Chair of Petroleum and Geothermal Energy Recovery

Master's Thesis

Investigation of the Digital Platform for Wellbore Centric Data; End-to-End Seamless Integrated Data Flow Concept Development

Felix Waltenberger, BSc

August 2023

MONTANUNIVERSITÄT LEOBEN
www.unileoben.ac.at

AFFIDAVIT

I declare on oath that I wrote this thesis independently, did not use other than the specified sources and aids, and did not otherwise use any unauthorized aids.

I declare that I have read, understood, and complied with the guidelines of the senate of the Montanuniversität Leoben for "Good Scientific Practice".

Furthermore, I declare that the electronic and printed version of the submitted thesis are identical, both, formally and with regard to content.

Date 15.08.2023

____
Signature Author
Felix Waltenberger

Felix Waltenberger, B.Sc.
Master Thesis 2023
Petroleum Engineering

Investigation of the Digital Platform for Wellbore Centric Data; End-to-End Seamless Integrated Data Flow Concept

Supervisor: M.Sc. Peter Berger / Univ.-Prof. PhD Keita Yoshioka / M.Sc. Ameneh Sobhani
Co-supervisor/Advisor: Univ. – Prof. Jim Crompton / M.Sc. Gisela Vanegas Cabas

OMV Energy, Department EDSA-I Integrated Services Montanuniversität Leoben, Chair of Petroleum and Geoenergy Recovery

This thesis is dedicated to my parents

for their unconditional support during my studies.

Acknowledgements

Deepest gratitude goes to the highly skilled and incredibly supportive people at OMV as well as OMV EDSA-1 Integrated Services department, especially Advisor Petrophysics Data Integration and Senior Petrophysicist Peter Berger, who was the main supervisor behind this thesis and from whom I learned countless, valuable insights in different wellbore centric data types, particularly log- mudlog- and core related data. Peter Berger taught me a lot during an internship at OMV even before the thesis project. As the creator of WLMS, without him the thesis would not exist in this form, and he forged the structure and gave direction to the thesis. I want to thank my OMV IPA scholarship mentor Senior Reservoir Engineering Adviser Torsten Clemens, who gave me over the course of two years guidance during my studies on how to develop my career and on career options as a graduate in the oil - and gas - & data management domain. Gisela Vanegas Cabas as former line manager enabled the thesis project and was very supportive at all times.

Professor Crompton at Colorado School of Mines volunteered as an external advisor for this thesis, giving his experience and sacrificing his time for a student he did not personally know prior to agreeing to this partnership. He agreed to generously support me as a foreign student not enrolled at Colorado School of Mines, which is incredible and cannot be stated enough. I want to thank Ameneh Sobhani at Leoben University for her continuous supervision and valuable advice on the thesis.

Many thanks go to Sergey Darovskikh, whose practical lessons complemented my university knowledge on using Techlog.

I want to thank OMV for a) deciding to sponsor me during my Master’s studies at Leoben University of Mining and b) for allowing such industry thesis projects to happen in the first place, putting trust into the students to deliver high quality work, that will hopefully fruitfully enrich both the industry - as well as the scientific university domain in the future.

Abstract

Nowadays, Energy companies face ever-increasing volumes and high variety of data. Wellbore centric data and more specifically WLMS (Well Log Management System) data types are an essential input to be used by OMV subsurface specialists in their daily work. WLMS is the global leading system in OMV and OMV Petrom and is the future planned single data source for all Log-, Mudlog- and Core data within the organization.

The optimization and automation of the existing E2E (End to End) data handling workflow is an ongoing process in OMV. The thesis was initiated to further optimize, automatize and to continue the development of a seamless integrated environment with a potential transition to DELFI and OSDU™ Data Platform. Special focus for the overall concept development was the consideration of all needed WLMS data conditioning workflows, including data gathering, data standardization, data preparation and serving the business as fast as possible with high company value data (standardized, complete, basic QC data) via relevant working environments.

An extensive literature review on well log management, data gathering, and audit was done in advance. Interviews with various specialists have been conducted to understand priorities and pain points. Contact with the service contractor supporting OMV with data conditioning activities was established. Feedback was continuously implemented in this thesis as an ongoing process. WLMS as a data foundation is presented. Fundamentals on DELFI and OSDU™ Data Platform are emphasized on. An approach to establishing a single data entry point is presented: Via a Hybrid Cloud Ecosystem and OSDU™ Data Platform ingestion. A proposal for a data gathering concept and audit process for timely and complete data delivery is presented considering OMV defined data handling processes and regulations. An E2E data flow diagram split into the different WLMS relevant business workflows for holistic data lineage and includes a concept on data tracking and - monitoring. The E2E diagram will allow OMV to translate the workflows existing in the current IT architecture to workflows in an OSDU™ / DELFI ecosystem. Different options for integrating the IT architecture of the service contractor for data conditioning into an OSDU™ / DELFI ecosystem are laid out. Respective advantages and limitations are compared.

A business value proposition shall give the reader information on the context of why wellbore centric data management is important for Energy companies, highlighting how WLMS data types are involved in all stages of exploration, field development, drilling, and production of hydrocarbons as well as in Low Carbon Business (LCB) at OMV. A clear list of actionable recommendations for OMV synthesizes the key findings and how to implement the solutions in the OMV environment. A proposed action plan on a possible way forward concludes this thesis.

Zusammenfassung

Heutzutage sind Energiefirmen mit stets wachsenden Volumina sowie Vielfalt an Daten konfrontiert. Bohrlochbezogene Daten und insbesondere WLMS (Well Log Management System) Datentypen stellen eine essenzielle Basis für die tägliche Arbeit von Fachpersonal in der Erdölbranche dar. WLMS ist das weltweit führende System von OMV, OMV Petrom und affiliierten Vertragspartnern, welches in Zukunft als einzige Datenquelle für alle Bohrlochlogdaten, Bohrspülungslogdaten sowie Bohrkerndaten in der Organisation dienen soll.

Die Optimierung und Automatisierung der existierenden E2E (End zu End) Datenaufbereitungs-Arbeitsabläufe ist ein laufender Prozess innerhalb der OMV. Die Masterarbeit wurde initiiert, um weiter an der Optimierung, Automatisierung zu arbeiten, bzw. an der Entwicklung einer nahtlosen integrierten Umgebung mit einem potenziellen Übergang zu DELFI und der OSDU™ Datenplattform. Spezieller Fokus lag während der ganzheitlichen Konzeptentwicklung auf der Berücksichtigung aller benötigten WLMS Datenaufbereitungs-Arbeitsabläufe inklusive Datensammelkonzept, Datenstandardisierung, Datenvorbereitung und einer möglichst schnellen Bereitstellung von hochwertigen Daten (standardisiert, vollständig und einer grundlegenden Qualitätskontrolle unterzogen) für die Organisation über alle relevanten Arbeitsumgebungen.

Eine umfassende Literaturanalyse zu Bohrlochlogdatenmanagement, Sammlung von Daten und Prüfung von Datenbereitstellung wurde im Vorhinein durchgeführt. Der Kontakt zum Dienstleistungs-Kontraktor, welcher die OMV mit Datenaufbereitung unterstützt wurde hergestellt. Feedback ist kontinuierlich in die Masterarbeit eingeflossen. WLMS als Datengrundlage wurde präsentiert. Grundlagen zu DELFI und OSDU™ Datenplattform wurden erklärt. Ein Ansatz, um allgemein im hybriden Cloudökosystem und in der OSDU™ Datenplattform einen einheitlichen Dateneintrittspunkt herzustellen wurde präsentiert. Ein Vorschlag zum Sammeln von Daten und Prüfung des pünktlichen und vollständigen Lieferns von Daten wurde unter Berücksichtigung der in OMV definierten Datenaufbereitungsprozesse – und Regulierungen eingebracht. Ein End-zu-End Diagramm, sowie ein Konzept für Nachverfolgung und Überwachung von Datenflüssen aufgeteilt in verschiedene WLMS relevante Geschäfts-Arbeitsabläufe für holistische Datenherkunft sind in die Arbeit integriert. Das E2E Datenflussdiagramm wird es der OMV in Zukunft ermöglichen, die in der aktuellen IT-Architektur existierenden Arbeitsabläufe in Arbeitsläufe innerhalb des OSDU™ / DELFI Ökosystems zu übersetzen. Verschiedene Optionen für die Integrierung der IT-Architektur des Dienstleisters für Datenaufbereitung ins OSDU™ / DELFI Ökosystem wurden ausgelegt. In diesem Zusammenhang wurden entsprechende Vorteile und Beschränkungen verglichen.

Eine Nutzendarstellung für die OMV soll vermitteln, warum bohrlochbezogenes Datenmanagement wichtig für Energiefirmen ist und dass bohrlochbezogene Daten in allen Phasen der Exploration, Entwicklung von Feldern, Bohrung und Produktion von Kohlenwasserstoffen sowie im Low Carbon Business (LCB) einen hohen Stellenwert einnehmen. Eine strukturierte Auflistung von verwertbaren Empfehlungen stellt eine Synthese der wesentlichen Erkenntnisse dar und soll helfen, die ausgearbeiteten Lösungen in der OMV-Umgebung zu implementieren. Ein Aktionsplan für das mögliche weitere Vorgehen schließt diese Masterarbeit ab.

Table of Contents

Chapter 1 Introduction

Introduction

Nowadays, Energy companies find themselves confronted with ever-increasing volumes of data. This data not only comprises of historic data such as seismic - or reservoir data of brownfield wells, but also of data that is generated daily through the lifecycle of oil and gas wells including exploration, drilling, and production activities. Mentioned data types are essential for field development planning, reservoir monitoring and production optimizations (e.g., workover). To be able to understand the subsurface regarding subsurface/ reservoir characterization and modelling, reservoir performance and forecasting activities, wellbore centric data types (e.g., log-, mudlog- and core data) are an important input to be used by geoscientists, reservoir engineers, petrophysicists and other subsurface specialists in their daily work.

