

SEA4DQ 2021

# The Development of Data Quality Management System for Ship IoT Data - Perspective of Ship Owner and Operator

August 24<sup>th</sup>, 2021

**Putu Hangga** | Unit Leader | Monohakobi Technology Institute - NYK Group

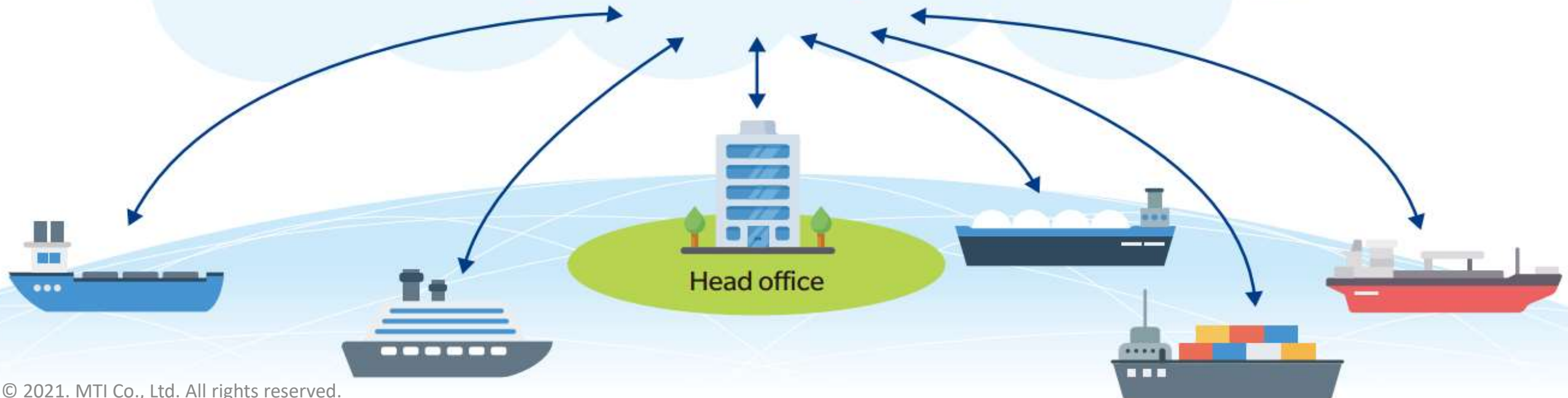
**Shogo Yamada** | Manager | Marine Group - NYK Line

# Data is in the core of future NYK operation



Information is collected in the cloud and shared with land

**Information is shared as knowledge for  
avoiding risks and ensuring safe operation**

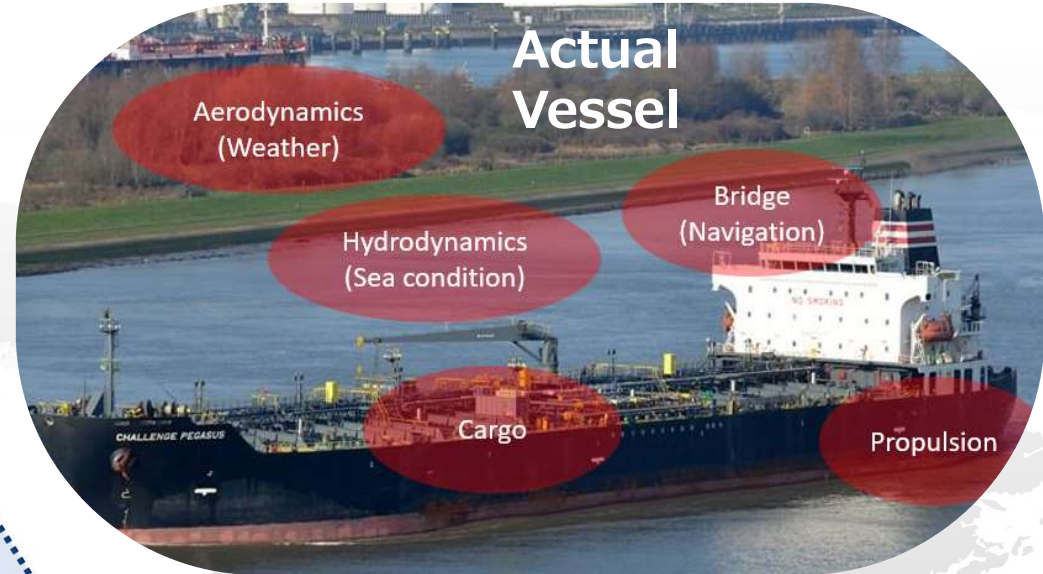


## NYK Effort to Ensure Safe Ship Operations by Enhancing Capabilities That Produce Safety

### SIMS Ship Information Management System

#### LIVE for Shipmanager

An application that compiles various ship operation data and processes it into data that can be used to predict danger, thereby contributing to safe ship operations and reducing fuel expenses



## SIMS

Measuring ship operation status and fuel efficiency data

Utilizing weather data to calculate efficient routes

Sharing of data on ship operations, fuel efficiency, and sea conditions

NYK / Ship-management companies



Ships



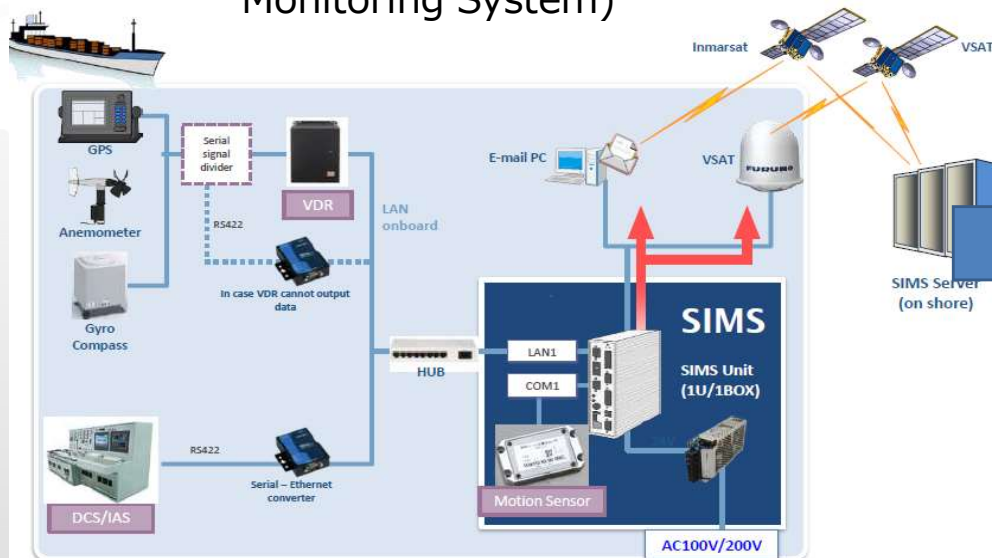
Introduced on  
Approximately  
**200 ships**



# Using Ship IOT to Ensure Safe Ship Operations by Enhancing Activities That Improve Safety

Industrial IoT system to collect and compile all digital data onboard a ship

Also called as VPMS (Vessel Performance Monitoring System)



## Data Viewer Application



consumed by the end-user.

## Data Analytics Application

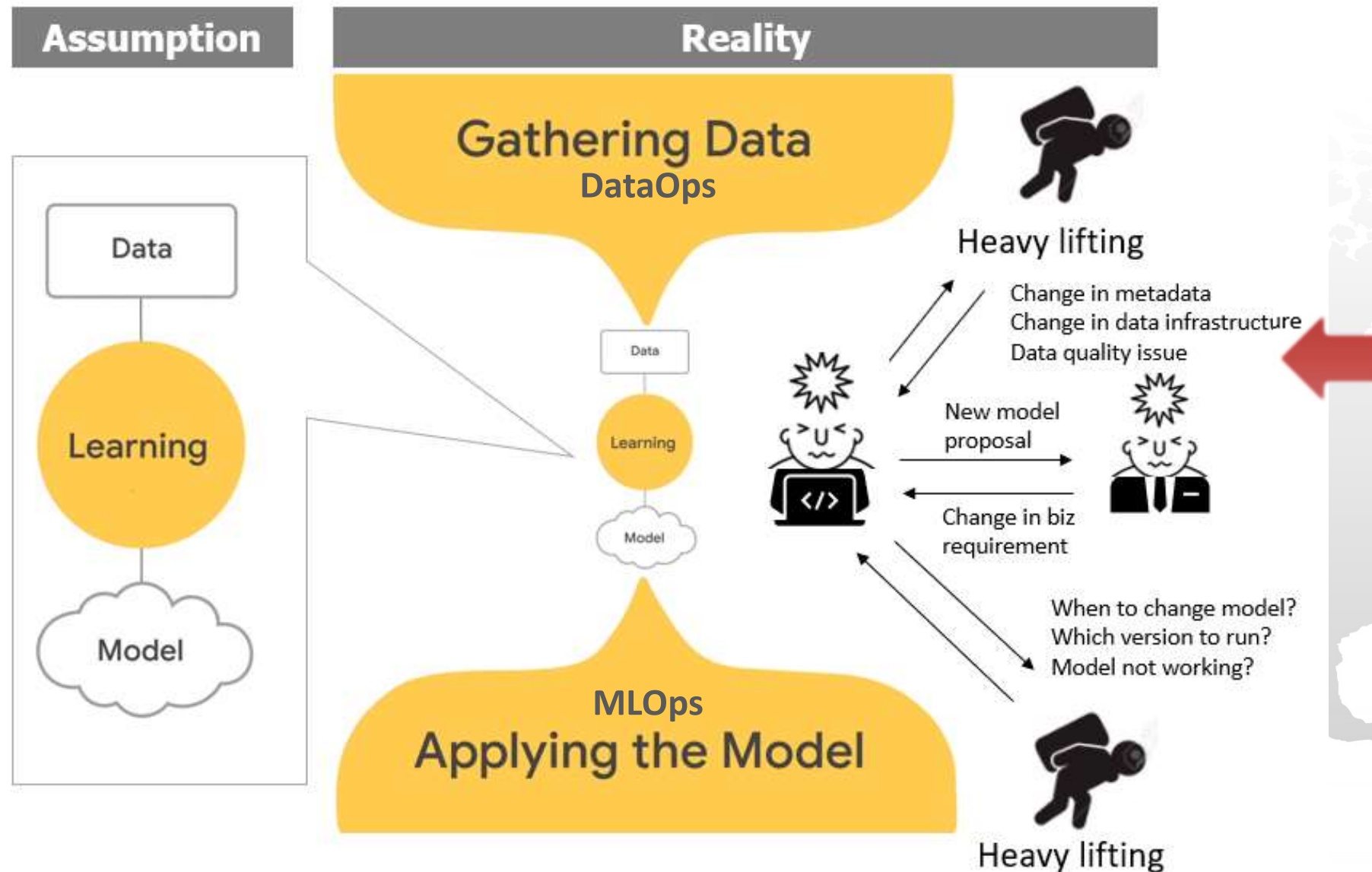


Used for many value-added activities

Can I rely on these data for analysis?

Are these analysis reliable?

