV_SU<br>DCS_INV_CO | MANUFACT<br>MANUFACT<br>MANUFACT | | | Check that there is no BLANK cell for MANUFACT, use UNKNOWN instead. | Use the COUNTBLANK function with the range as specified on the comment description. |
| 19 | ALL | DCS_INSTALLATION_HISTORY<br>DCS_INV_EC_HISTORY<br>DCS_INV_SU_HISTORY<br>DCS_INV_CO_HISTORY | I_START_REASON<br>IEC_START_REASON<br>ISU_START_REASON<br>ICO_START_REASON | | | Check if DCSTA (Data Collection Start-Up) is unused for START_REASON | Use the IF function as stated on the comment description. |
| 20 | ALL | DCS_MAINT_EVENT | MNT_CONSEQUENCE | DOWNTIME | | Check if MNT_CONSEQUENCE = NONE (No consequence), then DOWNTIME = 0. | Use the IF function as stated on the comment description. |
| 21 | CONNECTOR<br>HYDR_CONN<br>CHEM_CONN<br>VV_CHOKE<br>VV_ISOL_PR<br>VV_ISOL_UT | DCS_INV_CO | MANUFACT | MANUF_CODE | | Check if the manufacturer for the components follow DCS_CO_MANUF_CODE_CO from the latest taxonomy | Use the conditional formatting function : create new rules for duplicate values. |
| 22 | ALL | DCS_INV_EC | MANUFACT | MANUF_CODE_EC | | Check if the manufacturer name is not the yard or installation company | N/A |
| 23 | CS | DCS_INV_CO<br>DCS_FAILURE_EVENT<br>DCS_MAINT_EVENT | EC_CODE<br>EC_CODE<br>EC_CODE | COMP_CODE<br>COMP_CODE<br>COMP_CODE | | Check if EC_CODE = CS then COMP_CODE is not SUTU (Subsea Umbilical Termintion Unit) nor SUBSEA_C (Subsea Cabling). | Use the IF function as stated on the comment description. |
| 24 | CS | DCS_INV_CO | EC_CODE | COMP_CODE | | Check if EC_CODE = CS and COMP_CODE = PR_TE_SENS (combined pressure and temperature sensors) are collected according to guideline recommendations. | N/A |
| 25 | CS & WC | DCS_INV_CO_HISTORY | EC_CODE | COMP_CODE | NO_COMPONENTS | Check if NO_COMPONENT have a 'right' number when EC_CODE = CS or WC and COMP_CODE = CONNECTOR or VALVE. | N/A |
| 26 | WC | DCS_INV_EC_HISTORY | I_START_DATE | | | Check if I_START_DATE is different from various WC. | N/A |
| 27 | Hierarchy | ALL TABLES | | | | Check the hierarchy structure based on latest taxonomy | N/A |

= Unclear checklists.
= Applicable on the merged checklist.
= Extra attention on quantitative indication.
= Missing Checklists