1.1 Background and Context

Proper data handling and management of historically acquired and newly generated data is necessary to ensure that the data are accessible, quality assured, complete, utilized and meet future needs. The seamless integrability of well bore centric data with other data types, data accessibility for data analytics workflows and further automatization of interpretation workflows is in focus for OMV and many other Energy companies since the last years (digitalization) and gains more and more importance.

OMV and OMV Petrom implemented globally a system, called Well Log Management System (WLMS), to manage log-, core- and mud log data and prepare/ make data accessible for todays and future business needs. It is an important part of a planned, companywide standardized End-to-End (E2E) data flow, see Figure 2. The overall vision/ goals for the WLMS program are illustrated in Figure 1 below.

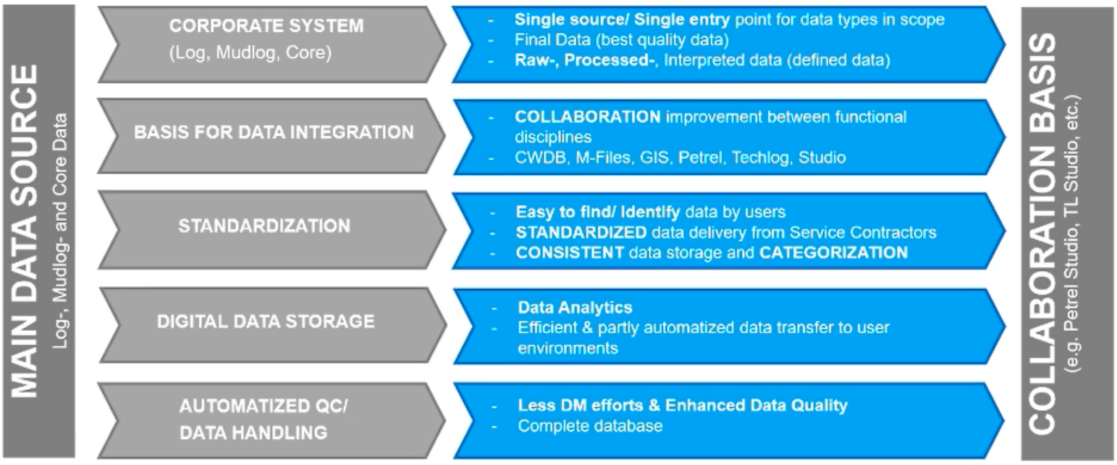


Figure 1: *Overall Goals for WLMS program. (OMV Energy)*

Since system Go-Live, OMV executed numerous data conditioning and loading projects to fill WLMS with business-critical data and make data available for end users. In parallel, the optimization of a seamless E2E data flow (from Service Contractor into end user working environments), further wellbore centric data standardization, adaptions to a modern IT landscape (e.g., Cloud), data flow automatization and optimization is progressing.

The aim of this thesis is to investigate a platform for wellbore centric data with an E2E **integrated seamless data flow** concept development with the ultimate goal of providing a common basis for collaboration of various subsurface disciplines.

The core problem is that in the current state, a single point of data entry, a data gathering concept, application integration, , automatic data transfer mechanisms as well as data volume tracking and auditing are only partly existing or missing (marked in red in Figure 2).

Figure 2 shows the overall E2E dataflow OMV and OMV Petrom is planning to implement. Several components such as Data Conditioning, Data Loading & Validation and Corporate Data Store are already existing but not yet fully integrated (green). Connected components would help to optimize the data flow, increase data quality/ completeness, efficiency (reduced manual efforts and cost) and potential further automatization. For easier readability, the upper section of Figure 2 reappears in certain chapters, to reference each step in the overall E2E data flow with the respective chapter.

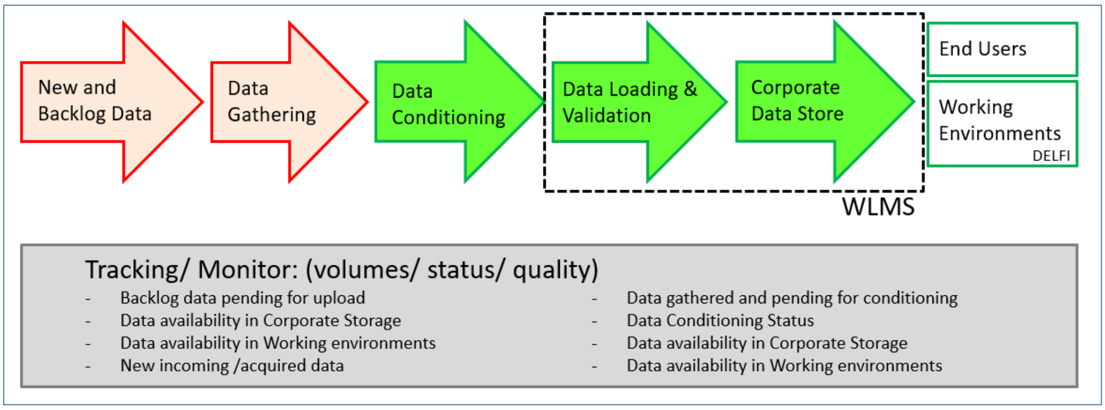


Figure 2*: End to End Data flow from Data Source to Working Environments (OMV Energy)*

E2E Workflow components description and implications:

New and Backlog Data: OMV, OMV Petrom and relevant ventures have several locations where WLMS data types are stored and enter the company. Problem: Missing Single Point of Data Entry and data collection point; overview is missing about data volumes (backlog data and new incoming data) to be loaded into the Corporate Data Store (Final Deliverables versus other non-relevant data).

Data Gathering: Data are distributed over several storage locations, different versions, and several copies. Problem: A data gathering concept is missing which ensures that all incoming data (backlog data gathering, new acquired data and other) are handled in a consistent and standardized manner. It must be ensured that all data are received by the company and entering the leading system according to the WLMS Regulation.

Data Conditioning: Various developed APPS to fulfill data conditioning needs (data classification, data digitization, data QC, data correction, data grouping, etc.). Problem: Applications developed by HOL (Heinemann Oil GmbH) such as: Data Entry, Digitizer are not integrated in the OMV environment.

Data Loading and Validation (WLMS): After data are prepared and ready for WLMS load, data are loaded into the system via a data "Drop Box" and validated following an approval process. Problem: Not integrated into an E2E workflow.

Corporate Data Storage (WLMS): All data are stored in WLMS and made available to the end user via Pro Source (PS) Web. Problem: The current WLMS system, based on PS Web, is an on prem solution and an adaption to the new OMV IT landscape is necessary (transition of PS Web into DELFI/ OSDU™ Data Platform) and must be considered in the targeted E2E workflow.

Working Environments: Only "High Company Value Data" are transferred to the working environments. Problem: Currently there is no automatic mechanism in place which triggers the data transfer or necessary data updates into the Working environments such as DELFI.

Data Flow Tracking/ Monitoring (Audit): Currently only a report is available based on Power BI showing data volumes and status of data loaded in the Corporate Data Store (WLMS). Problem: An overall report is needed showing data volumes, data status and data completeness for the different data processing steps. These reports should help to enforce data availability for end users via a streamlined and audited process within set timelines KPI (Key Performance Indicator).

1.2 Scope and Objectives

Key items to be considered:

- **Concept development of missing data flow components.**
  - **New and Backlog Data:** An overview of new incoming data (operated and non-operated boreholes, etc.) and backlog data located in various locations (OMV and OMV Petrom and its affiliates), structured by data types and business priority.
  - **Data gathering:** A global data gathering concept that honors the WLMS dataset concept and loading standards to be developed. Data completeness is an important aspect to be considered.
  - **Tracking/ Monitor:** A fully integrated auditing concept should be developed considering the different stated components above.
- **Integration of different components, see figure 2 above, into a seamless workflow considering:**
  - **Current and near future OMV and OMV Petrom IT architecture** (e.g., WLMS based on PS Web, DELFI/PTS used as working environment, OSDU™ Data Platform.
  - **Software solution by HOL** currently used for data conditioning, integration, or replacement with other tools.
  - **New proposed components:** Single Point of Data Entry, Data volume overview, Data status changes, Data Gathering and Tracking/ Monitor

Prerequisites:

- Review of the inhouse software resources available at OMV
- Familiarization with existing WLMS Regulations and Processes and other relevant documents
- Overview about data location and current data flows in OMV Head Office (HO), OMV Petrom and relevant ventures

Deliverables:

- E2E Data Flow Diagram:
  - including defined data types
  - including all necessary components
  - data flow reflecting different data statuses (hardcopy material, digital new acquired data, etc.)
- Architecture map including all components, interfaces, and relevant software tools.
- Concept for:
  - Tool to see gathered and new incoming data (wells drilled operated, wells drilled non-operated and other incoming data) considering different wellbore centric data types in OMV ventures and HO
  - data gathering from different storage locations in OMV ventures and HO
  - Component integration: one environment versus different connected environments
  - Overall data flow tracking/ monitoring
  - Automated auditing to ensure all defined data end up in the corporate system and working environments within defined timelines.
- Creation of a business case: What kind of value is generated?
- Possibility of writing a paper about the most important findings