# The Heavy lifting is not the analytics part

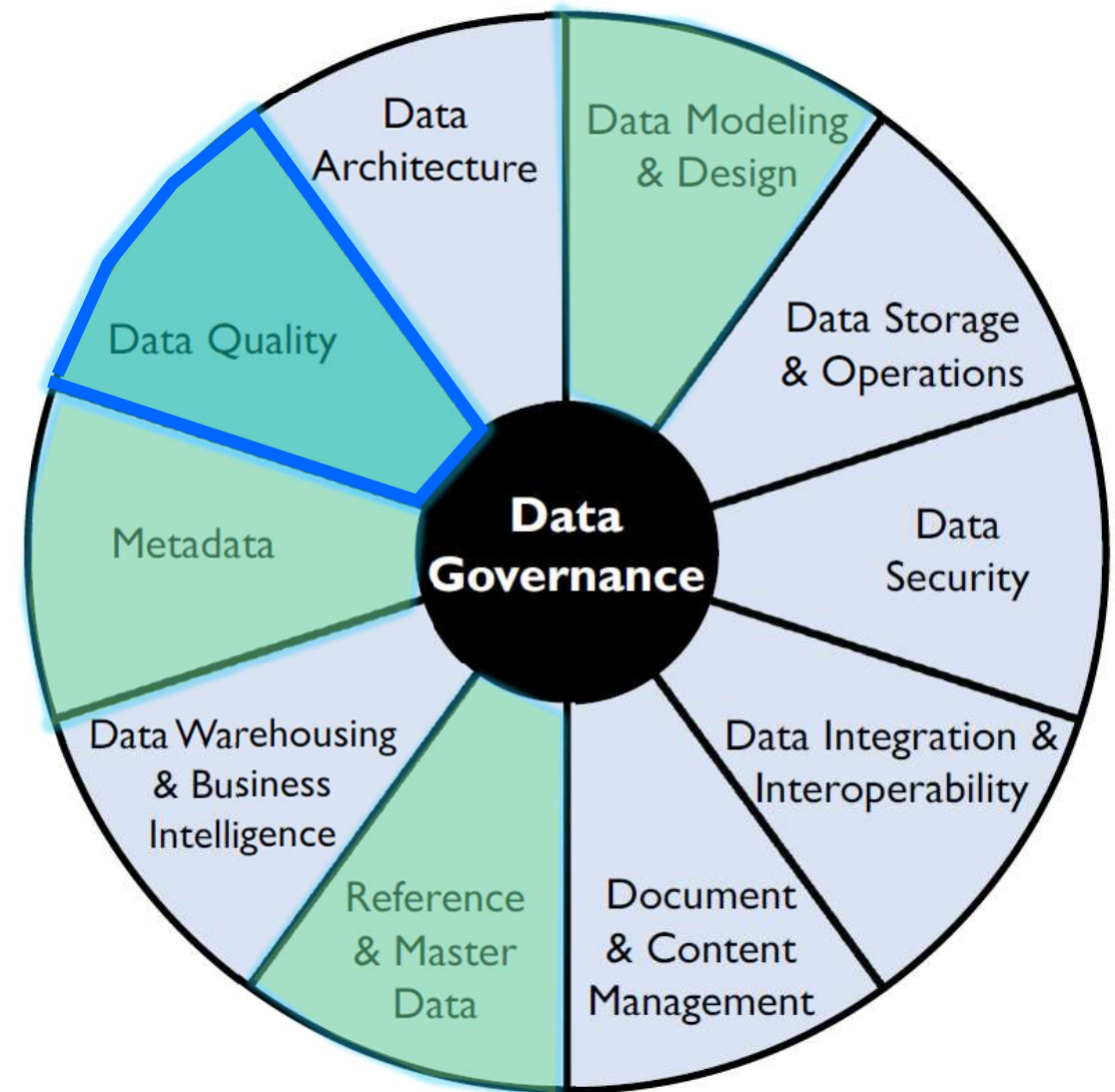


# The Journey of NYK IoT Data Quality Management System



## Scoping of Data Quality Management Projects

- **Data management is a complex**, with data governance at the center and the other knowledge areas are balanced around the it.
- NYK focuses on **data quality monitoring and management** since there are immediate needs to ensure good quality data supplied to data-driven analytic solution.
- DQMS is organized as **development of system, and continuous coverage of process & people**
- **All internal stakeholders need to commit** and we try to provide the values with immediate needs while touching other part of BOK



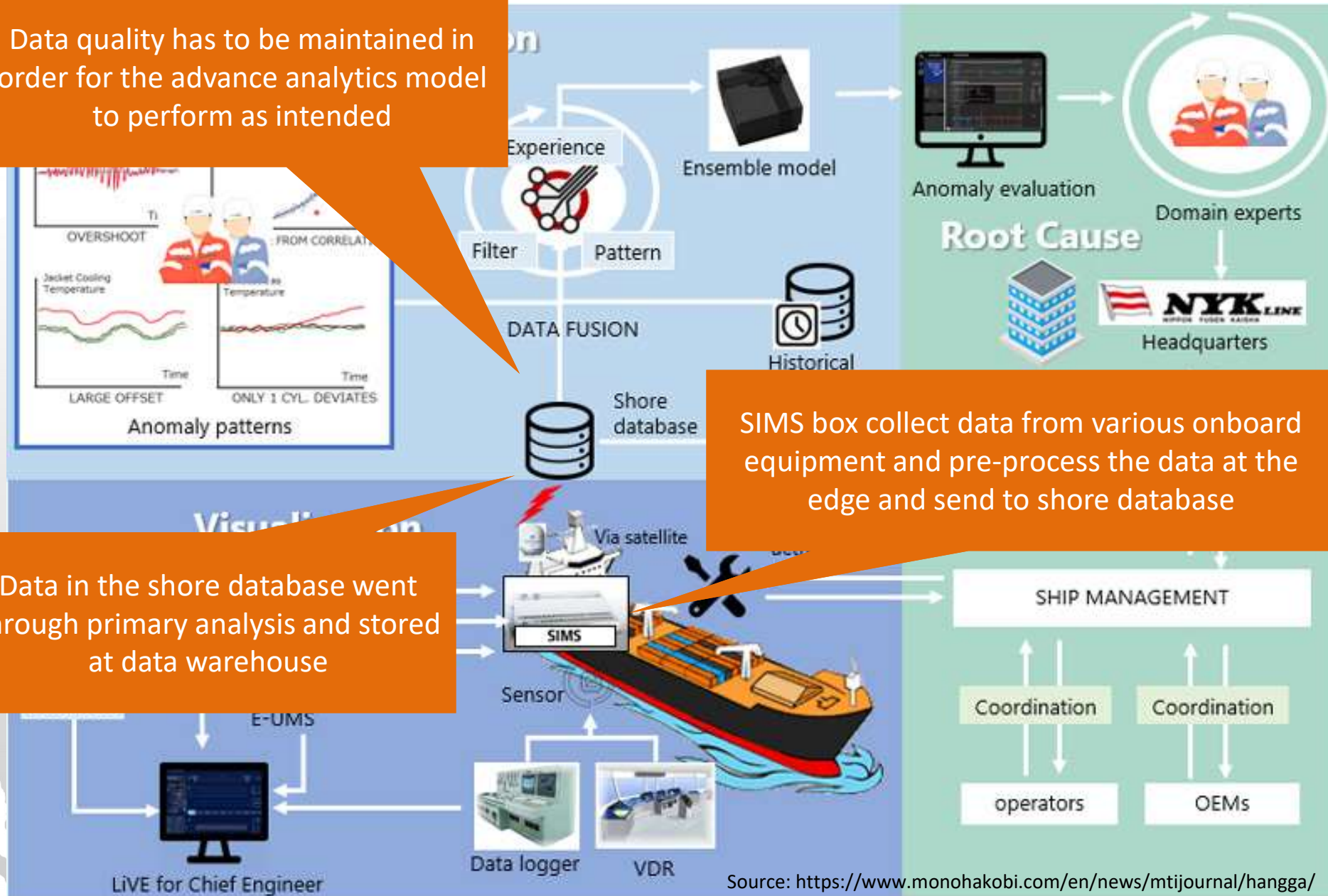
**Data Management Knowledge Areas**

Source: DAMA



# A Specific Purpose

Data quality has to be maintained in order for the advance analytics model to perform as intended



SIMS box collect data from various onboard equipment and pre-process the data at the edge and send to shore database

Data in the shore database went through primary analysis and stored at data warehouse

Source: <https://www.monohakobi.com/en/news/mtijournal/hangga/>

\* Potential data source in edge-computed anomaly detection



# Purpose: High Data Quality Level to Support NYK Digitalization Activity

Current  
State

Digitalization Activities

Digital  
Transformation  
Of Ship Operation

Threats



**GARBAGE  
DATA**



**GARBAGE  
RESULTS**

## Business Target

Reduce workload of crew  
Reduce downtime of asset  
Improve Collaboration  
New Revenue Generation  
Etc.

Things that keep the  
“Goal” afloat

**1 International Standards**

Digital Class  
Class  
DNV  
Data Governance  
ISO 8000

**2 Data Buyer (New Business)**

**ShipDC**  
Shore DC

**SP**  
Solution Provider

セキュアな陸上...を活用した分析・  
保管とアクセス制御...分析などの高付加価  
きデータ提供...値サービス提供事業者

**DB**  
Data Buyer

自社製品へのフィードバックの  
ためにデータ利用権を購入  
例えば  
造船所、メーカー等

**3 Condition Based Maintenance**

Preventive Maintenance (PM)  
Reliability Centred Maintenance (RcM)  
Predictive Maintenance (PdM)

- User Defined Formula
- Anomaly Detection
- AOB

# Many other use cases



## Shipping

- Safety operation
- Vessel performance analysis
- Fleet operation optimization
- Weather routing



## Shipyard

- In-service performance analysis of delivered ships
- Feedback to new ship design