### C.0.3 Extended Quality Assurance Checklists on The Merged Database

| NO | Equipment | Table | Fields | | | Comment Description | Automation Process Description |
|---|---|---|---|---|---|---|---|
| 1 | ALL | ALL TABLES | | | | Evaluation of format alteration (e.g. from "string" to "list box"). | N/A |
| 2 | ALL | ALL TABLES | | | | Update the existing used code with the code on the latest taxonomy. | N/A |
| 3 | CO | DC_INV_M02 | UOM | | | Change the unit to KJ/Kg. | Use Find & Replace function based on statement on the comment description. |
| 4 | PU | DC_INVENTORY | DC_CODE | | | Change Diaphragm "DI" into Reciprocating "RE" and Screw "SC" into Rotary "RO". | Use Find & Replace function based on statement on the comment description. |
| 5 | PU, WC TM | DC_INVENTORY | | | | New DB fields (not specified which fields). | N/A |
| 6 | ALL | DC_FAILURE EVENT<br>DC_MAINT_EVENT<br>DC_INVENTORY | ELC_STATUS_ID<br>ELC_STATUS_ID<br>ELC_STATUS_ID | | | removed all unnecessary records. Evaluate ELC_STATUS_ID = "2" (Interpretation problem), "3" (New code proposed), "4" (Missing data). | Existing script (Delete.wat) is available. |
| 7 | ALL | ALL TABLES | INST_ID | | | Check that each companies has unique INST_ID. | Use the conditional formatting function : create new rules for unique values. |
| 8 | ALL | Statoil Vs Exxon | | | | Relevant for phase IV data only. | N/A |
| 9 | ALL | ALL TABLES | | | | Check that no duplicate records exist. | Use the conditional formatting function : create new rules for unique values. |
| 10 | ALL | ALL TABLES | | | | NOT CLEAR. | Existing script (Itemmain.wat) is available. |
| 11 | ALL | DC_INV_SUB_UNIT | ISU_NO_INSTALLED | | | Check if ISU_NO_INSTALLED = 0, Then "Figure 1" should be inserted. | Existing script (Subunit.wat) is available. |
| 12 | ALL | DC_FAILURE EVENT<br>DC_MAINT_EVENT<br>DC_INVENTORY | ELC_STATUS_ID<br>ELC_STATUS_ID<br>ELC_STATUS_ID | | | Change ELC_STATUS_ID = "1" (ongoing) & "5" (No QA) into "6" (completed). | Existing script (Fixstate.wat) is available. |
| 13 | ALL | DC_MAINT_EVENT | M_TOTAL_MANHOUR | | | Check if M_TOTAL_MANHOUR = MEC + EL + INST + OTHER. | Existing script (Totalman.wat) is available. |
| 14 | ALL | DC_INVENTORY | I_MANUFACT | I_MANUFACT_CTRL_SYS | I_MODEL | Check if the spelling is consistent. | Existing script (Manuf.wat) is available. |
| 15 | ALL | ALL TABLES | | | | Compare and update existing used code with the code on the latest taxonomy | N/A |
| 16:01 | ALL | DC_FAILURE EVENT<br>DC_INVENTORY HISTORY | F_DETECTED_DATE<br>I_SURV_START_DATE | I_SURV_END_DATE | | Check if I_SURV_START_DATE <F_FAILURE_DATE <I_SURV_END_DATE. | Use the IF function as stated on the comment description. |
| 16:02 | ALL | DC_INVENTORY<br>DC_INVENTORY HISTORY | I_INSTALLED_DATE<br>I_SURV_START_DATE | | | Check if I_INSTALLED_DATE <I_SURV_START_DATE | Use the IF function as stated on the comment description. |
| 16:03 | ALL | DC_MAINT_EVENT<br>DC_INVENTORY HISTORY | M_MAINT_DATE<br>I_SURV_START_DATE | I_SURV_END_DATE | | Check if I_SURV_START_DATE <M_MAINT_DATE <I_SURV_END_DATE | Use the IF function as stated on the comment description. |
| 16:04 | ALL | DC_MAINT_EVENT<br>DC_FAILURE EVENT | MC_CODE<br>F_DECECTED_DATE | M_MAINT_DATE | | Check if MC_CODE = "CORRECT", then M_MAINT_DATE >= F_DETECTED_DATE | Use the IF function as stated on the comment description. |
| 16:05 | ALL | DC_INVENTORY HISTORY | I_SURV_START_DATE | I_SURV_END_DATE | I_OPER_TIME | Check if I_SURV_END_DATE -I_SURV_START_DATE >= I_OPER_TIME | Use the IF function as stated on the comment description. |
| 16:06 | ALL | DC_INVENTORY HISTORY<br>DCS_INV_SU_HISTORY<br>DCS_INV_CO_HISTORY | I_SURV_START_DATE<br>ISU_START_REASON<br>ICO_START_REASON | I_SURV_END_DATE | I_OPER_TIME | NOT CLEAR | N/A |

= Unclear checklists.
= Applicable on the received data checklist.
= Not applicable anymore. To be removed.

# Bibliography

Adams, S. (2018). Is the full text the full answer?–considerations of database quality. *World Patent Information*, 54:S66–S71.