Chapter 2 The Data Foundation

In the context of a digital oilfield, the question arises on which components are important in building a solid data foundation. Data can be seen both as an opportunity as well as a challenge for modern Energy companies. The characteristics of big data (volume, variety of data types, velocity) are accompanied by the term quality of data. (Jim Crompton 2023b, March 6)

Quality data as a corporate asset for wise decisions

It is common sense that assets such as money, hydrocarbon reserves in the reservoir formation or physical facilities are considered objects that are valuable. Nowadays data holds the same value and is becoming an increasingly important, complimentary corporate asset. Even though data matters, the decisions that can be made based on that data matter even more. As corporations face increasingly high data volumes with lots of different types of data (data variety), with a lot less time to analyze it, making the right decisions based on the data is getting harder. Data quality also matters! Even when the best and most advanced artificial intelligence algorithms are being used, but fed with bad data, it will still result in wrong outputs being produced by these algorithms. Data is not just a single persons' effort but rather an enterprise effort. Over time, the understanding of what gives data the attribute of good quality data may change as the complexity of data changes as well. (Jim Crompton 2023b, March 6)

Information intensity

Within the last decades, the amount of data the industry is collecting has increased by 2-3 orders of magnitude. Examples include seismic recordings, reservoir characterization - or field instrumentation & control data. Data variety increases with new data sources and data velocity increases, as the number of sensors and data frequency from field sensors increases. The Energy industry is rich in data but poor in information. Information implies insights and knowledge; however, employees still spend lots of time on manually analysis in spreadsheets and inflexible, static reports. This is an obstacle that must be overcome to build a data foundation and digital twin. (Jim Crompton 2023b, March 6)

Digital Twin

Creating a digital twin involves the challenge of mirroring the physical world in a digital one, i.e., matching the physical world to a digital representation of it. Many critical elements must be monitored, such as vibrations, emissions and even money (costs, value, …). With sensors, one can cover only so much and if the workers' perception of physical condition of equipment, processes and gauge readings is suddenly removed, many operational areas become dark. Sensors only deliver a small glimpse of the real world, like a window which gives the observer a limited view, but not a holistic 360° picture of reality. The value of the digital oilfield (DOF) is not derived from its' existence alone, but rather from the insights and the decisions one can make based on these insights. Also, a digital twin requires complex technology. (Jim Crompton 2023b, March 6)

Technology behind the Digital Twin and the Data Foundation

A digital platform involves more than just a single functional application. It is composed of information from different systems (finance, supply chain, sensors, subsurface, …). Elements of an industrial internet of things (IIOT) digital platform may include:

- Operations technology
- Oil field services
- Enterprise-resource-planning (ERP)
- Information technology
- Engineering technology

Ultimately, the required solution is an IT infrastructure enabling:

- Management of high data volumes with variety and velocity.
- Workflow orchestration capability in an integrated, seamless way.
- Intuitive and simple human interface at the front of the model, despite the inherent complexity at the back of the model.

To fulfill these requirements and manage all the information generated, organizations need software to manage structured data (rows and columns, data in orderly form) such as master data of critical information objects, attributes, units of measure and others. Services around a digital platform may include among others: microservices, application programming interfaces (APIs) and ETL scripts. A data foundation needs a good design. This is achieved via an appropriate information architecture. (Jim Crompton 2023b, March 6)

Information architecture

An information architecture is characterized by:

- Identified, trusted system of record being available for each type of information.

- Information model supporting sharing and integration through workflows.
- Common definitions and semantics.
- Information services delivered through defined data access interfaces. (Jim Crompton 2023b, March 6)

Structured vs unstructured databases

Using a structured query language (SQL) helps handle the complexity introduced by big data. By creating a meaningful model beforehand, one can put data into the right places. SQL is an approach of modelling data in a structured way. However, lots of incoming data is unstructured and requires a suitable information architecture for non-relational databases, such as NoSQL. Storing high volumes of data is possible with SQL but storing a large variety of formats of unstructured data requires a new technology called data lakes, whose concept relies on the principal software NoSQL. An example includes Apache Hadoop. (Jim Crompton 2023b, March 6)

Data lake

- Large volumes & varieties of data, both structured and unstructured in a repository.
- Distributed file structure where data of different types can be stored.
- Analytics can be performed on data lakes by data specialists.
- Staging area for data warehouses (DWH). A data warehouse hosts more carefully treated data for the purpose of reporting and analysis.
- Relies on commodity cluster computing techniques with the goal of massively scalable, low-cost storage of a variety of data formats.

Good accessibility via the front end of a data lake for users, which would otherwise be done in Python or R. (Jim Crompton 2023b, March 6)

Enterprise Data Warehouse

Is an SQL database, which allows for creating, modelling and ingesting data. (Jim Crompton 2023b, March 6)

Data modelling: Semantics

The goal of data modelling is to translate business rules and definitions to technical underlying systems. Semantics is a structured approach with a new data modelling paradigm. It helps describe the meaning of data; definitions, attributes of an information object, relationships between such objects (e.g., How does a borehole relate to a well?). Semantics data modelling is less fixated on the conventional concept of a single structural data model. (Jim Crompton 2023b, March 6)

Data standards

Data standards are the common language between functional oil and gas disciplines and other Energy divisions, which help create more efficient workflows. For simplicity reasons, it is recommended to use the least number of standards as possible. (Jim Crompton 2023b, March 6)

Building an appropriate Data Foundation

- Incorporating volume, velocity and variety to the Data Foundation is essential.
- Good data quality assurance is the basis. Mapping against data quality standards is therefore required.
- Data Governance provides the means to deliver Data Modelling, which in turn scopes and helps prioritize Data Quality, which drives the need for Data Governance.
- A data foundation is not a one-time effort, but rather needs continuous management.
- A data foundation relies on the trust of the users that they are using best data for quality insights into the digital oilfield. (Jim Crompton 2023b, March 6)

Data Quality

Is considered high when the data is demonstrably fit for purpose. (Jim Crompton 2023b, March 6)

Data Governance

The ongoing process of managing and improving data so that it can benefit all stakeholders of an organization. (Jim Crompton 2023b, March 6)

The role of OSDU™ Data Platform as a Data Foundation

With quality data gaining increasing importance in most E&P companies and at the same time increasing information intensity, there is the demand for a data foundation that unifies all disparate data sources for storage of subsurface data into one platform while at the same time breaking data silos. OSDU™ Data Platform is aiming to solve following challenges:

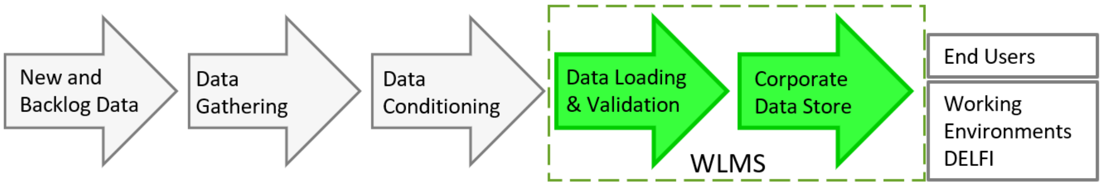
- A single source of truth ("system of record") adds context to the data,
- Removal of data silos and unifying all data formats. Insight can then be won from previously siloed data via analytics, advanced computing, etc. …
- De-duplication of data, which reduces data storage costs,
- Interoperability: Hassle free haring and exchange of data. Enables easy access to usable data, irrespective of its source, enabling them to work on better solutions to solve operators' problems,
- Openness: Adaption of open data standards,
- Scalability: Compatibility with the cloud,

- Extensibility: A data foundation that can be extended and adapted to business rules in a sustainable way.

2.1 WLMS Data Fundamentals

This section explains the technical processes involved in well logging, an explanation of mud logging and coring, a differentiation between Measurement While Drilling (MWD) and Logging While Drilling (LWD), as well as a comparison between different data formats involved in well logging. The full section is located in Appendix 1 - WLMS Data Fundamentals.

2.2 Data Foundations by the example of WLMS



Wellbore Centric data, a corporate asset

OMV WLMS bundles the main data sources: Log-, mudlog - and core data. This data is regarded as a corporate asset and loaded in a corporate system. The ultimate goal of WLMS is to enable making better and faster decisions based on information gained from data as a corporate asset. The data types are currently loaded via one single point of entry, called PS Web into the system.

WLMS hosts three types of data:

- RAW
- PROCESSED
- INTERPRETED

In general, final deliverables from log-, mudlog – and core Service Contractors (external and internal) are stored in WLMS. Parts of the data (fully standardized based on OMV and OMV Petrom standard definitions) are called High Company value data. High Company value data is transferred from WLMS (Data Warehouse) to the working environments, e.g., Techlog Reference Projects or Techlog Studio and Petrel Studio. From there the end users can retrieve the data for their user projects. All published data in WLMS are available for all end users, broad user community (e.g., subsurface specialists, employees in the field, non - Techlog and non - Petrel users, employees in the laboratories).

WLMS includes manual and automatic quality control to ensure that data is submitted following minimum data completeness and quality criteria. The system itself does several automatic data integrity checks and executes quality checks based on defined business quality rules starting from data submission until publishing.

The OMV and OMV Petrom goal is to improve overall data quality and provide high company value data in a standardized data format. Monitoring data quality involves the use of business intelligence (BI) dashboards on data volumes, data status and completeness audits. This is still executed with a high degree of people involvement (manual and time-consuming efforts).