## Manufacturer

- Remote condition monitoring
- Remote diagnostics
- After service support



## Class Society

- Utilization in class inspection

## Insurance

- New services

## Regulatory use

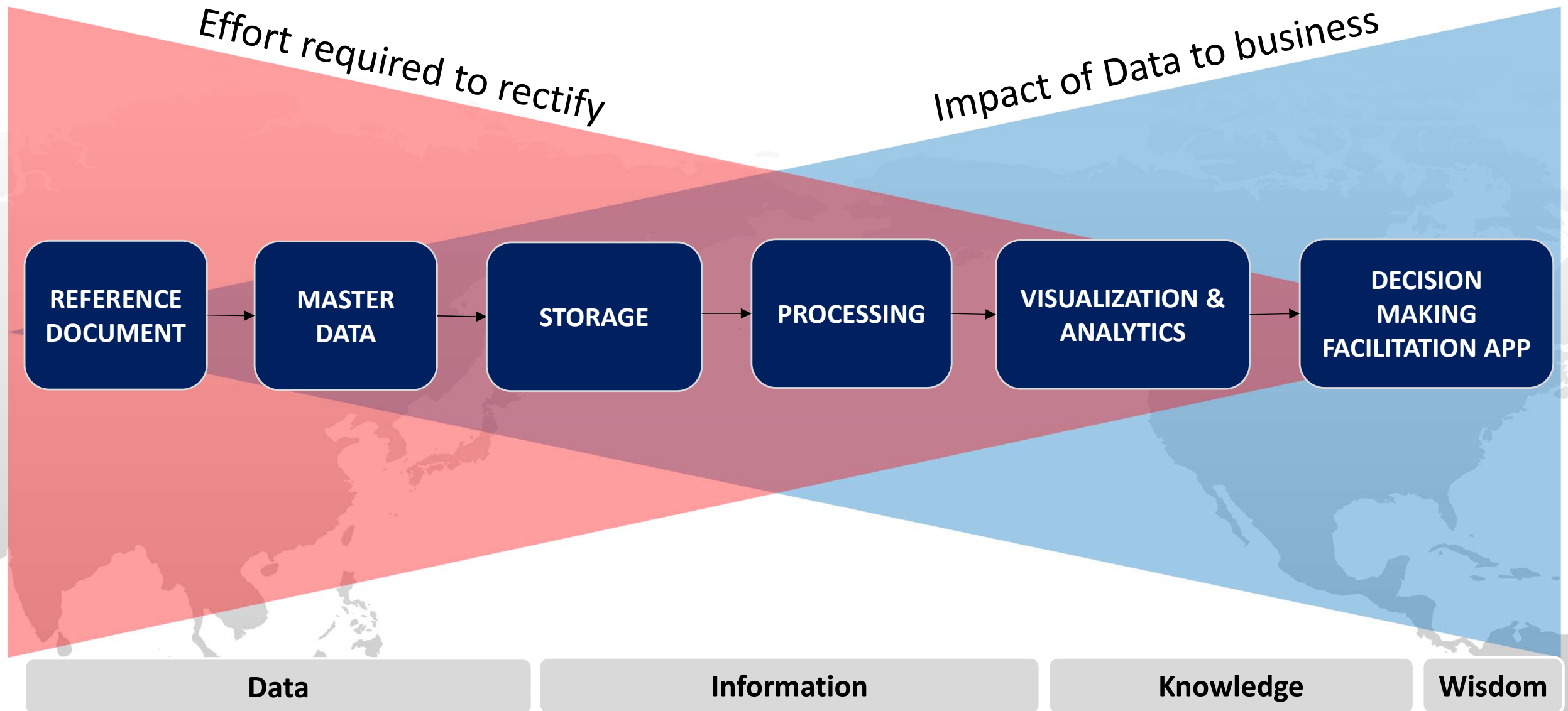
- Data reporting

# The flow of NYK IoT Data Quality Management System





# Risk of Data Error in each data lifecycle

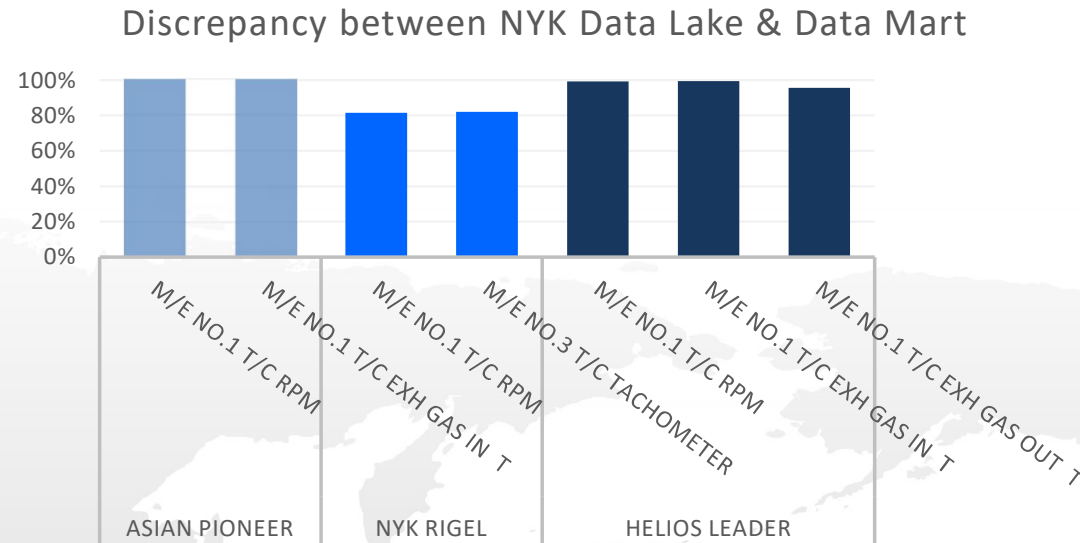


# Findings : Poor quality data leads to snowball effect

## [1] misleading vessel information

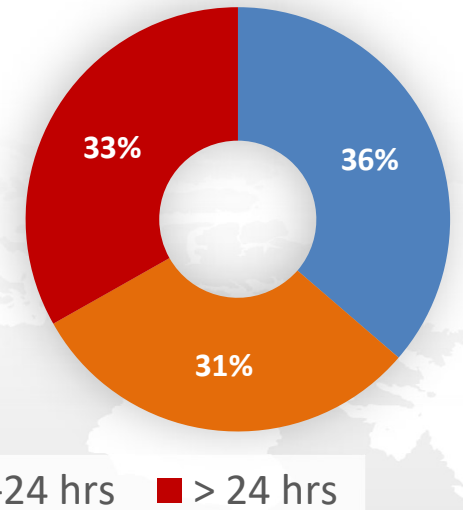


## [2] Inconsistency of data

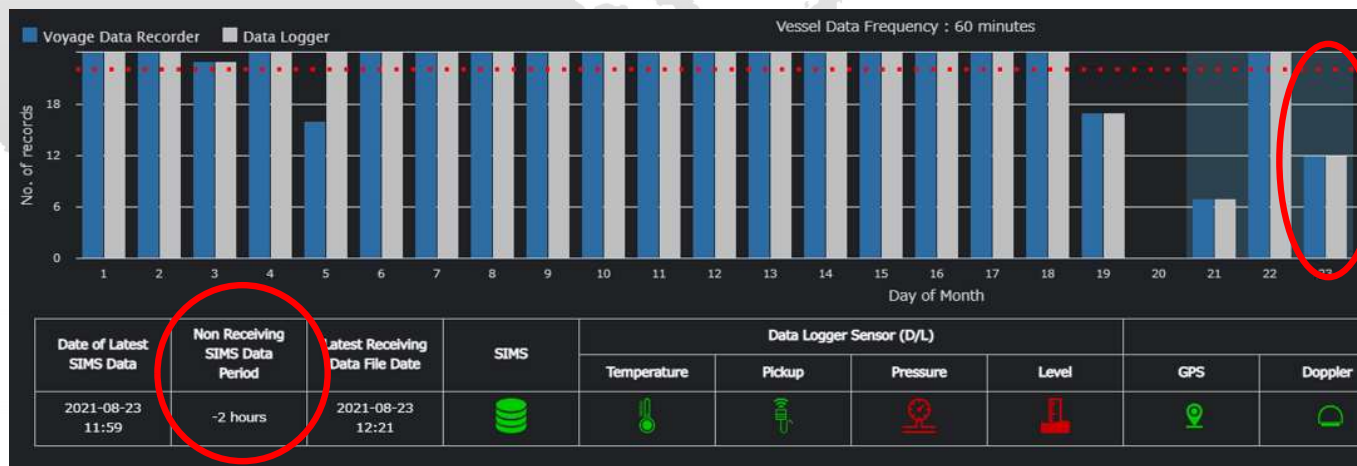


## [3] Delay in receiving the data

Data is not fresh to do condition monitoring

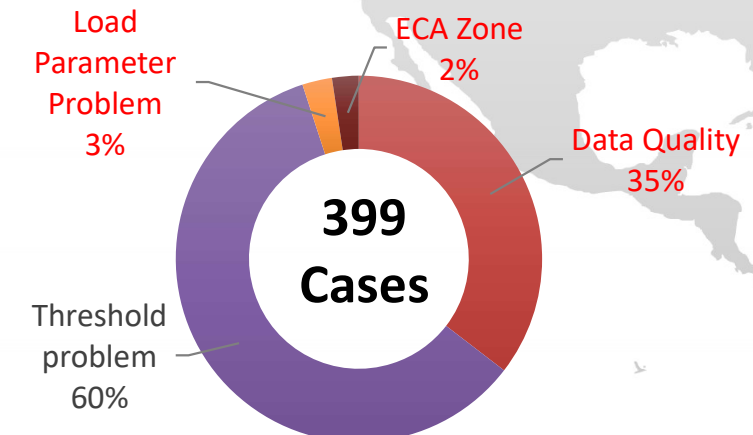


## [4] Back to the future

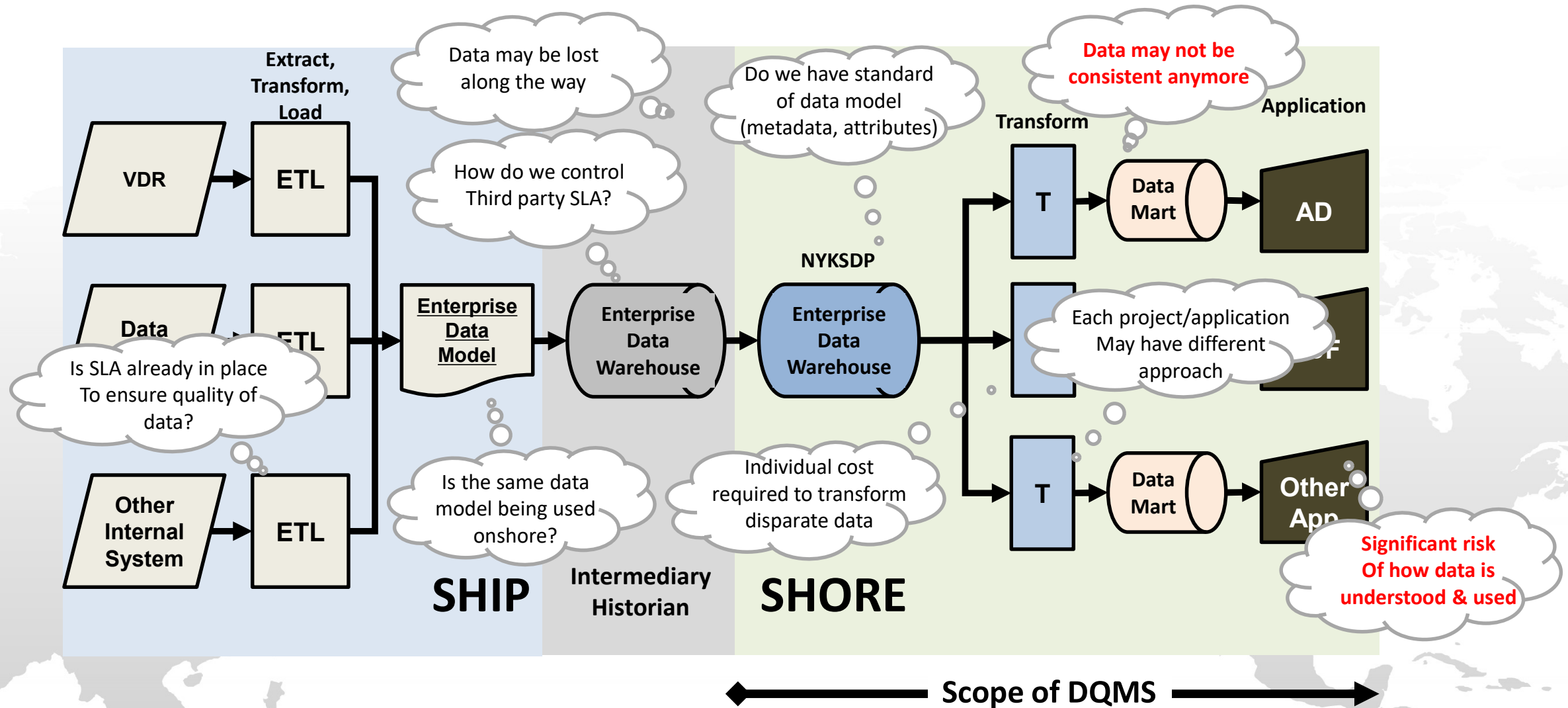


## [5] Unclear judgment for decision maker

Root cause of Expert unclear judgment from Anomaly Detection System



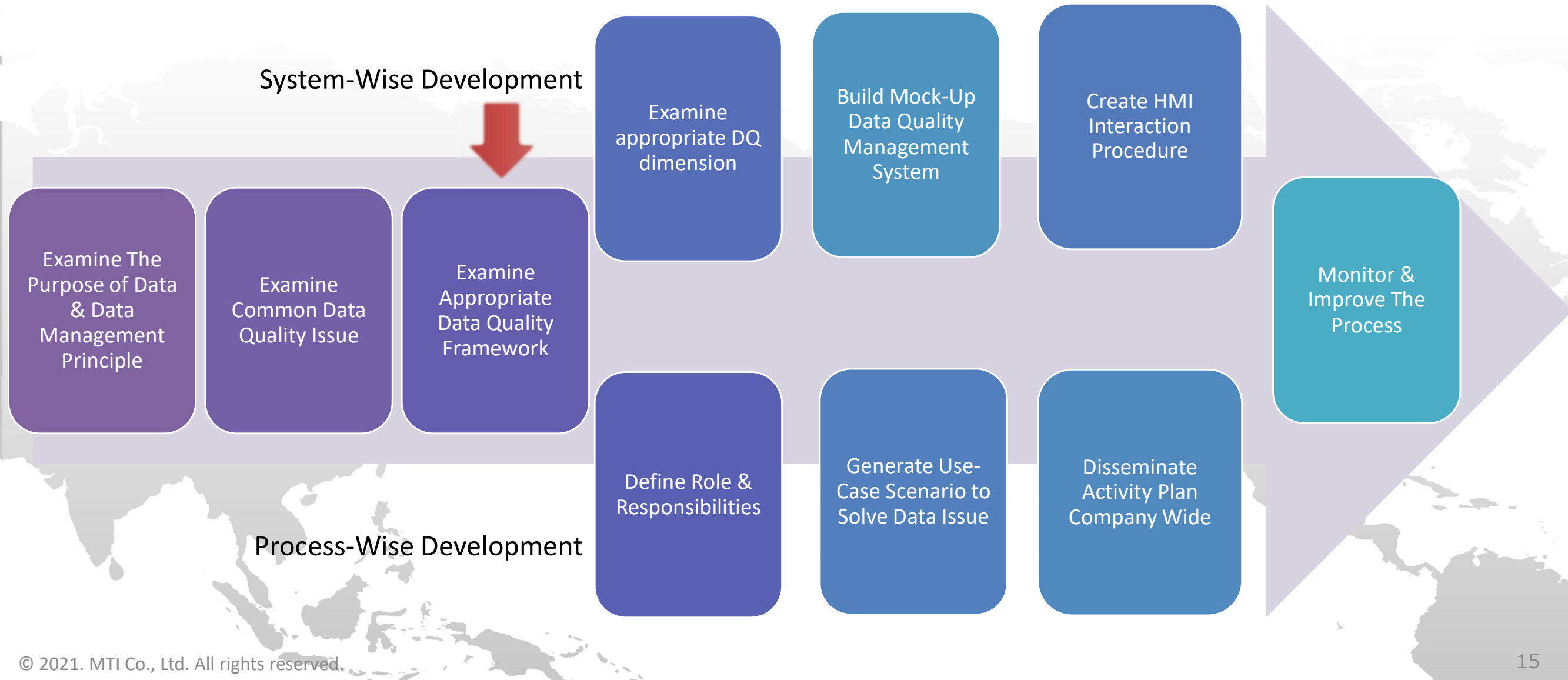