Alizamini, F. G., Pedram, M. M., Alishahi, M., and Badie, K. (2010). Data quality improvement using fuzzy association rules. In *2010 International Conference on Electronics and Information Engineering*, volume 1, pages V1–468. IEEE.

Bailey, S. F., Scheible, M. K., Williams, C., Silva, D. S., Hoggan, M., Eichman, C., and Faith, S. A. (2017). Secure and robust cloud computing for high-throughput forensic microsatellite sequence analysis and databasing. *Forensic Science International: Genetics*, 31:40–47.

Baker, R. D. (2000). Calculating the expected failure rate of complex equipment subject to hazardous repair. *International Journal of Production Economics*, 67(1):53–61.

Barabadi, A. (2014). Reliability analysis of offshore production facilities under arctic conditions using reliability data from other areas. *Journal of Offshore Mechanics and Arctic Engineering*, 136(2):021601.

Barabadi, A., Gudmestad, O. T., and Barabady, J. (2015). Rams data collection under arctic conditions. *Reliability Engineering & System Safety*, 135:92–99.

Barabadi, A. and Markeset, T. (2011). Reliability and maintainability performance under arctic conditions. *International Journal of System Assurance Engineering and Management*, 2(3):205–217.

Brissaud, F., Charpentier, D., Fouladirad, M., Barros, A., and Bérenguer, C. (2010). Failure rate evaluation with influencing factors. *Journal of Loss Prevention in the Process Industries*, 23(2):187–193.

Bunea, C., Mazzuchi, T. A., Sarkani, S., and Chang, H.-C. (2008). Application of modern reliability database techniques to military system data. *Reliability Engineering & System Safety*, 93(1):14–27.

Cats, O. and Loutos, G. (2016). Evaluating the added-value of online bus arrival prediction schemes. *Transportation Research Part A: Policy and Practice*, 86:35–55.

Chen, S., Zhang, S., Zheng, X., and Ruan, X. (2019). Layered adaptive compression design for efficient data collection in industrial wireless sensor networks. *Journal of Network and Computer Applications*, 129:37–45.

Chen, W., Zhou, K., Yang, S., and Wu, C. (2017). Data quality of electricity consumption data in a smart grid environment. *Renewable and Sustainable Energy Reviews*, 75:98–105.

Ciliberti, V. A., Østebø, R., Selvik, J. T., Alhanati, F. J., et al. (2019). Optimize safety and profitability by use of the iso 14224 standard and big data analytics. In *Offshore Technology Conference*. Offshore Technology Conference.

Cressent, R., David, P., Idasiak, V., and Kratz, F. (2013). Designing the database for a reliability aware model-based system engineering process. *Reliability Engineering & System Safety*, 111:171–182.

DNV (1999). *Handbook on quality of reliability data*. Den Norske Veritas.

DNV-GL (2016a). Oreda guideline for data collection - topside equipment. 11:Appendix B.

DNV-GL (2016b). Oreda guideline for data collection - topside equipment. 11:Appendix A.

Duarte, J. C., Cunha, P. F., and Craveiro, J. T. (2013). Maintenance database. *Procedia CIRP*, 7:551–556.

Fanailoo, P., Andreassen, G., et al. (2008). Improving reliability and reducing intervention costs of ultradeep subsea technology at the design stage. In *Offshore Technology Conference*. Offshore Technology Conference.

Fernández-Medina, E. and Piattini, M. (2005). Designing secure databases. *Information and Software Technology*, 47(7):463–477.

Fidel, R. and Green, M. (2004). The many faces of accessibility: engineers' perception of information sources. *Information Processing & Management*, 40(3):563–581.

Gens, F. (2008). It cloud services user survey, pt. 2: Top benefits & challenges. *IDC eXchange*.

Haegemans, T., Snoeck, M., and Lemahieu, W. (2018). Entering data correctly: An empirical evaluation of the theory of planned behaviour in the context of manual data acquisition. *Reliability Engineering & System Safety*, 178:12–30.