Overall, WLMS acts as basis for data integration and enabling collaboration between functional disciplines (e.g., data exchange between Techlog and Petrel). (OMV Energy)

Well Log Management System is a global system of OMV and OMV Petrom, acting as a Corporate Data Source consisting of different data types:

- **Well log data:** Measurements of in-situ downhole rock properties over depth, conducted via different physical principles and acquired during or after drilling.
- **Core data:** Physical rock samples obtained from a well during the coring process to analyze the rock with different mechanical and physiochemical methods.
- **Mud log data:** During drilling, mud is circulated, and cuttings are brought to the surface. In mud logs, this cutting - and drilling related data is captured and analyzed.
- **Other subsurface data.** (OMV Energy)

Data standardization in WLMS

In terms of standardization, WLMS facilitates:

- One single data source for log-, mudlog- and core data in OMV Classic (including all Branch Offices) and OMV Petrom,
- Quality controlled final deliverables from Service Contractors,
- Global consistent classified and named data based on company defined catalogues.
- Fully standardized High Company Value Data: WLMS enforces that data is treated in a consistent and standardized manner already during the data gathering - (e.g., Service Contractor deliverables) and generation stage (e.g., in-house generated Composite logs and Computed Petrophysical Interpretations - CPIs). This includes backlog data internally distributed across many storage locations and varying versions, but also newly generated data (operated and non-operated wells). (OMV Energy)

Focus on data quality in WLMS data integration and - submittal

All data loaded to WLMS must be assigned to the correct borehole (based on corporate well database: CWDB - Central Well Database), grouped based on the dataset concept (linking all data belonging to a delivery; including digital data and support documents), quality rated reflecting the overall data quality (HIGH, MEDIUM and LOW quality) and the quality information which specifies what the quality rate reflects (e.g. different criteria for digital data and scanned material). High Quality Company Data (i.e., fully standardized data based on OMV definitions) has the highest quality rating. In order to allow a seamless integration, incoming data is submitted via a data drop box. During the submission process, data undergoes an automatized first quality control and is transferred to WLMS staging area. (OMV Energy)

The person who submits the data must select the domain specialist (e.g., Petrophysicist), who is responsible for the first approval step by ensuring that all data ordered from the Service Contractors are fully delivered in good quality and complete. The second approval step is fulfilled by a data manager whose role is to verify that all metadata is populated in a consistent way (correctly named and classified). After the second approval step, data is published and online for end users. If during one of the two quality control processes the data does not fulfill the requirements, data is rejected, and users are notified by email. In total there are 6 data quality dimensions: (OMV Energy)

Submitted data must fulfill the following three data quality conditions: (1) Completeness, (2) Accuracy, (3) Validity (uncorrupted data). For data loaded via loading templates (unstructured data formatted into defined loading templates, which is High Company Value data), checks for whether duplicates exist. However, these duplicate checks do not exist for DLIS, LAS, LIS, TIF or document files. Hence, the person who submits the data types mentioned before must check for duplicates manually. This quality control of avoiding duplicates refers to another data quality standard: (4) uniqueness. Uniqueness is also ensured by using unique borehole identifiers (UBHI). Timeliness (5) and consistency (6) are the other dimensions which are essential to good data quality. (OMV Energy)

Industry standard formats versus unstructured data

- **Unstructured data** in general includes various file formats (e.g., PDF, PTS, .doc, .xls, .txt files) and non-standardized file content (e.g., excel files from different service contractors with varying content format). In OMV and OMV Petrom including all their affiliates, unstructured data exists in various archives and other locations either in digital or analog form. As an example of how unstructured data is handled in OMV to bring it into a usable, end-user friendly format is the digital transformation process from paper logs into digitized, available data usable by end-users in their applications like Techlog or Petrel. They can then be used for further data analytics and interpretation workflows. Another example for unstructured data handling is the formatting of various core laboratory deliverables, e.g., RCA (Routine Core Analysis) results, into predefined standardized loading templates, so that this data can be loaded as structured digital data into the Data Warehouse and further transferred into the working environments. (OMV Energy)

- **Structured data** generally consists of data in rows and columns, making it storable in databases such as SQL. One can distinguish between industry standard structured data (e.g., LAS) and company-specified structured data. For well log data, industry standard formats are already well established, while for core and mudlog data delivery by contractors in structured form is not so common. In OMV and OMV Petrom, log data are mainly received in common industry digital standard formats (DLIS, LAS or LIS). (OMV Energy)

Unfortunately, for **core data** there is no such an industry standard as for well log data. Each core laboratory delivers core plug sample analysis in different ways, therefore consistency between the output core analysis data is missing. For this reason, WLMS contains loading templates to transform this unstructured data into structured, pre-defined formats. (OMV Energy)

This allows to store the data digitally in predefined tables in the database.

WLMS information architecture

In addition, WLMS acts as a digital data storage to enable data analytics with increasingly, partly automatized data transfers to user environments. In terms of corporate data storage, currently WLMS is based on an application called ProSource. Alternative information architectures include DELFI or OSDU™ Data Platform. Technology behind the data foundation relies on services. WLMS uses service apps for data conditioning, such as data classification, digitization, quality control, data correction or data grouping (following the dataset concept). The next step for WLMS is to integrate these apps into a seamless E2E workflow considering the OMV information architecture. (OMV Energy)

The goals of WLMS are summarized in Figure 1. WLMS shall enable the collaboration between different functions, with increased data quality and faster retrieval of data, optimized as well as automated data flows and better data integration by making data available.

WLMS is part of Central Well Database CWDB. CWDB stores primarily: well header information, well deviation surveys, formation tops, DST data, etc... Figure 3 and 4 show both data submission, quality control and loading as well as IT architecture, especially how WLMS is embedded in CWDB and how it is related to PSFO and the Seabed data model. (OMV Energy)

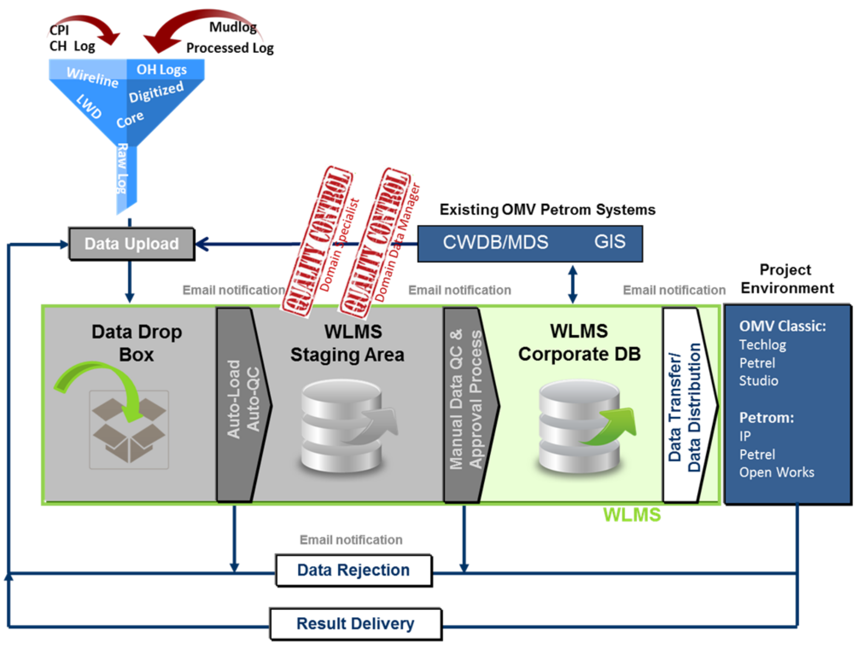


__Figure 3:__ The original process of data upload via the data drop box, quality control, staging, upload to WLMS and data delivery or - rejection. (OMV Energy)

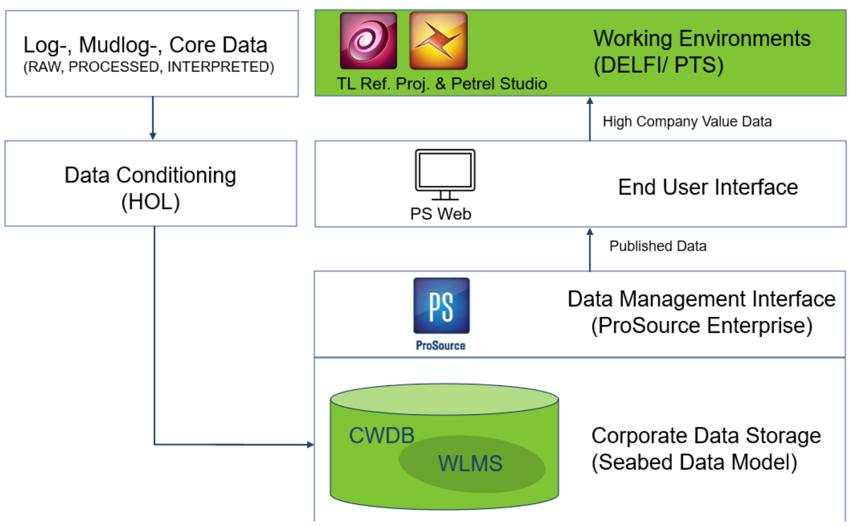


__Figure 4:__ WLMS is part of CWDB, which acts as corporate data storage system. The database is provided by SLB and is based on Seabed data model. ProSource acts as the data management interface, while PS Web is the end user interface. (OMV Energy)

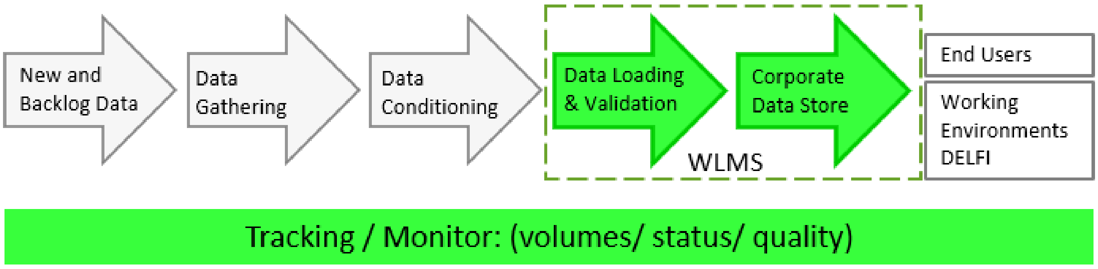
The Dataset Concept

The dataset concept (data grouping) is important as in the past digital data (DLIS, LAS or LIS) was separated from support data (reports, log plots) by storage in different systems/ locations which has an implication that data related to each other are very difficult to find by end users. As an example, CPI (Computed Petrophysical Interpretations) reports and plots were stored in Document Management System, while LAS files (input - and output data) were stored in the legacy log database (LogDB). This caused a lot of difficulties and efforts in finding information belonging to each other. Another example is the separate storage of scanned log plots from digitized las files.