# Risk in the journey of NYK's Ship IoT Data



An Ideal way is to fix any data quality issues at the source.  
But operation is not an ideal world and we have to compromise things to achieve the goal.

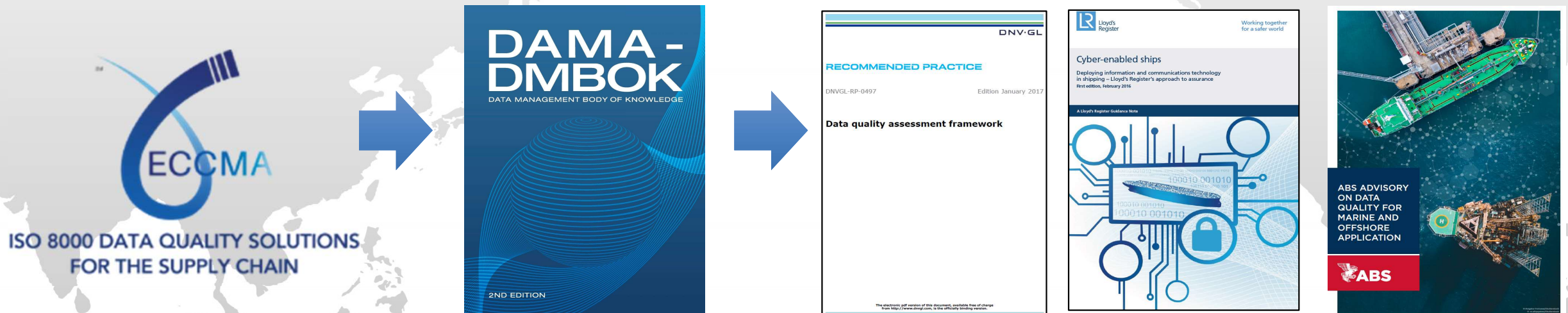


# The flow of NYK IoT Data Quality Management System

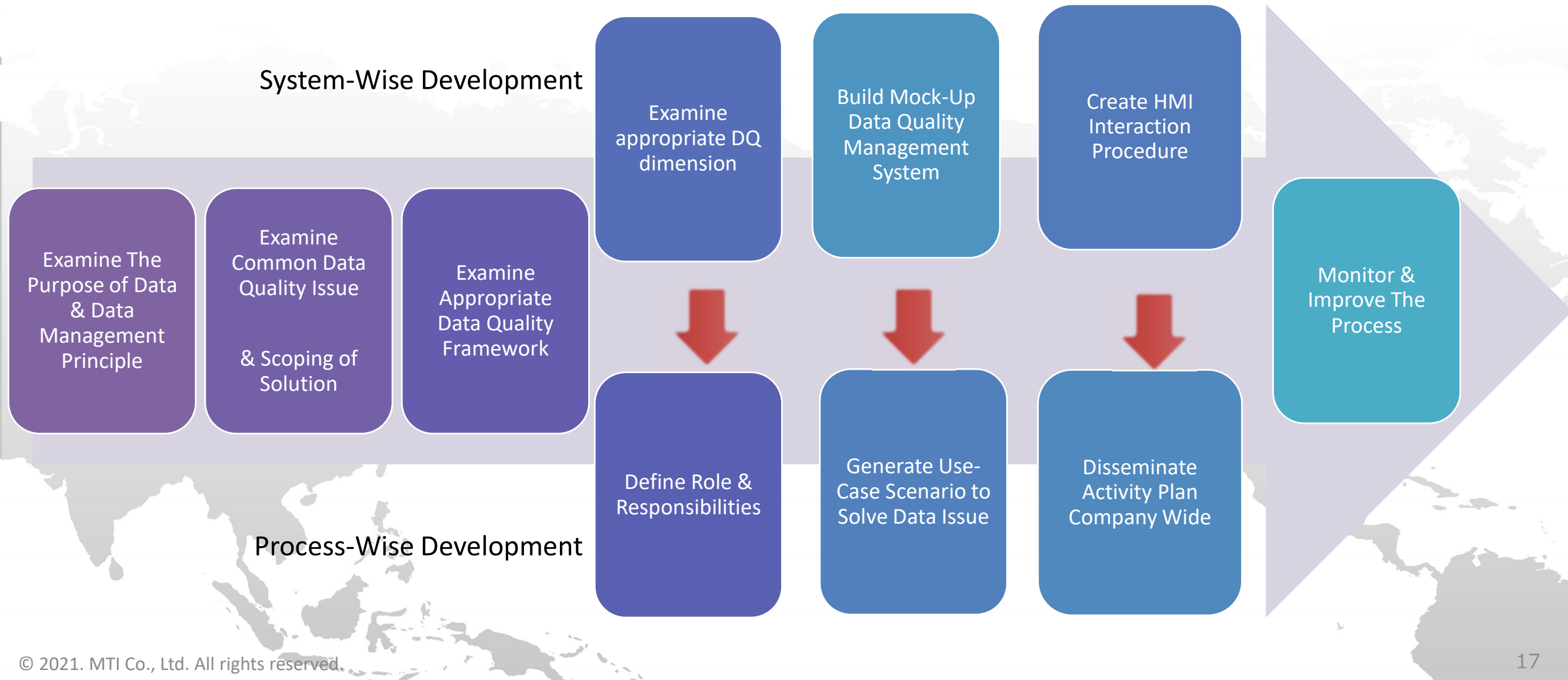


# The DQMS has been built with reference to various guidelines and best practices of classification societies.

- **The ISO 8000** families only state **high-level requirement** and framework to do IoT data quality management
- **The DAMA DMBok** provides a more **detailed methodology** but is cross-industry generic.
- **Classification societies** have developed **practical guidelines** for maritime applications and smart ships.
- The DQMS development are referring to the following guideline/practice:
  - **DNVGL-RP-0497**: Data quality assessment framework
  - **DNVGL-CG-0564 Section 9**: Data quality management for Smart Ship
  - ABS Advisory on Data Quality for Marine and Offshore Applications
  - LR Cyber-enabled ship