Haegemans, T., Snoeck, M., and Lemahieu, W. (2019). A theoretical framework to improve the quality of manually acquired data. *Information & Management*, 56(1):1–14.

Hahn, B., Welte, T., Faulstich, S., Bangalore, P., Boussion, C., Harrison, K., Miguelanez-Martin, E., O'Connor, F., Pettersson, L., Soraghan, C., et al. (2017). Recommended practices for wind farm data collection and reliability assessment for o&m optimization. *Energy Procedia*, 137:358–365.

Harahap, V. A. (2018). Automation on oreda data analysis process in subsea application.

Hernandez, M. J. (2013). *Database design for mere mortals: a hands-on guide to relational database design*. Pearson Education.

Hodkiewicz, M., Kelly, P., Sikorska, J., and Gouws, L. (2006). A framework to assess data quality for reliability variables. In *Engineering Asset Management*, pages 137–147. Springer.

IEC61508 (2010). Functional safety of electrical/electronic/programmable electronic safety-related systems (part 1–7). *International Electrotechnical Commission Geneva*.

ISO14224 (2016). Petroleum, petrochemical and natural gas industries–collection and exchange of reliability and maintenance data for equipment. *International Organization for Standardization (ISO)*.

Koronios, A., Lin, S., and Gao, J. (2005). A data quality model for asset management in engineering organisations. In *ICIQ*.

Kozak, M., Krzanowski, W., Cichocka, I., and Hartley, J. (2015). The effects of data input errors on subsequent statistical inference. *Journal of Applied Statistics*, 42(9):2030–2037.

Langseth, H., Haugen, K., and Sandtorv, H. (1998). Analysis of oreda data for maintenance optimisation. *Reliability Engineering and System Safety*, 60(2):103–110.

Larsen, I. K., Småstuen, M., Johannesen, T. B., Langmark, F., Parkin, D. M., Bray, F., and Møller, B. (2009). Data quality at the cancer registry of norway: an overview of comparability, completeness, validity and timeliness. *European journal of cancer*, 45(7):1218–1231.

Lee, Y. W. (1996). Why 'know why'knowledge is useful for solving information quality problems. In *Americas Conference on Information Systems*, pages 200–202.

Lee, Y. W. and Strong, D. M. (2003). Knowing-why about data processes and data quality. *Journal of Management Information Systems*, 20(3):13–39.

Lin, S., Gao, J., and Koronios, A. (2008). A data quality framework for engineering asset management. *Australian Journal of Mechanical Engineering*, 5(2):209–219.

Linsday, J., Briand, D., Hill, R. R., Stinebaugh, J. A., and Benjamin, A. S. (2008). Wind turbine reliability: a database and analysis approach. Technical report, Sandia National Laboratories.

Liu, P., Qiu, Y., Hu, J., Tong, J., Zhao, J., and Li, Z. (2018). Expert judgments for performance shaping factors' multiplier design in human reliability analysis. *Reliability Engineering & System Safety*.

Molina, R., Unsworth, K., Hodkiewicz, M., and Adriasola, E. (2013). Are managerial pressure, technological control and intrinsic motivation effective in improving data quality? *Reliability Engineering & System Safety*, 119:26–34.

Murphy, G. D. (2009). Improving the quality of manually acquired data: Applying the theory of planned behaviour to data quality. *Reliability Engineering & System Safety*, 94(12):1881–1886.

Naseri, M., Baraldi, P., Compare, M., and Zio, E. (2016). Availability assessment of oil and gas processing plants operating under dynamic arctic weather conditions. *Reliability Engineering & System Safety*, 152:66–82.

Nassiri, H., Machkour, M., and Hachimi, M. (2018). One query to retrieve xml and relational data. *Procedia Computer Science*, 134:340–345.

Neto, A. P., Moreno, U. F., and Orth, A. (2018). Failure distribution analysis of a novel subsea valve actuator concept based on reliability database. *IFAC-PapersOnLine*, 51(24):1247–1254.