To solve this problem in WLMS, when a Final Delivery Package (e.g., log interpretation, wireline logging run, LWD logging run) is received from the Service Contractor, all files belonging to the delivery are grouped into one dataset and classification attributes (metadata) are added to each file to specify information and context about the specific file. Attaching metadata and the dataset ID together facilitate retrieval of files by the end user. Additionally, metadata supports the setup of automatized data flows and offer the end user to search specific data based on standardized companywide catalogues. Three examples are presented to illustrate the dataset concept:

- A Final Delivery Package of core laboratory (one specific order with specific work scope) is a single dataset including all relevant files, for a delivery package, linked together via the corresponding unique dataset ID. (CORE_12345)
- A Final Delivery Package of a log interpretation data is a single package containing digital input/output data, log plots, reports, etc. linked together via the corresponding unique dataset ID. (LOG_56789)
- A final OMV processed mudlog package is a single package containing standardized mud log data, including gas parameters, drilling parameters, lithology, fluorescence, composite mudlog, mudlog plot. (MUDLOG_13898) (OMV Energy)

WLMS Data Loading Statistics Reporting



The objective of data loading statistics reporting and visualization via Power BI is bridging the challenges of:

- Data type complexity
- Data volumes

These challenges are at the same time an opportunity, as they allow us to tell a data story. Especially the visualization of patterns is an opportunity to work with algorithms and illustrate patterns and stories. Data visualization attempts to dive down into the data and get more detail on what story the data is conveying. (Jim Crompton 2023a, March 6)

Currently used WLMS dashboards and reports are an example for effectively telling a data story. On top of the existing WLMS system, two existing PowerBI dashboards are generated allowing monitoring of: Data loading statistics and data completeness. The three central objectives of the PowerBI dashboards are (OMV Energy):

- Data loading progress monitoring for WLMS: It allows checking on and monitoring the data loading progress to see how many data elements have entered WLMS system (see Figure 5, records loaded by time).
- Dynamic visualization and filtering for specific countries, fields, or wellbores, enabled by prior data classification. For this feature to work, consistent data classification is the key.
- Audit: The overview of data loading activities (statistics and completeness) is used to monitor data loading progress and for checking if the WLMS Regulation (REAL) data loading deadlines are executed as defined. If deadlines are not met, people from branch offices are contacted and reminded about missing data. Currently this is done with extensive manual efforts.

Data loading statistics of WLMS: Countries in which OMV, OMV Petrom and affiliates are active are presented on a map and highlighted in green. A report showing data volumes and the status of data loaded into the Corporate Data Store (WLMS) is shown in Figure 5 below. Data loading into the warehouse is only one step (step four) in the overall E2E Data flow, from data

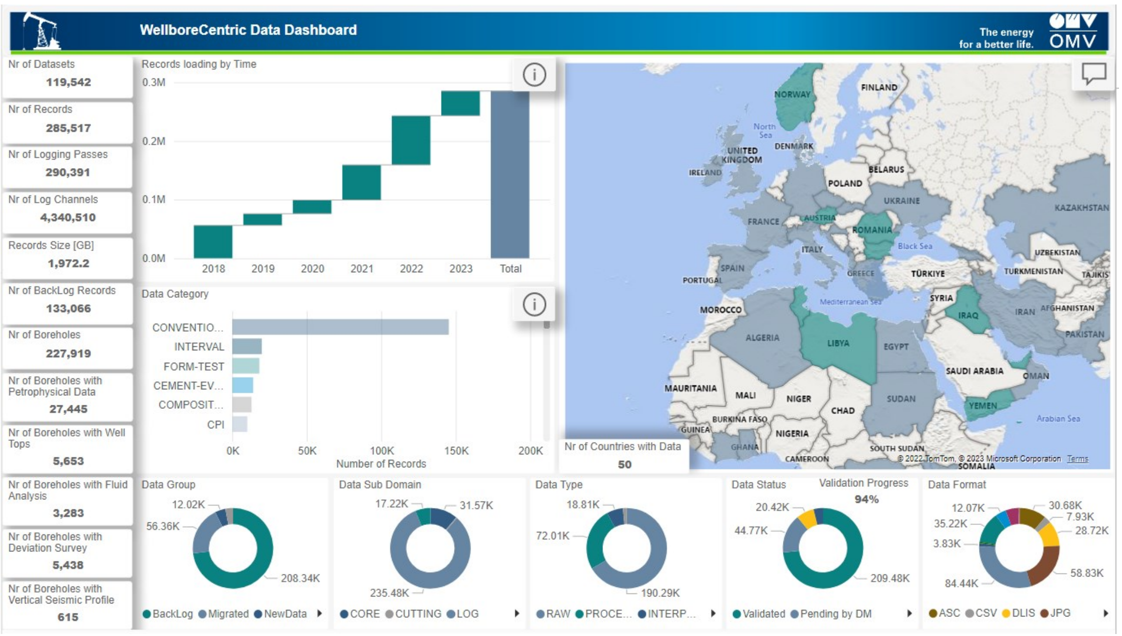


Figure 5: *Wellbore centric data dashboard with an overview of OMV datasets, showing the amount of datasets loaded by time (top), data category, data group (Backlog, migrated, new data, unknown), data sub domain (core, cutting, log, mudlog), data type (raw, processed, interpreted, unknown), data status (validated, pending by data manager), validation progress (%) and data format (ASCII, CSV, DLIS, JPG, LAS, LIS, ODF, PDF, PDS, TIF, XLSM, XML). (OMV Energy)*

source to Working Environments. The goal is to extend the PowerBI dashboard or to make another report to monitor and track the status of other steps in the E2E Data flow.

The following WLMS data elements can be monitored among many others:

- Data Group (Backlog, New acquired, Migrated)
- Data Sub Domain (Core, Log, Mudlog)
- Data Type (RAW, PROCESSED, INTERPRETED)
- Data Status
- Number of files with specific Data Format
- Number of data datasets
- Number of Logging passes
- Number of Log channels
- etc.

Data completeness in WLMS: A second report showing data completeness for recently drilled or selected boreholes of OMV, OMV Petrom including all their affiliates. Figure 6 below gives an exemplary impression of the completeness report which is currently mainly filled with extensive manual efforts. The most challenging part is to verify what data were acquired in the different boreholes and are expected to arrive in WLMS. (OMV Energy)

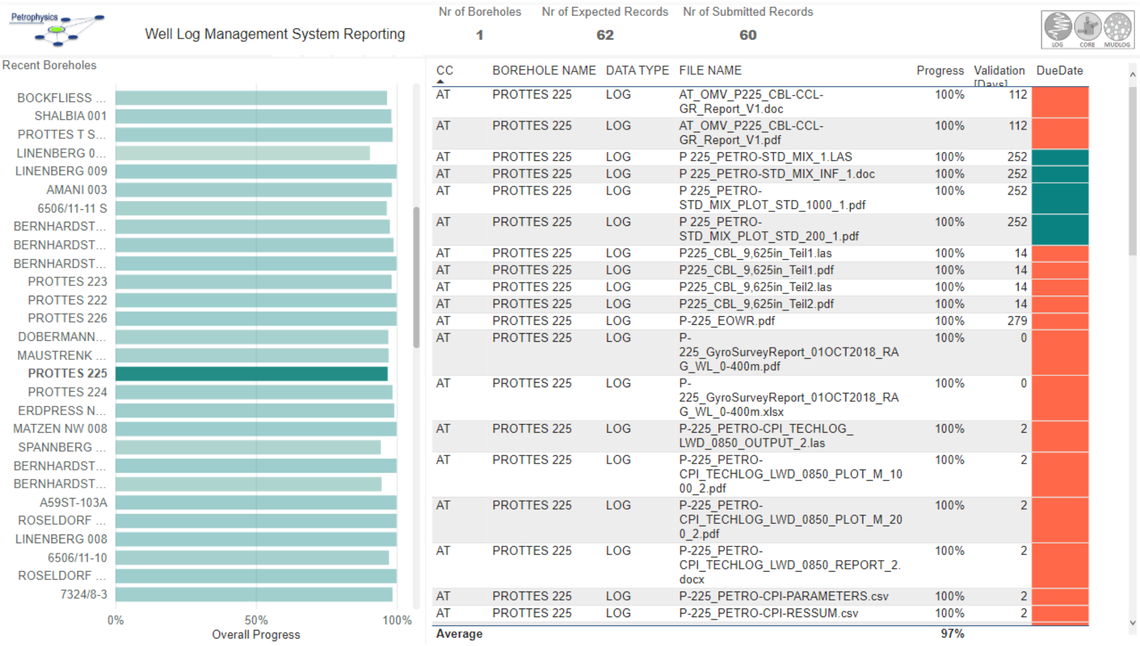


***Figure 6**: Data completeness by the example of Prottes 225 borehole (Austria). In the top section, the number of submitted vs the number of expected records is shown. The columns include borehole location (country code), borehole name, data type (LOG, MUDLOG, CORE, CUTTINGS), file name, progress (%), validation (days) and due date (a status represented by orange or blue color). (OMV Energy)*

These dashboards help users monitor the progress of data loading and completeness of loaded data, while at the same time helping people understand the overall aim of WLMS and the requirement of having complete data as a Final Delivery package in the system. Even more importantly, new insights about the data can be drawn, such as:

Which OMV affiliates from certain countries have problems with data preparation and loading to be able to support the relevant branch office? Identification of missing data to ensure data completeness and follow up of expected incoming data. How did the amount of loaded data volumes change over time? What is the dominant well logging format loaded into WLMS? What is the ratio of raw to interpreted or processed data loaded into WLMS? What is the percentage of different data types (logging data compared to core or mudlog data) in WLMS? Why is data completeness from a specific well low?