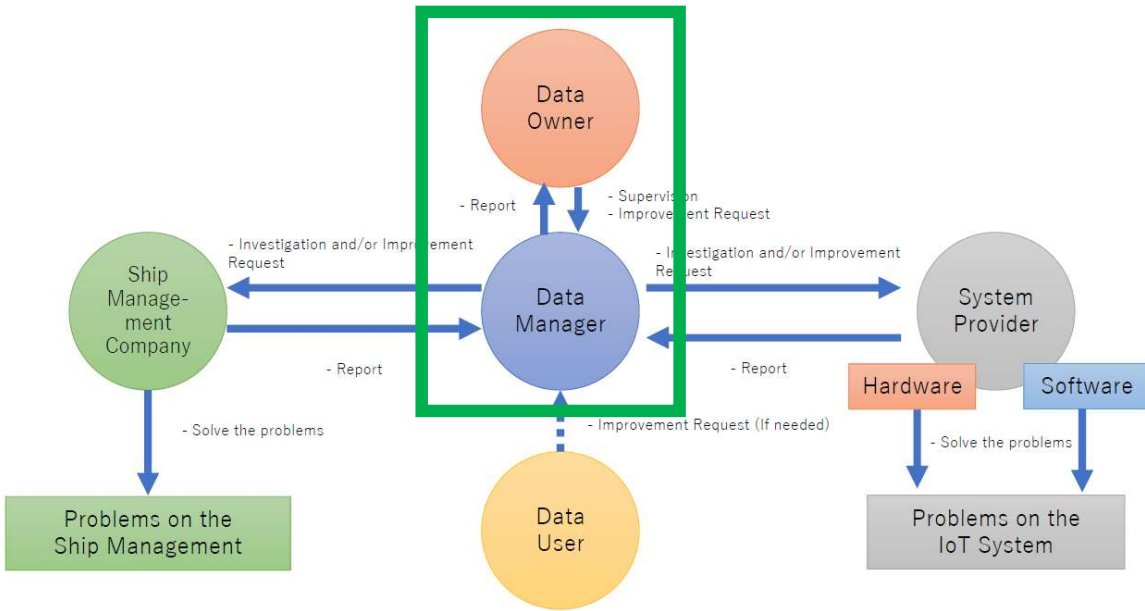


# The flow of NYK IoT Data Quality Management System





# Data Quality Monitoring Team



Data Manager at NYK HQ (Tokyo) designate the monitoring activity to NYK RDC (Remote Diagnostic Center) in Manila

## Data Quality

Monitoring and Maintaining data quality at a certain level

Toward higher safety levels

## Anomaly Detection

Using AI to leverage problem solving and decision making

## Data accuracy monitoring

RDC

"Expert in the Loop"

Use Of data

Assessment by experts

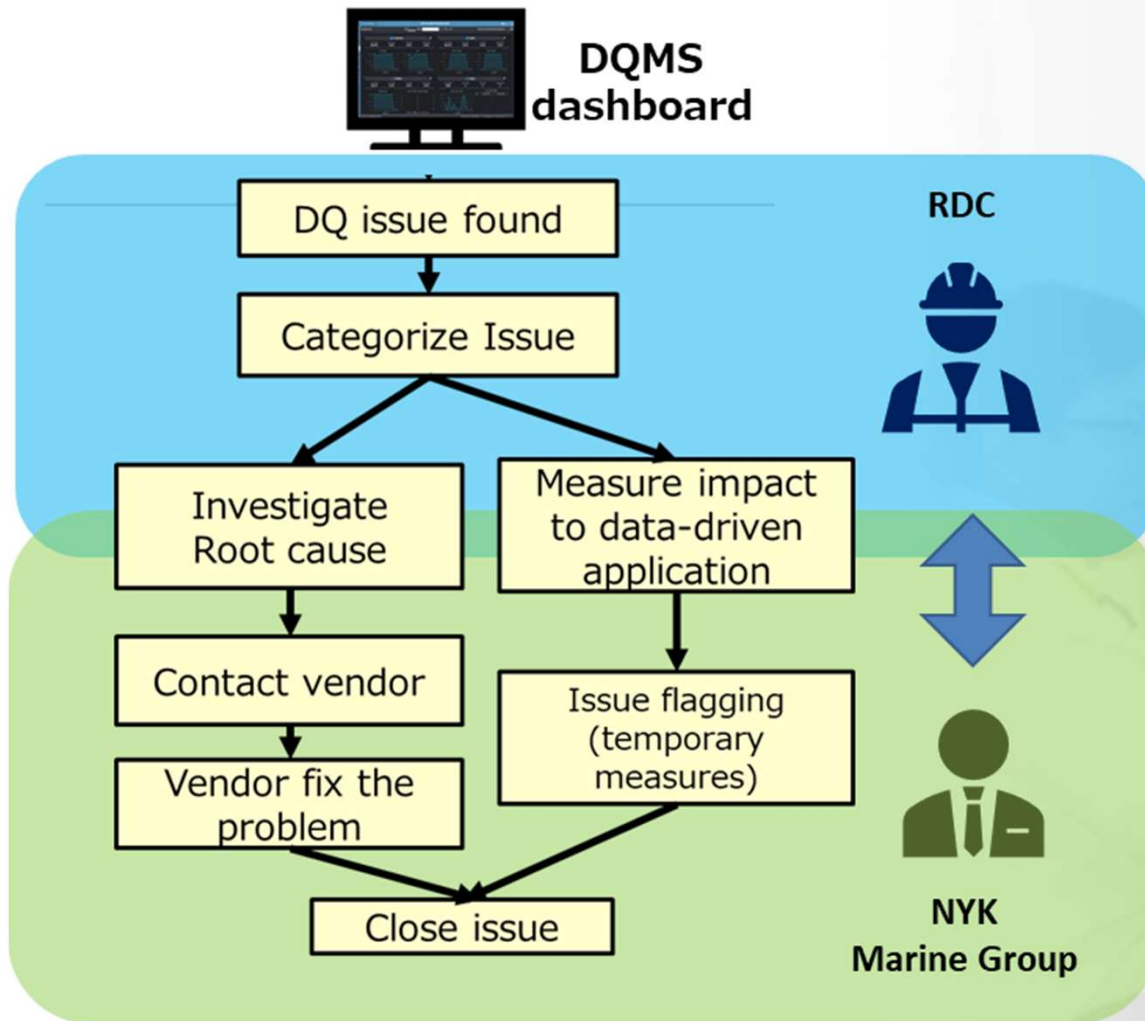
Domain experts conduct analysis



© NYK Line

# Use-Case Scenarios to Solve Data Issue

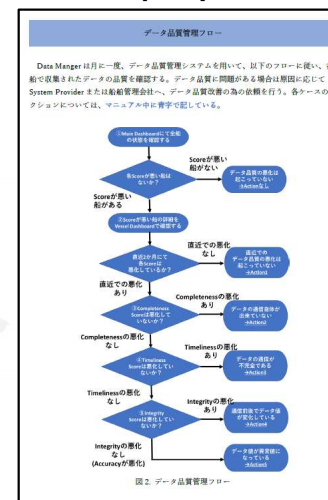
- Role of **data quality analyst** was assigned to RDC expert to do the use case scenario



- ▶ Company-wide circular to ensure good data input

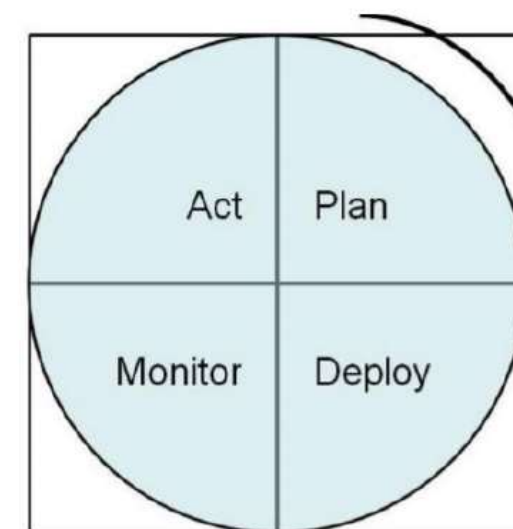
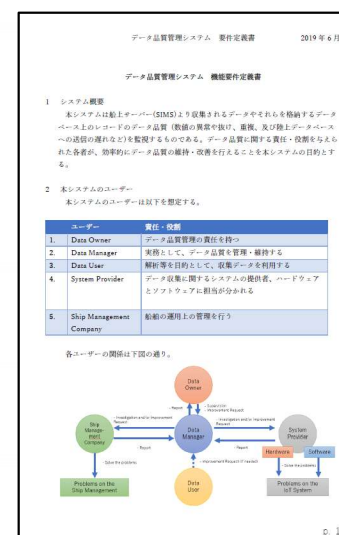
► DQM roles & responsibility, and  
Change Management Procedure

NYK collects various operational data from vessels through SIMS, e-UMS, and SPAS, and uses it for various purposes. It is expected that the data collected by these continuous monitoring systems will be used for verification in future GHG regulations. Therefore, the accuracy and timeliness of the data is very important. But there are some cases where the data is not fully utilized due to input errors, so we would like to ask for your cooperation in the following areas for future data reporting.