Okaro, I. A. and Tao, L. (2016). Reliability analysis and optimisation of subsea compression system facing operational covariate stresses. *Reliability Engineering & System Safety*, 156:159–174.

Oliva, J. T., Lee, H. D., Spolaor, N., Takaki, W. S. R., Coy, C. S. R., Fagundes, J. J., and Wu, F. C. (2019). A computational system based on ontologies to automate the mapping process of medical reports into structured databases. *Expert Systems with Applications*, 115:37–56.

OREDA (2015). *Offshore reliability data handbook*. OREDA.

Parasuraman, R. and Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human factors*, 52(3):381–410.

Pelzer, P., Geertman, S., van der Heijden, R., and Rouwette, E. (2014). The added value of planning support systems: A practitioner's perspective. *Computers, Environment and Urban Systems*, 48:16–27.

Sağlam, R. B. and Temizel, T. T. (2014). Automatic information timeliness assessment of diabetes web sites by evidence based medicine. *Computer methods and programs in biomedicine*, 117(2):104–113.

Saha, B. and Srivastava, D. (2014). Data quality: The other face of big data. In *2014 IEEE 30th International Conference on Data Engineering*, pages 1294–1297. IEEE.

Sharma, K. (2016). *Overview of industrial process automation*. Elsevier.

SINTEF (2007). *Data Collection Quality Assurance Procedures and Checklist*. SINTEF.

SINTEF (2015). *OREDA Offshore and Onshore Reliability Data 6th Edition – Volume 2 - Subsea Equipment*. DNV GL.

Spüntrup, F. S., Londono, J., Skourup, C., Thornhill, N., and Imsland, L. (2018). Reliability improvement of compressors based on asset fleet reliability data. *IFAC-PapersOnLine*, 51(8):217–224.

Strong, D. M., Lee, Y. W., and Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, 40(5):103–110.

Sun, Y. (2004). Methods for automated concept mapping between medical databases. *Journal of biomedical informatics*, 37(3):162–178.

Tapia, C., Dies, J., Abal, J., Ibarra, Á., and Arroyo, J. M. (2011). Exploration of reliability databases and comparison of former ifmif's results. *Fusion Engineering and Design*, 86(9-11):2726–2729.

Unsworth, K., Adriasola, E., Johnston-Billings, A., Dmitrieva, A., and Hodkiewicz, M. (2011). Goal hierarchy: Improving asset data quality by improving motivation. *Reliability Engineering & System Safety*, 96(11):1474–1481.

Uyiomendo, E. E. and Tore, M. (2015). Subsea maintenance service delivery: A multi-variable analysis model for predicting potential delays in scheduled services. *Journal of Quality in Maintenance Engineering*, 21(1):34–54.

Vroeijenstijn, A. I. (1995). *Improvement and Accountability: Navigating between Scylla and Charybdis. Guide for External Quality Assessment in Higher Education. Higher Education Policy Series 30.* ERIC.

Walter, M., Siegle, M., and Bode, A. (2008). Opensesame—the simple but extensive, structured availability modeling environment. *Reliability Engineering & System Safety*, 93(6):857–873.

Wang, R. Y. and Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4):5–33.

Wickens, C. D., Hollands, J. G., Banbury, S., and Parasuraman, R. (2015). *Engineering psychology & human performance.* Psychology Press.

Wilton, P. and Colby, J. W. (2005). *Beginning sql.* John Wiley & Sons.

Yokogawa (2019). Yokogawa pressure transmitters evolution.

Zanetti, R., Schmidtmann, I., Sacchetto, L., Binder-Foucard, F., Bordoni, A., Coza, D., Ferretti, S., Galceran, J., Gavin, A., Larranaga, N., et al. (2015). Completeness and timeliness: cancer registries could/should improve their performance. *European journal of cancer*, 51(9):1091–1098.

Zhang, F., Liu, Q., and Zeng, A. (2017). Timeliness in recommender systems. *Expert Systems with Applications*, 85:270–278.

Zio, E. (2009). Reliability engineering: Old problems and new challenges. *Reliability Engineering & System Safety*, 94(2):125–141.