Data stories in OMV WLMS dashboards are enablers of new, actionable insights.

Below, an outlook for future visualization feature implementations is presented:

- As a next step, it is planned to implement in ProSource dynamic visualization of well log and core data in the form of cross plots, core images, log tracks, etc. … An integrated, interactive and intuitive data visualization for different data types and formats is needed. The key is the implementation and selection of modern visualization tools which can be easily used by end users with minimum training efforts. The goal is to provide high volume and complex data to end users in the simplest way.

- Currently, data completeness monitoring dashboards are built manually and are very time consuming to feed. Integrating with other existing systems/ software tools (e.g. Drill Plan for OMV and OMV Petrom operated wells) could help to automatize certain steps and therefore reduce manual efforts. Consistency across other systems is another important aspect.

WLMS achievements and pain points

Over the years, WLMS has contributed significant value to OMV, OMV Petrom and its affiliates. Some activities take long term and have a long-term impact whose value is not possible to measure immediately. See also Chapter 5 – Economic Evaluation and especially the section "Data Governance to power data discovery, - measure and – valuation" for more information on the challenge of putting a value tag on data initiatives. Below is a list of key achievements from the last years since the WLMS program was started, reflecting on how WLMS has contributed over the last years to improve the overall situation of wellbore centric data (log-, mudlog – and core data):

Holistic implementation taking most critical components into consideration to make it functioning in the organization.

Key achievements behind WLMS include:

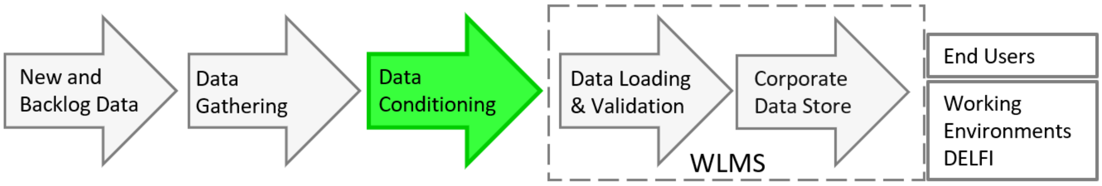
- WLMS business processes,
- WLMS related business regulations (data delivery requirements, standards, etc.),

- Systems design and implemented to support business defined processes and defined minimum standards,
- Software development to support data condition in an efficient way, but still fulfilling business needs,
- Continuous data quality improvements,
- Defining data standards and minimum quality criteria (e.g., High Company Value Data),
- Data conditioning and loading initiatives to get business critical data into the system (singly data source!),
- Execution of audits to ensure that regulation is followed (could be improved as it is a lot of manual efforts),
- All new acquired data must be loaded into the system (operated wells)
- High Company Value Data transferred to working environments (e.g., Techlog and Petrel) after published in WLMS.

Key pain points:

- Not all BOs are following the WLMS regulation and process as requested,
- Data still available at several other locations resulting in the problem that end users are using data from those locations and not WLMS,
- Limited resources (with domain expertise) for data condition projects in HO,
- Limited resources and budget for data gathering activities in BOs. Generally, the buy in from BOs is missing,
- Missing Wellbore centric Data Integration team taking care of all needed activities to lead WLMS further successfully,
- No dedicated resources to fulfill relevant work tasks. People are only provided if there is nothing else to do,
- Budget constrains to start data loading and cleaning activities,
- No budget for business to keep the system properly managed (dictionaries, business rules, system optimizations, data quality control on database level, etc.) and maintained.

2.3 Data Conditioning Fundamentals



As described in Chapter 2.2, WLMS is the company wide database for storing Log-, Mudlog and Core data in OMV, which is based on SLB's ProSource (PS)/ Seabed database software.

In order to enable integrated operations and the digital oilfield, leading databases for key data types play an important role. WLMS is one of OMVs leading systems, storing important data types for interpretation and analysis specialists.

Background and motivation: Pain points

In OMV and OMV Petrom Geoscientists and other specialists still spend a massive amount of their time (~80%) to gather data, format data and prepare data to make it usable for their interpretational work in the respective applications. (CDA) Following examples list some reasons (pain points) for that huge amount of time spent on data preparation.

- Data is located at different locations (several physical paper archives, different share drives, DVDs/CD ROMs, etc.), different formats (hardcopy, digital, etc.), different quality and different data status (preliminary data: field data, real time data, edited data with no documentation about what editing was done, multiple versions with small differences, intermediate data, … as well as final data: final deliverables from contractors, etc.). This makes it difficult to access the data, find the correct data to be used for specific workflows and be able to work with the data in an efficient way.
- Huge amount of data is only available in hardcopy format in the different physical archives as OMV and OMV Petrom has a long-term history of drilled wells. To be able to work in the digital world with those data, data digitization is the key. This digitization must be done in a consistent and acceptable quality with full transparency of origin and quality.
- High volume of digital data is only available on DVDs/ CDs. In addition, a big amount of high important information is only captured in scanned reports or another unstructured format. This information must be extracted, formatted, and loaded to a system so that the data are available for future data analytics workflows and other activities.

- The lack of standard data naming and classification convention complicates in addition to work efficiently with the correct data.
- Data belonging to each other (data belonging to a dataset) are located at different locations and databases (scanned logs and digitized logs)
- Data duplications as data are duplicated in different data sources (Head Office vs Branch Office, etc.)
- Data located in different shared drives or other locations which are not necessarily accessible for the broad user community. Therefore, people are often not aware of the existence of the data.

The goal of the overall WLMS program is that data are stored in a single data storage and in a clearly defined/ classified manner. This ensures that data are loaded following minimum quality rules, that all necessary data are available via one interface and data are easy to find. Additionally, high company value data (most important data) are stored in a structured and standardized way enabling the contribution to the digital oilfield and allow integrated operations in a controlled manner ("compare apples with apples").

To overcome some of the above-mentioned challenges, partnerships between service contractors and OMV play an important role. One example is the goal to establish an efficient workflow for the preparation of all WLMS related data types. It is important that all WLMS standards are honored, to be able to load the data without implications and efficiently into the system. This allows to make data available in a format that can be easily used and integrated with other data types from other disciplines.

Figure 7 and Figure 8 describe the high-level data conditioning workflows from data gathering, data preparation until data is published in the leading system WLMS. From there relevant data is ready for distribution to working environments.

The following four example workflows are an important step towards digitalization and digital transformation. Therefore, such workflows are important to provide data in a format useable for integrated operations and the digital oilfield:

1. **Data gathering and structuring:**
    - Assign data to correct Borehole (UBHI, based on well/ borehole leading system),
    - Group data by data type (Log, Mudlog and Core, Image, etc. ...), group data by Service Contractor final deliverables (complete packages if possible),
    - Data quality (correctness, completeness) following "the dataset concept". (OMV Energy 2021a)