製品名・製品提供形態	データセンター	データ管理	データ分析	クラウドサービス	セキュリティ
製品概要	データセンターの運用と管理を行う。	顧客データの収集・管理、データの統合・分析、顧客との連携を行う。	顧客データの分析・予測、顧客の行動・嗜好の分析、顧客のニーズの把握を行う。	顧客データのクラウド化、データのバックアップ・復元、データの移行を行う。	顧客データのセキュリティ対策、顧客のプライバシーの保護を行う。
製品名	Amazon Web Services (AWS)	Amazon Redshift (Amazonのデータウェアハウス)	Amazon Athena (Amazonのデータ分析サービス)	Amazon S3 (Amazonのクラウドストレージ)	Amazon IAM (Amazonのアイデンティティ管理サービス)
製品概要	Amazon Web Services (AWS)は、クラウド上の様々なサービスを提供する。AWSは、世界中の様々な地域にデータセンターを有し、世界中の様々な企業にサービスを提供している。	Amazon Redshiftは、Amazonのデータウェアハウスサービスである。Amazon Redshiftは、Amazonのクラウド上のデータウェアハウスであり、Amazonのクラウド上のデータを分析・処理するためのサービスである。	Amazon Athenaは、Amazonのデータ分析サービスである。Amazon Athenaは、Amazonのクラウド上のデータを分析・処理するためのサービスである。	Amazon S3は、Amazonのクラウドストレージサービスである。Amazon S3は、Amazonのクラウド上のデータを格納・管理するためのサービスである。	Amazon IAMは、Amazonのアイデンティティ管理サービスである。Amazon IAMは、Amazonのクラウド上のアクセスを管理するためのサービスである。
製品名	Microsoft Azure	Microsoft Azure Data Lake Storage (Azure DLS)	Microsoft Azure Synapse Analytics (Azure Synapse)	Microsoft Azure Cosmos DB (Azure Cosmos DB)	Microsoft Azure Active Directory (Azure AD)
製品概要	Microsoft Azureは、クラウド上の様々なサービスを提供する。Microsoft Azureは、世界中の様々な地域にデータセンターを有し、世界中の様々な企業にサービスを提供している。	Microsoft Azure Data Lake Storage (Azure DLS)は、Microsoft Azureのデータウェアハウスサービスである。Microsoft Azure DLSは、Microsoft Azureのクラウド上のデータウェアハウスであり、Microsoft Azureのクラウド上のデータを分析・処理するためのサービスである。	Microsoft Azure Synapse Analytics (Azure Synapse)は、Microsoft Azureのデータ分析サービスである。Microsoft Azure Synapseは、Microsoft Azureのクラウド上のデータを分析・処理するためのサービスである。	Microsoft Azure Cosmos DB (Azure Cosmos DB)は、Microsoft Azureのクラウドデータベースサービスである。Microsoft Azure Cosmos DBは、Microsoft Azureのクラウド上のデータを格納・管理するためのサービスである。	Microsoft Azure Active Directory (Azure AD)は、Microsoft Azureのアイデンティティ管理サービスである。Microsoft Azure ADは、Microsoft Azureのクラウド上のアクセスを管理するためのサービスである。
製品名	Google Cloud Platform (GCP)	Google Cloud BigQuery (Google Cloud BigQuery)	Google Cloud Data Analytics (Google Cloud Data Analytics)	Google Cloud Storage (Google Cloud Storage)	Google Cloud IAM (Google Cloud IAM)
製品概要	Google Cloud Platform (GCP)は、クラウド上の様々なサービスを提供する。Google Cloud Platformは、世界中の様々な地域にデータセンターを有し、世界中の様々な企業にサービスを提供している。	Google Cloud BigQuery (Google Cloud BigQuery)は、Google Cloudのデータウェアハウスサービスである。Google Cloud BigQueryは、Google Cloudのクラウド上のデータウェアハウスであり、Google Cloudのクラウド上のデータを分析・処理するためのサービスである。	Google Cloud Data Analytics (Google Cloud Data Analytics)は、Google Cloudのデータ分析サービスである。Google Cloud Data Analyticsは、Google Cloudのクラウド上のデータを分析・処理するためのサービスである。	Google Cloud Storage (Google Cloud Storage)は、Google Cloudのクラウドストレージサービスである。Google Cloud Storageは、Google Cloudのクラウド上のデータを格納・管理するためのサービスである。	Google Cloud IAM (Google Cloud IAM)は、Google Cloudのアイデンティティ管理サービスである。Google Cloud IAMは、Google Cloudのクラウド上のアクセスを管理するためのサービスである。
製品名	IBM Cloud	IBM Cloud Data Warehouse (IBM Cloud Data Warehouse)	IBM Cloud Analytics (IBM Cloud Analytics)	IBM Cloud Storage (IBM Cloud Storage)	IBM Cloud IAM (IBM Cloud IAM)
製品概要	IBM Cloudは、クラウド上の様々なサービスを提供する。IBM Cloudは、世界中の様々な地域にデータセンターを有し、世界中の様々な企業にサービスを提供している。	IBM Cloud Data Warehouse (IBM Cloud Data Warehouse)は、IBM Cloudのデータウェアハウスサービスである。IBM Cloud Data Warehouseは、IBM Cloudのクラウド上のデータウェアハウスであり、IBM Cloudのクラウド上のデータを分析・処理するためのサービスである。	IBM Cloud Analytics (IBM Cloud Analytics)は、IBM Cloudのデータ分析サービスである。IBM Cloud Analyticsは、IBM Cloudのクラウド上のデータを分析・処理するためのサービスである。	IBM Cloud Storage (IBM Cloud Storage)は、IBM Cloudのクラウドストレージサービスである。IBM Cloud Storageは、IBM Cloudのクラウド上のデータを格納・管理するためのサービスである。	IBM Cloud IAM (IBM Cloud IAM)は、IBM Cloudのアイデンティティ管理サービスである。IBM Cloud IAMは、IBM Cloudのクラウド上のアクセスを管理するためのサービスである。
製品名	Oracle Cloud	Oracle Cloud Data Warehouse (Oracle Cloud Data Warehouse)	Oracle Cloud Analytics (Oracle Cloud Analytics)	Oracle Cloud Storage (Oracle Cloud Storage)	Oracle Cloud IAM (Oracle Cloud IAM)
製品概要	Oracle Cloudは、クラウド上の様々なサービスを提供する。Oracle Cloudは、世界中の様々な地域にデータセンターを有し、世界中の様々な企業にサービスを提供している。	Oracle Cloud Data Warehouse (Oracle Cloud Data Warehouse)は、Oracle Cloudのデータウェアハウスサービスである。Oracle Cloud Data Warehouseは、Oracle Cloudのクラウド上のデータウェアハウスであり、Oracle Cloudのクラウド上のデータを分析・処理するためのサービスである。	Oracle Cloud Analytics (Oracle Cloud Analytics)は、Oracle Cloudのデータ分析サービスである。Oracle Cloud Analyticsは、Oracle Cloudのクラウド上のデータを分析・処理するためのサービスである。	Oracle Cloud Storage (Oracle Cloud Storage)は、Oracle Cloudのクラウドストレージサービスである。Oracle Cloud Storageは、Oracle Cloudのクラウド上のデータを格納・管理するためのサービスである。	Oracle Cloud IAM (Oracle Cloud IAM)は、Oracle Cloudのアイデンティティ管理サービスである。Oracle Cloud IAMは、Oracle Cloudのクラウド上のアクセスを管理するためのサービスである。

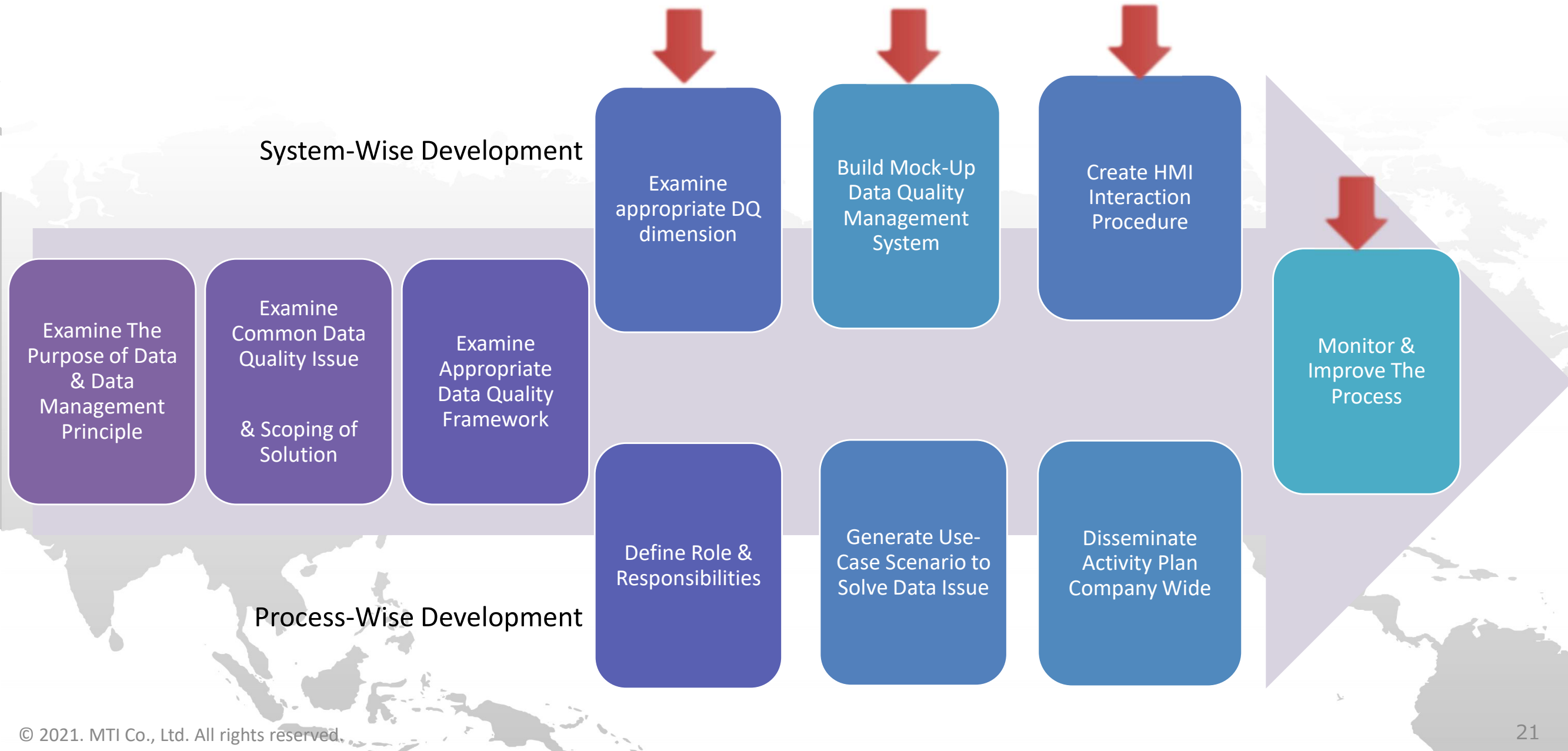
## ▶ Monitoring Flow and Escalation Process



## DMBOK DQM Cycle

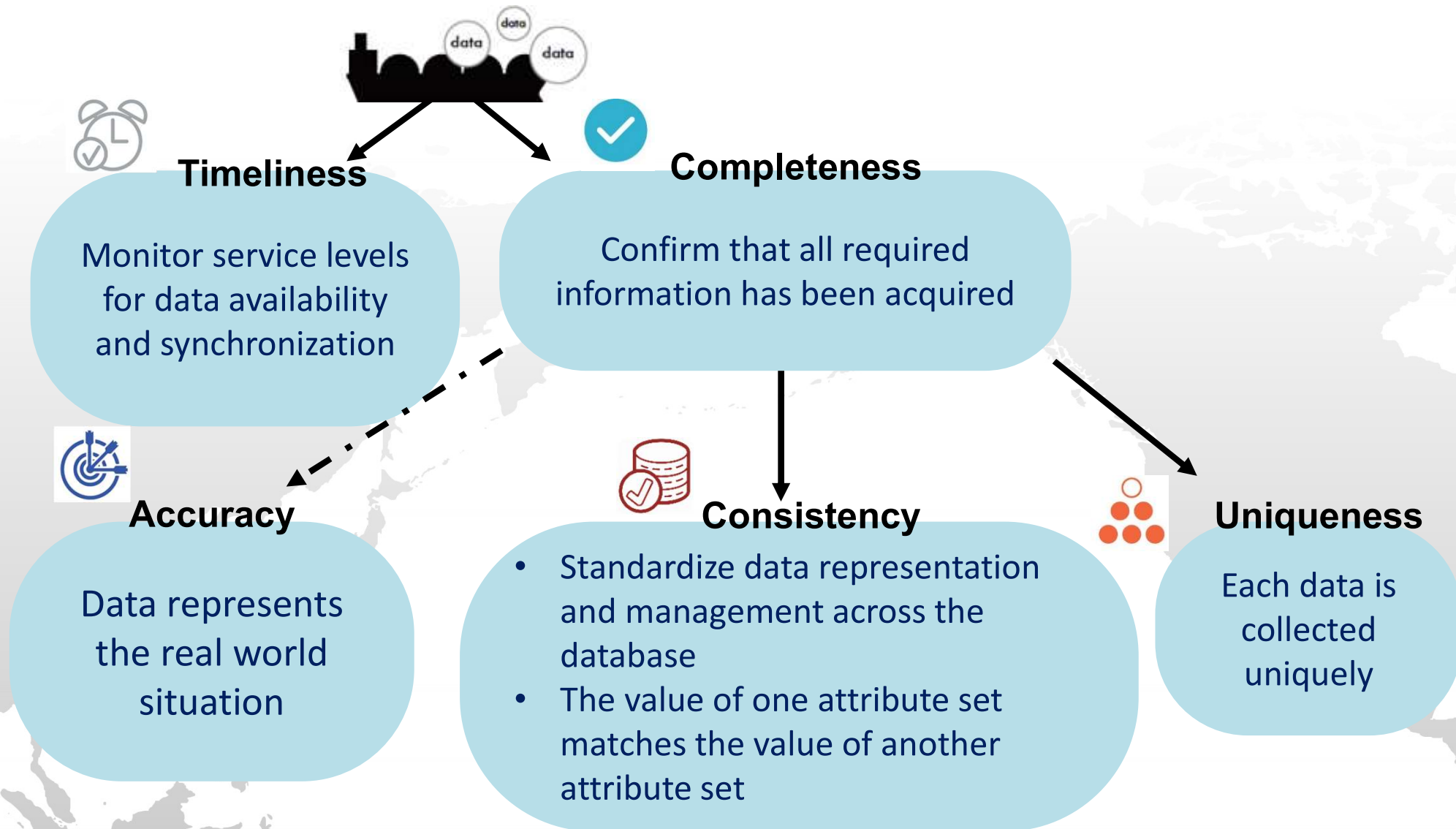


# The flow of NYK IoT Data Quality Management System





# Important DQ dimensions based on Common Issue



# Designed Capabilities of SIMS DQMS

## Data quality Repository

- Storage of definition of data quality rules and metrics that can be modified by business user

## Data Quality Reporting

- Automatically generate report and dashboard that can be utilized by business user
- Identify and support resolution of issue

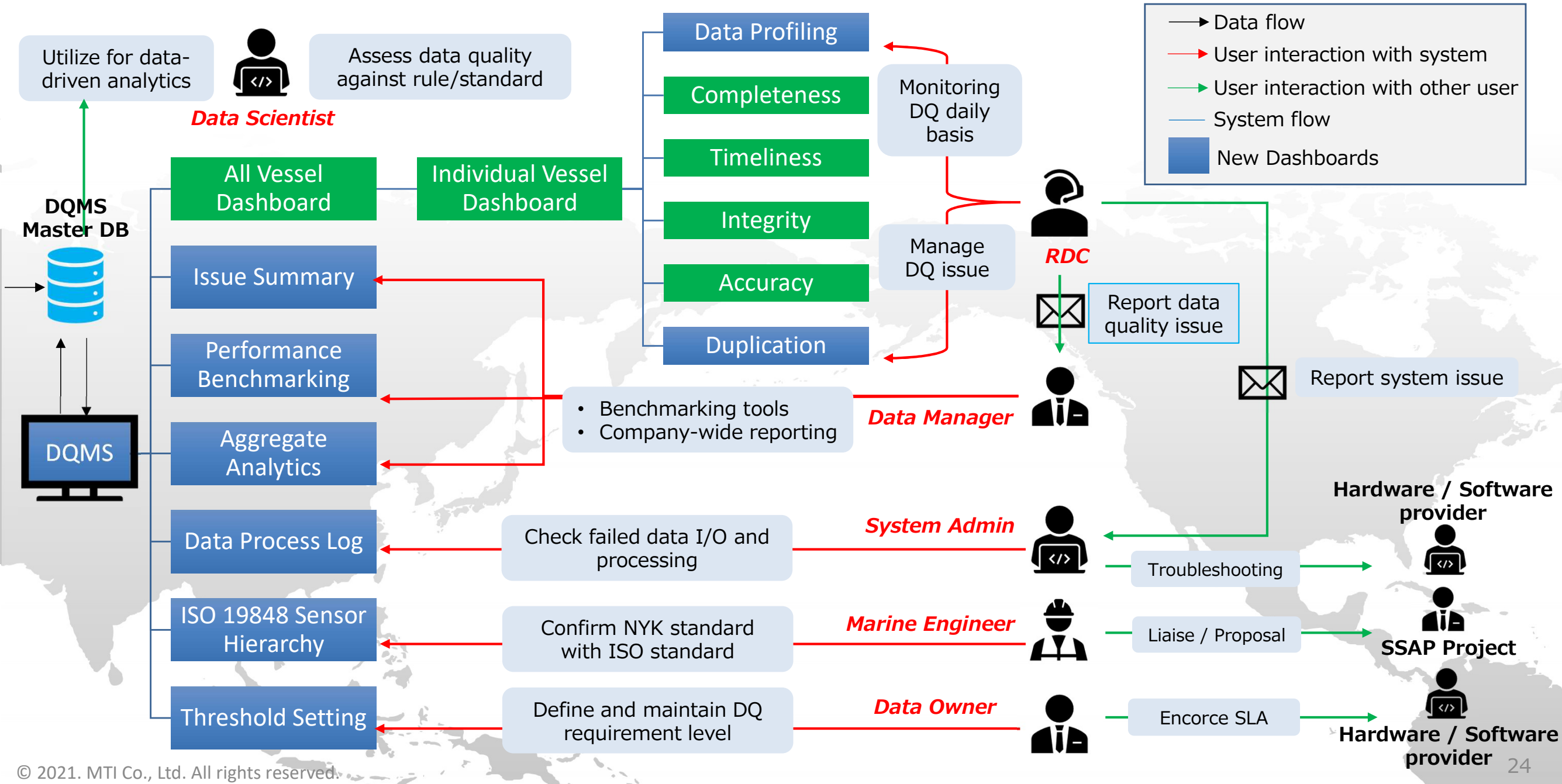
## Data Quality Monitoring

- Automatic calculation of data quality indicators based on event and trigger of data quality issue

## Data Quality Profiling

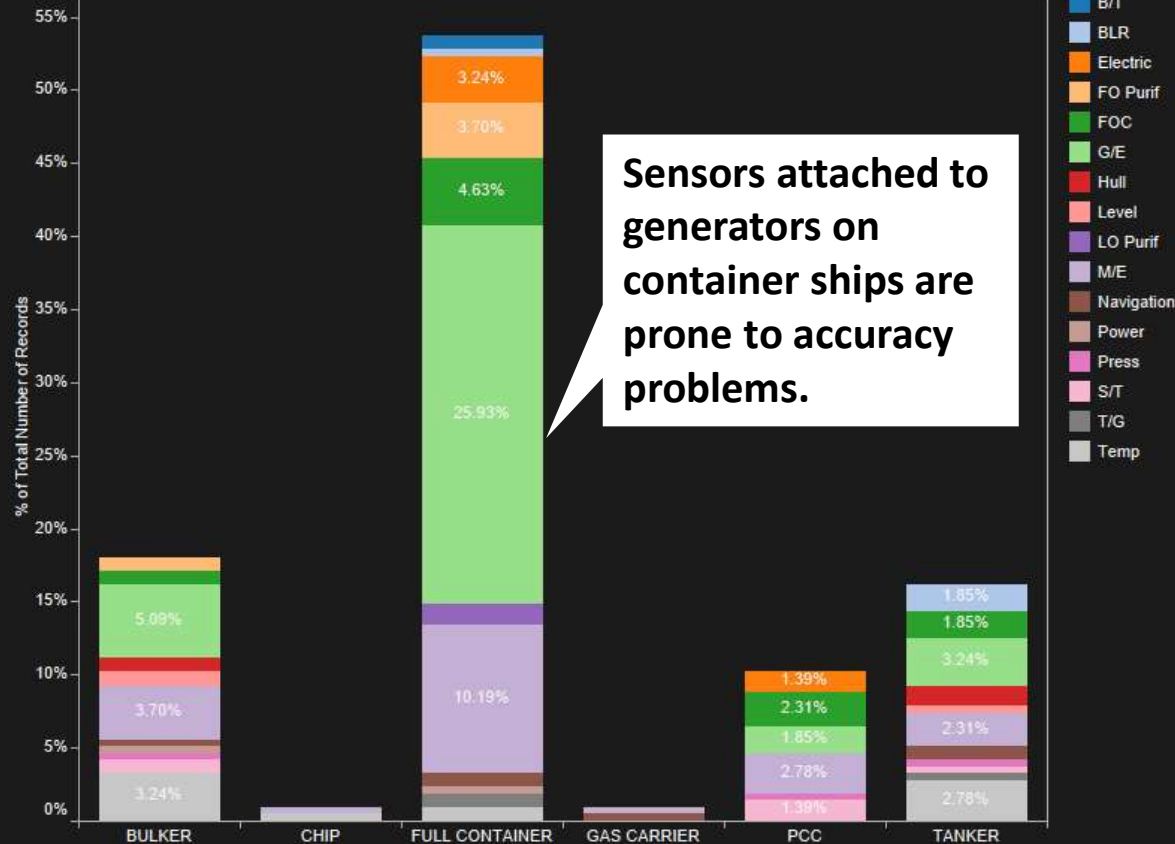
- On demand/ad-hoc analysis of content of data for advance use case (e.g. machine learning)