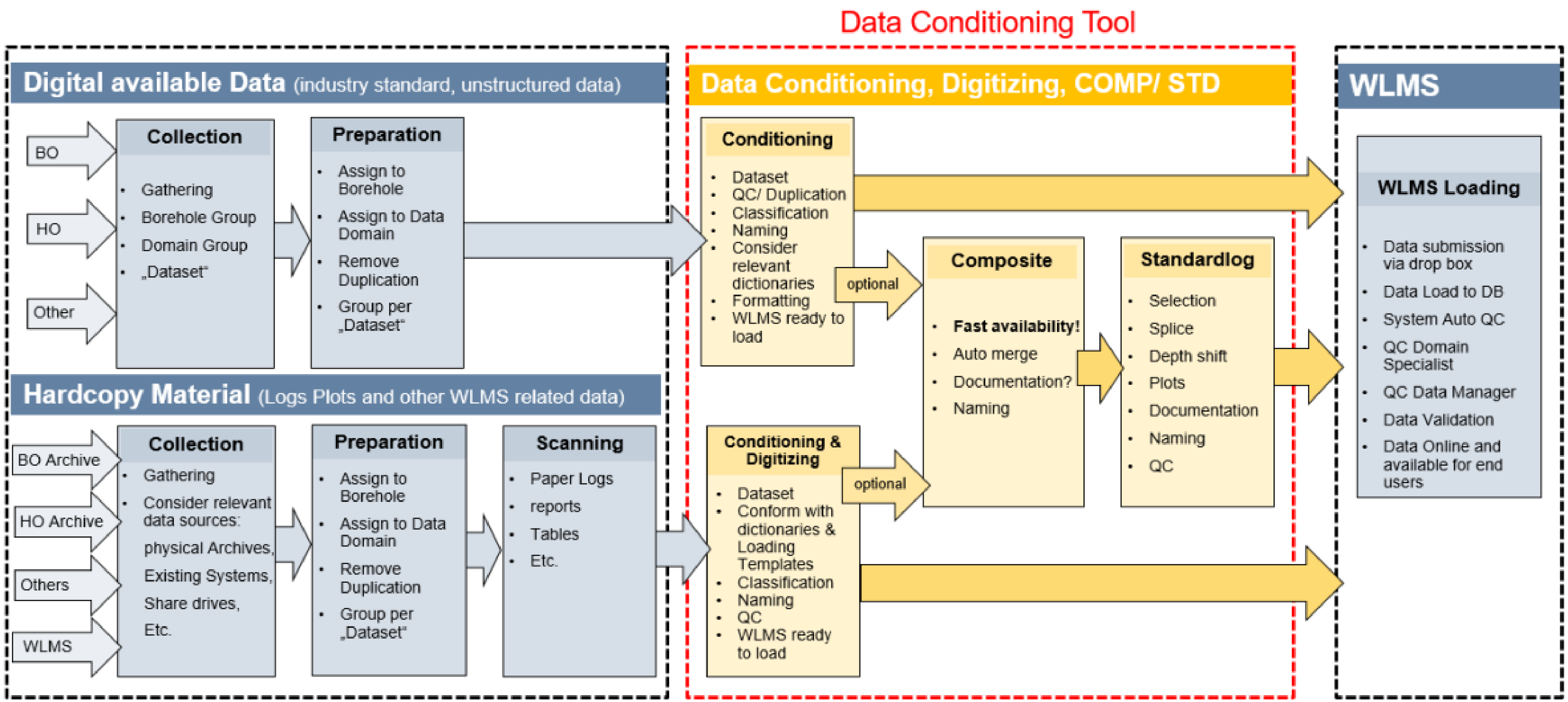


Figure 7: *Data gathering -, preparation - and conditioning workflow to prepare data for upload into WLMS. (OMV Energy 2021b)*

2. **Data Digitization with high quality delivery assurance (logs and unstructured data).**
    - Highest possible data quality following log data digitization rules/ standards: log straightening, digitization overlay, standardized final product including overlay for QC and standardized formats (e.g., las files, loading templates).
    - Long-term traceable results through a unique workflow and quality control process. (OMV Energy 2021a)

3. **Data Preparation, cleaning, structuring and classification and naming for digital data, and digitized data:**

    Basic pre-processing:
    - Ensure data uniqueness: Duplicate removal during data preparation but also considering already existing data in the data warehouse (WLMS);
    - Data grouping based on the WLMS dataset concept;
    - Data QC and correction (supported by automatized processes) to ensure that data are conform with industry standard formats but also OMV/ OMV Petrom defined standards;
    - Classification of digital files (e.g. las, dlis, lis, etc.) supported by automatized information extraction processes and transformation to strictly defined OMV/ OMV Petrom dictionaries;
    - Extract available information from digitized images (e.g. pdf, tif, jpg, ..., etc.) for data classification purposes supported by automatized information extraction processes and transformation to strictly defined OMV/ OMV Petrom dictionaries;
    - Extraction of information from unstructured data files and transformation of data into WLMS loading templates to be able to load data in a standardized and strictly defined digital format to be able to store the information digitally in the database and provide data for data analytics workflows and integration.

4. **(Semi)-Automatic Composite/ Standard log generation**

Figure 8 on the next page describes the data conditioning workflow.

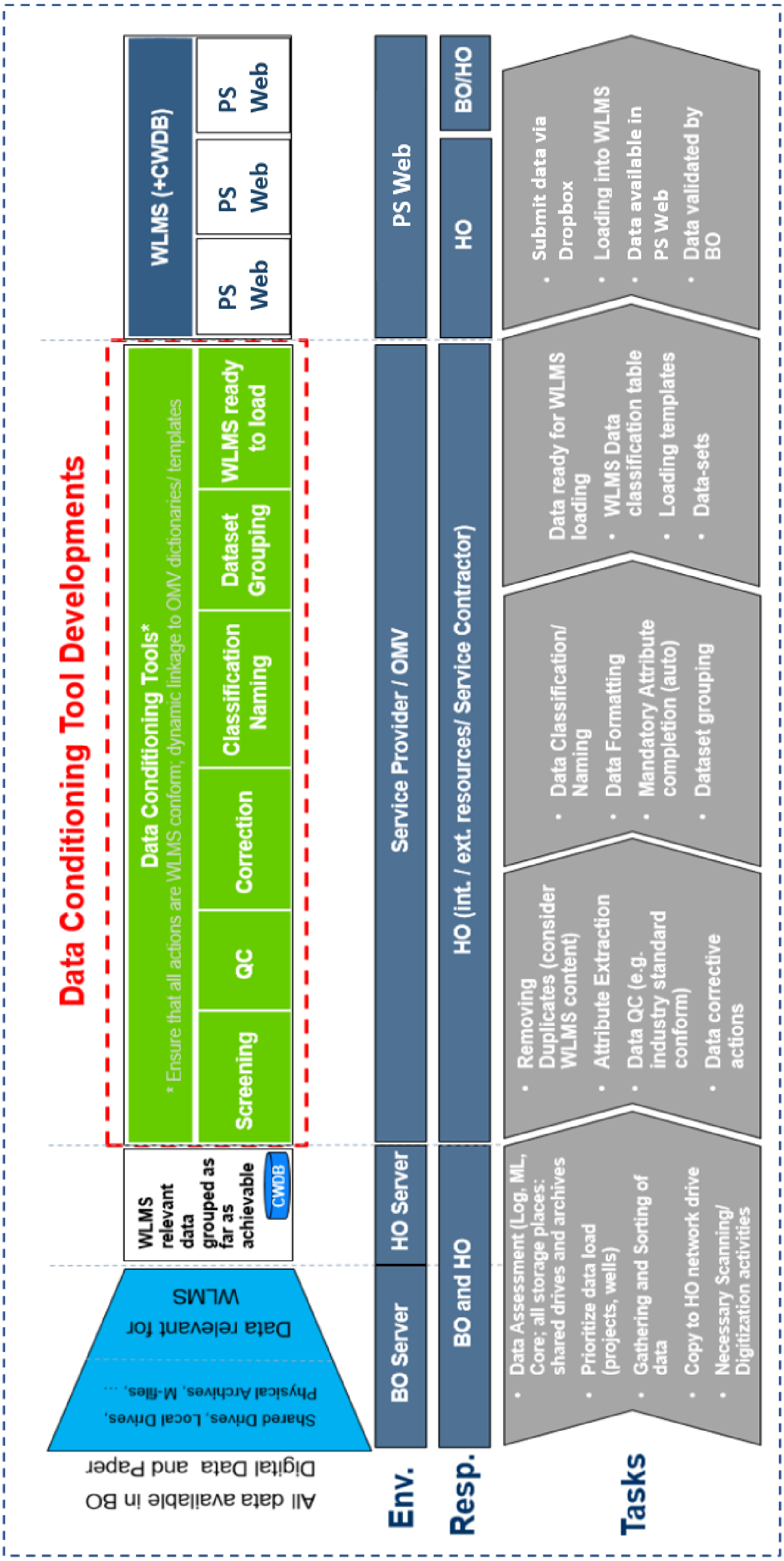


Figure 8: *Workflow for Data Conditioning (OMV Energy)*

Digitization, Digitalization and Digital Transformation

In the context of data conditioning, three essential processes termed digitization, digitalization and digital transformation should be defined, to avoid confusion :

- **Digitization:** Process of creating a digital (bits and bytes) version of analog/physical things such as paper documents, microfilm images, photographs by scanning them. In Digitization, Analog Information/Data is transformed into a Digital Image and Digital Information/Data.
- **Digitalization:** the use of digital technologies and of data (digitized and natively digital) in order to create revenue, improve business, replace/transform business processes. An example for digitalization would be the creation of a Company-wide Database (WLMS), which can be used for further processing within OMV.
- **Digital Transformation:** Creation of High Company Value data for direct usage by end users without additional data formatting or cleaning requirements (e.g., creation of a Composite or Standard log).

(OMV Energy 2021b)

The goals for those four example data processing workflows can be characterized by three categories: Data conditioning goals, digitizing goals, Standard – and Composite log generation goals:

Data conditioning goals:

- Reduction of loading errors in WLMS via semiautomatic data corrections,
- Minimization of data duplicates in WLMS,
- Via strictly defined rules, companywide reference tables and consistent classification/ naming conventions, automatic data classification is facilitated, and end users are confronted with consistent/ harmonized data. This enables the end users to execute company wide data browse or search in a consistent manner and automatized data transfers based on specific attributes.
- Classification loading templates and dataset concept ensure that data is received in a format that is loadable into WLMS and only high company value data are further transferred into working environments (confront end user with most important data and not all available data. This will avoid data misusage/ wrong interpretation, improve interpretation quality/ consistency, and reduce easy avoidable uncertainties). (OMV Energy 2021b)

Digitizing goals:

- Detailed quality control of digitization results, executed by Service Contractors, ensures high quality delivered products and reduces OMV quality control efforts (e.g., by graphical overlay of digitization data with scanned log images (rectified images),
- Naming of digitized curves according to specific rules (following service contractor specific tool and curve dictionaries). This ensures that the overall digitization process follows a consistent way, does not matter by whom the digitization is performed.
- Final LAS-files following OMV's defined standard, fully verified (if necessary corrected) and loadable by PS (WLMS). This ensures full transparence of data source and consistent curve naming convention for specific service contractors. (OMV Energy 2021b)

Standard – and Composite log generation goals:

- Alignment of data selection and data digitizing processes for fast Standard – and Composite log generation,
- Consistent generation of Standard – and Composite logs according to defined rules and aided by automatic processes,
- Automatic generation of info file describing how the Standard log or Composite log was generated and what data were used as data source. This documentation is essential to have full transparency where the data come from and what editing was done to the raw data,
- Possibility to continue work on the former generated Composite log and upgrade the Composite to a Standard log quality level according to the OMV regulation for Composite log and Standard log generation,
- Loading and transfer of standard log datasets into WLMS and further transfer in working environments with minimum manual efforts,
- Consistency (e.g., depth reference, common data quality) in interpretation workflows across all subsurface domains. Standardization will reduce the duplication of work. The work is done once properly, and all disciplines will have a good basis for their work and can use the resources in a more efficient way (not time extensive data gathering and formatting activities). (OMV Energy 2021b)

The high-level workflow from 1) input data to 4) final products looks as follows :

1) **Input data – data review:** Data storage locations & condition, grade of Digitization, necessity of Digitization, data types (structured and unstructured),

2) **Database entry:** Dataset classification, naming according to tools and curves, additional data, selection of digitization,
3) **Processing – Final product creation:** Preprocessing (e.g., rectify), digitization, preparation of non-digitized data, quality control, documentation,
4) **Final products:** Database of log plots, LAS files and other data, WLMS input.