# DQMS facilitate different stakeholders by various dashboards to do their jobs

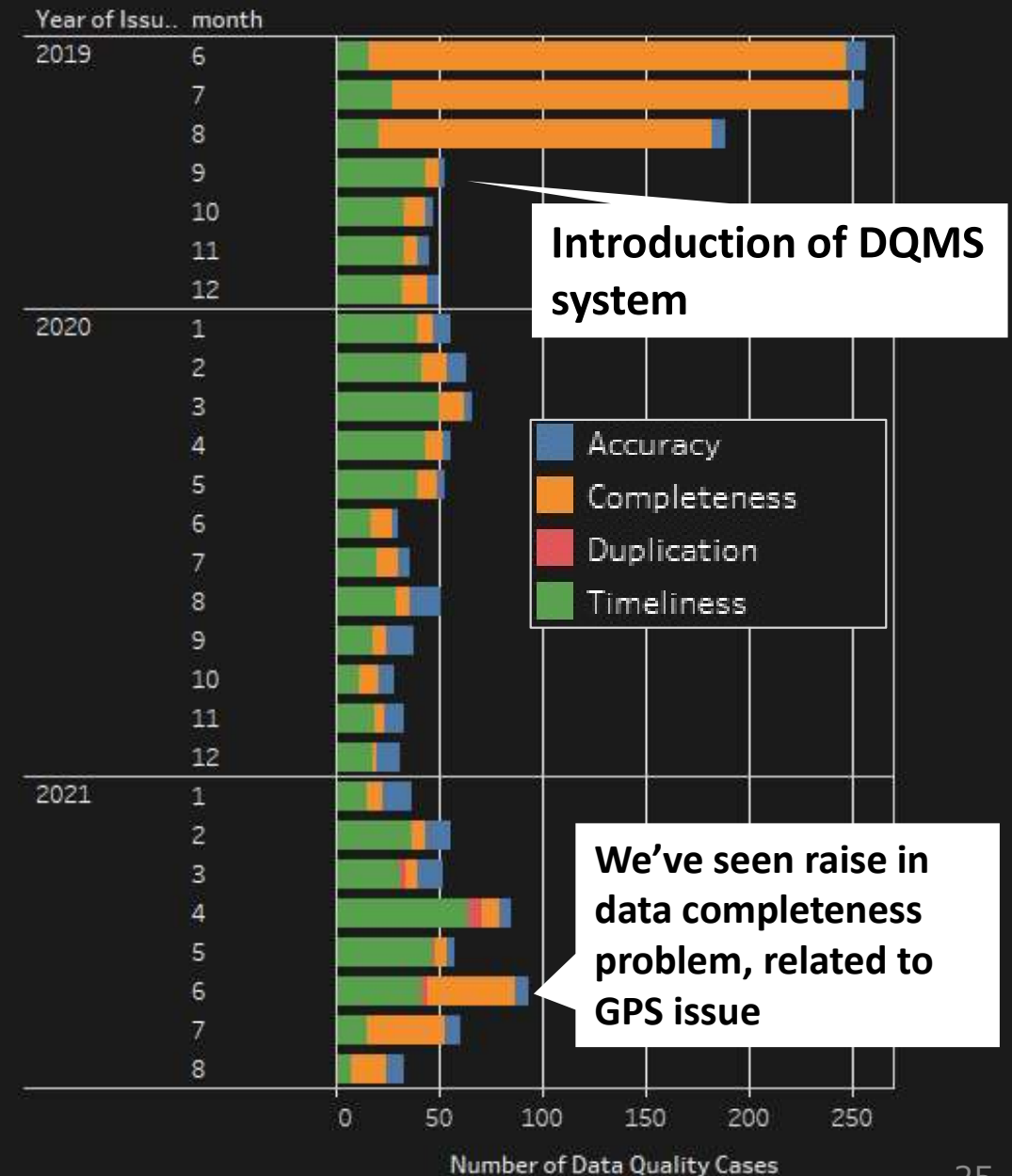


# We gained useful insights for our operations.

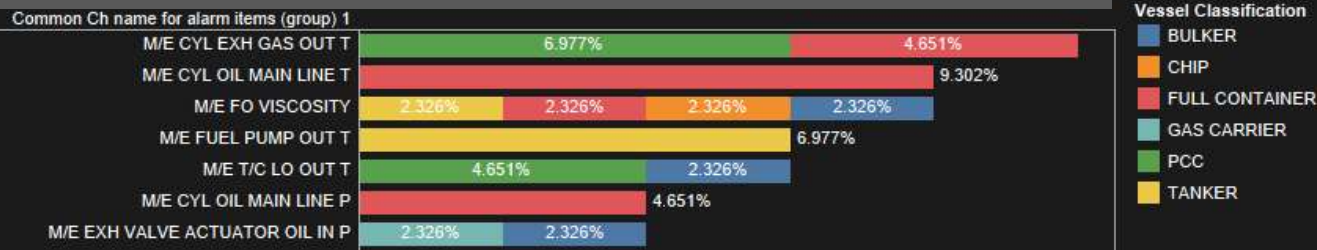
Number of Data Quality Issue by System Component



Trend of Data Quality Issue

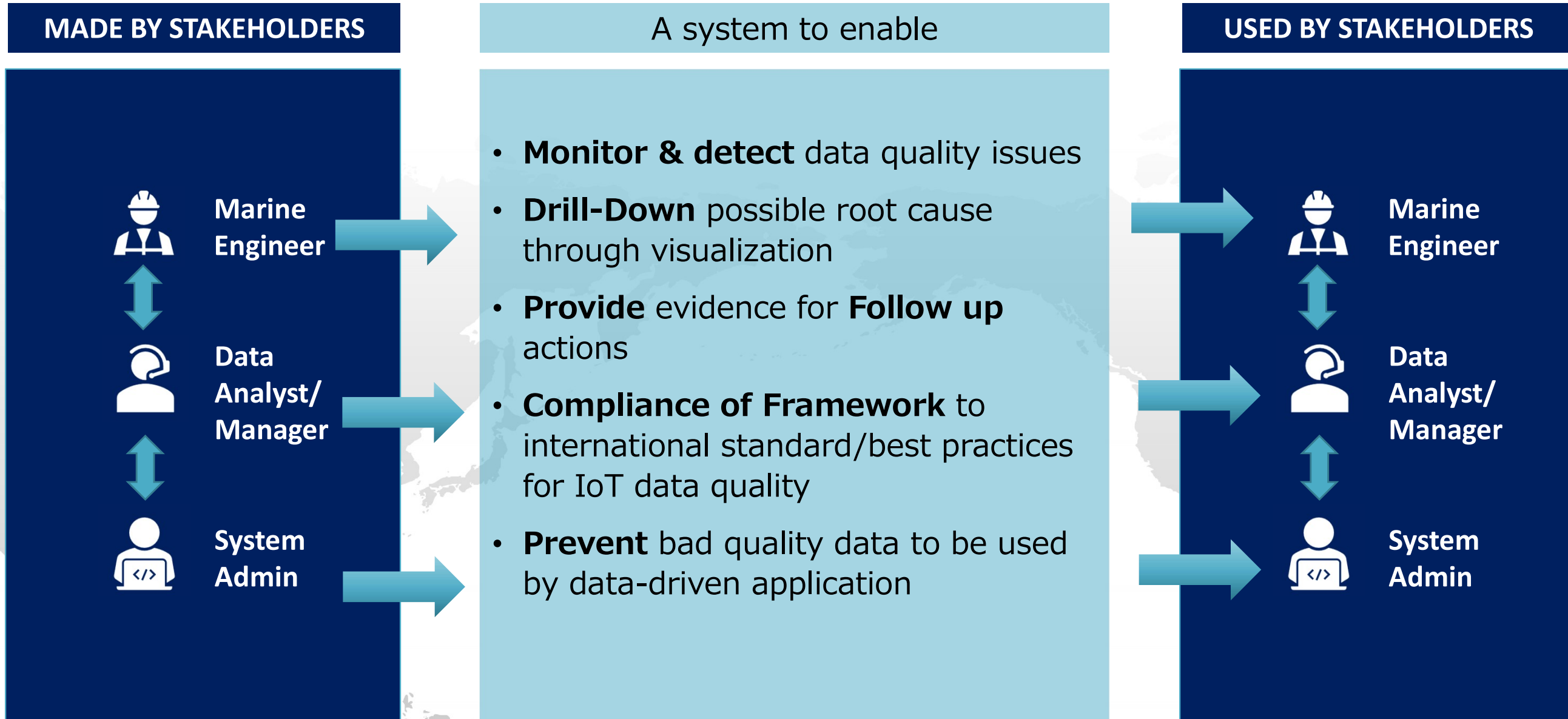


Number of Data Quality Issue by Sensor Name

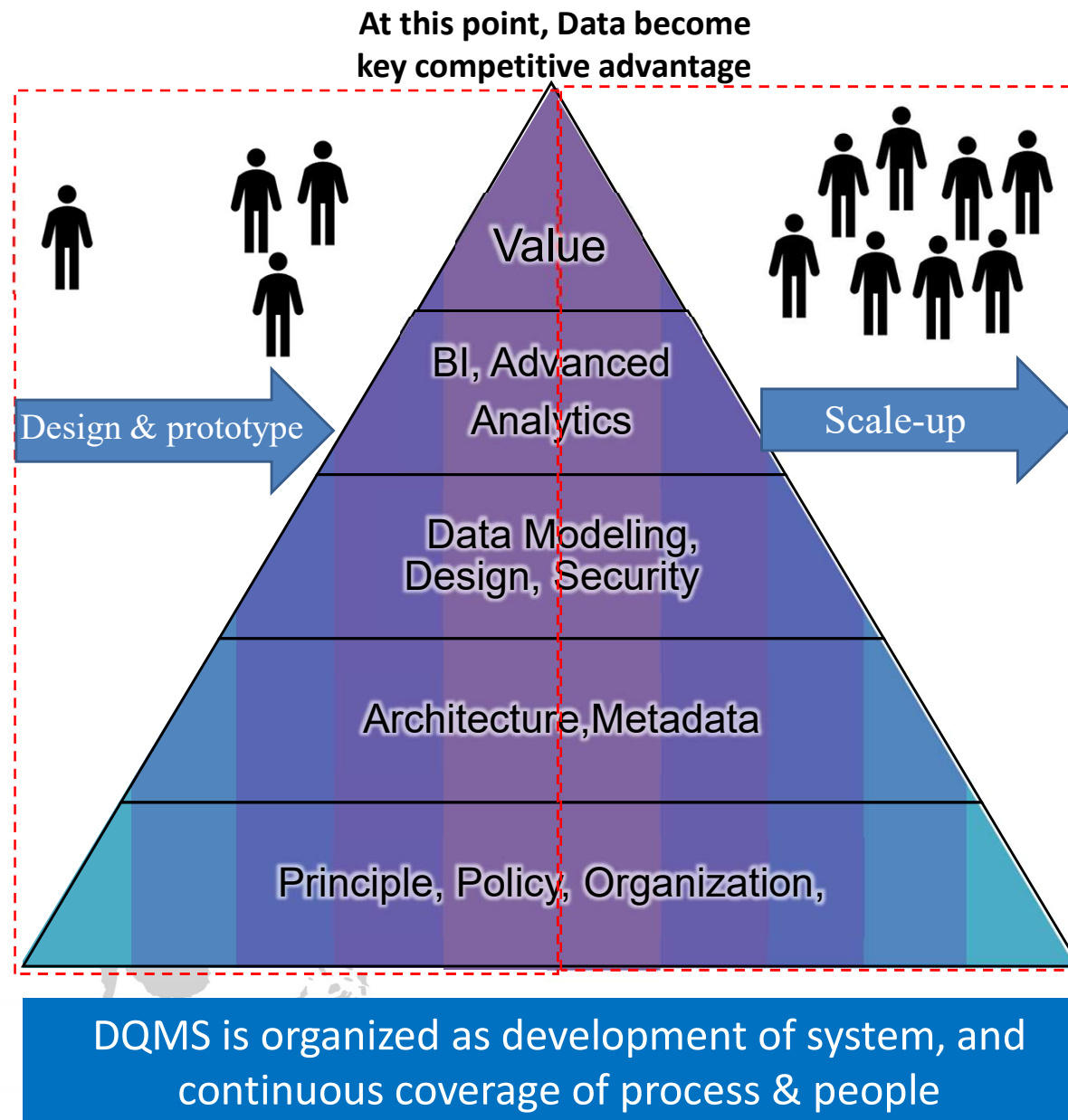




# Data quality management system from our perspective



# The building block of development of NYK DQMS



## Capabilities

### DATA AS ASSET

NYK leverages the benefits of well-managed data and advances its analytic capabilities.

### DATA UTILIZATION

NYK fully enable the advanced practices from high quality data

### DATA INTEGRATION

Get the system functioning within their environment and with their data requires work on integration and interoperability.

### DATA QUALITY

Find and fix challenges with the quality data. Getting to higher quality data depends on reliable Metadata and consistent Data Architecture

### DATA GOVERNANCE

Structural support for data management activities and enables execution of strategic initiatives

Source: Modified from Aiken's pyramid

# Benefit of Building Data Quality Management System for NYK

## SHORT term benefit

- INCREASE **COMPANY-WIDE AWARENESS** OF DATA QUALITY
  - MANAGEMENT COMMITMENT



## MID term benefit

- **CONFIDENCE** OF USER IN USING THE DATA
- **TRUST** IN DATA-ANALYTICS CAPABILITY



## LONG term benefit

- FIRM HOLD ON DATA WHILE CUTTING DOWN EXPENSES
- IMPROVED RISK MANAGEMENT AND **CONTROL OF PHYSICAL ASSET**



**Better customer service**

# Some final thoughts and lesson learnt

- On top of the non existing procedure, there is still a **lack of awareness about the importance of data quality**, and this is often the most important part.
- Our initiative to **start small** and build own system to some extent proven to bring **positive impact** to increase the awareness
- We had started with (finding)**people**,(developing)**process**,(using)**technology**, in that order and then try scale up.
- We are seeking more awareness from maritime industry practitioner to **involve in discussing** how software engineering can be used to manage data quality IoT data and how it should be practiced.
- A standard need to be set together from the **point of view of the industry** with clear examples and sequential steps, so that business can proactively adopt.



**ご清聴どうもありがとうございました。**  
**Thank you for your attention**