(OMV Energy 2021b)

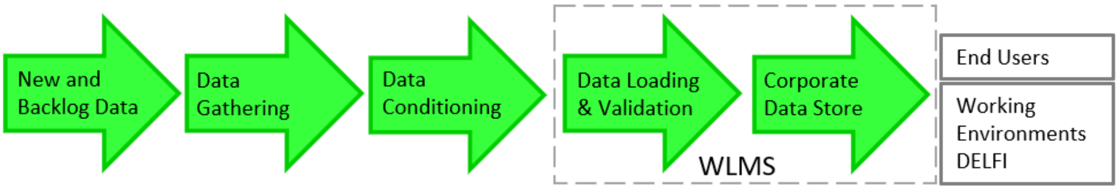
In stage 3) Processing – Final product creation, several steps such as rectification and log straightening are necessary, incorporated into APPS developed by HOL. The APPS presented below are also integrated into the E2E data flow diagram presented in chapter 4.1 E2E Data Flow Diagram. Some functionalities of these APPS are illustrated below:

- **Rectification:** Rectification is necessary due to varying log scanning quality from excellent to poor, sometimes extremely poor, having severe distortions (skewed, wavy, stretched or squeezed) in any direction. Therefore, rectification is applied to de-skew.
- **Log straightening:** Proper alignment of distorted logs is necessary due external effects on paper logs that lead to a decrease in paper quality over time. Therefore, efficient log straightening has been adopted by HOL.
- **Overlay:** The necessity to double-check the quality of the entire log requires the use of overlays. To achieve this, an overlay between Digitalization & External Software is made.
- **Header Extraction:** Header attributes must be extracted from scanned log plots, a process which used to be manually done and is nowadays a semi-automatic header attribute extraction process. Those header attributes are used for log classification, as header meta information inputs in the generated las header and several other workflow relevant product generation processes.
- **Paginated Stitching:** Paginated plots can be joined/stitched. This is important as OMV Petrom split the hardcopy log plot into several separate files during the scanning process. For storing in WLMS and digitizing it is important that the plot is combined beforehand into one plot. (OMV Energy 2021b)

Regarding stage 4) Final products, WLMS input is achieved with the support of the service contractors' tool of data entry:

- **Data Entry:** a database that holds all relevant information in alignment with CWDB well/ borehole header information and WLMS reference tables and dictionaries. The output includes standardizes file names, standardized and QC‘d LAS files loadable into WLMS (PS) and OMV naming conventions for Tools and Curves. Additionally, all data can be extracted following the dataset concept and WLMS loadable format (including the classification sheet and files following a standard naming convention in alignment with the classification attributes). (OMV Energy 2021b)

2.4 OSDU™ Data Platform Fundamentals



The mission of OSDU™ Data Platform is defined as follows: *"The Open Group OSDU Forum delivers an Open Source, standards-based, technology-agnostic data platform for the energy industry that stimulates innovation, industrializes data management, and reduces time to market for new solutions."* (OSDU)

The definition of OSDU™ Data Platform above can be broken down into its' key philosophies, which are summarized below:

- Data platform: OSDU™ Data Platform is neither a data storage nor a data management application, but a data platform. According to Einar Landre, *"Data platforms store, manage and serve data to consumers while adhering to the governance rules defined for the data."* (Landre 2022, November 29)
- Agile governance: ability to adapt to other branches apart from E&P and scalability.
- Open source: to facilitate broad access and collaboration of companies across different industries.
- Standards based: to ensure interoperability.
- Data type support: development and support of new data types.
- Speed: enabled real time data capture and streaming.
- Technological flexibility: agility through microservices that OSDU™ Data Platform is based on, full backward compatibility of APIs and technology agnostic: Cloud, On Premise and Edge.
- Cost effectiveness: with a changing energy landscape, systems most become more efficient to save costs.

It is designed to function as a single System Of Record (SOR), meaning that it is the single, definitive (leading) data source. OSDU™ Data Platform is designed to remove barriers between vertical data silos, planning and operations, operators, and suppliers as well as past and future knowledge.

(Epam b), (Epam a)

OSDU^{TM} Data Platform Components

Figure 9 shows the main OSDU^{TM} Data Platform software components.

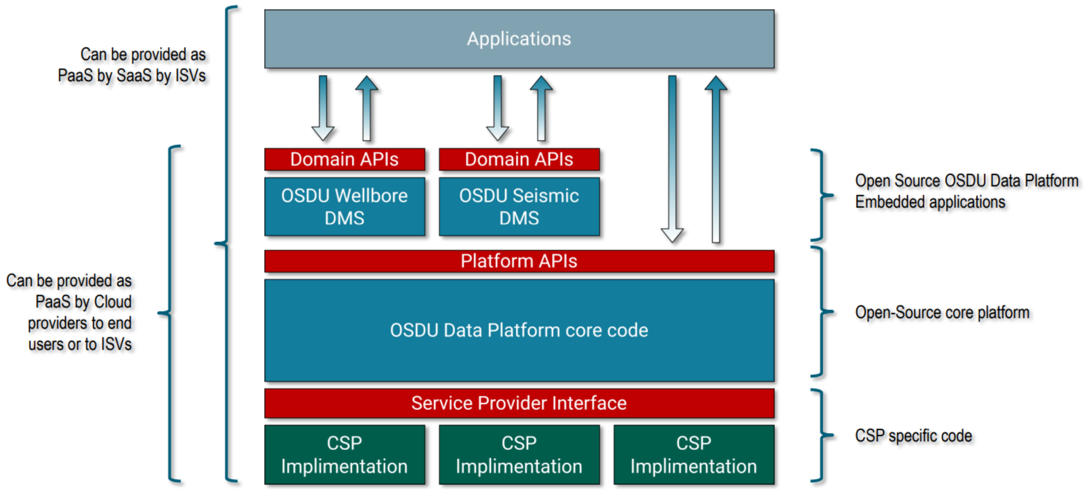


Figure 9: *A description of software components present in OSDU ^{TM} Data Platform. It has been designed in a way to provide maximum portability, flexibility and usability for end users and cloud service providers or ISVs. (Epam a)*

OSDU^{TM} Data Platform software components are structured from bottom to top as follows:

- **Cloud Specific Code (CSP) Implementation:** Optimization of code for different cloud service providers with the goal of improved portability. Portability is a measure for how well one can transfer applications between cloud service providers.
- **Service Provider Interface (SPI):** Optimization for the respective cloud infrastructure by various cloud service providers. SPI is the separation boundary between cloud implementation of service providers and the common cloud of OSDU^{TM} Data Platform. This configuration facilitates optimization of each cloud service providers' particular cloud infrastructure.
- **Core code:** The core code contains all microservices that may be accessed and that are exposed through the APIs. The core code is technology - as well as data agnostic. Implementations optimized for different domains, such as DMS (Domain Management Systems) are based on the APIs.
- **PaaS vs SaaS:** Access to OSDU^{TM} Data Platform can be either obtained through platform as a service model (cloud provider) or through a software as a service model (software providers). OSDU^{TM} Data Platform provides flexibility in users accessing the data platform as well as flexibility and usability via DMS for developers via access to specific data types.

  (Epam a), (Epam b)

OSDU™ Data Platform Principles

The four OSDU™ data principles are:

I) **Value All Data:** Knowledge capture and provenance are central elements of OSDU™ Data Platform. Provenance is information about the source and history of data as it evolves. This ensures that content and context of data ingested into OSDU™ Data Platform are conserved, which relates to the concept of data immutability (data cannot be modified or deleted).
Deleting data content at a present point in time may be problematic, as this data might have relevance for usage in the future.

II) **Minimum Viable Governance:** Ensure that data is classified, discoverable, consumable. Also, data may be identified from multiple data sources. OSDU™ Data Platform is very open towards new data types, as long as this data is classified and described as well as an underlying governance to have the right usage of the data at a future point in time.

III) **Data Security:** Ensure that there is access control in place for data and that there is appropriate data governance, policies, and regulations in place for the right use of data.

IV) **Continuous Data Improvement:** Data lineage and tracking, which is related to I), meaning that data versioning is done as metadata or data is improved. Continuous data improvement also evolves around updating the context of improved, new data as well as two modalities for data access: either through Domain DMS APIs for optimized access (usability) or generic APIs (as it was brought into the system) for normal access. (Epam a), (Epam b)

Principle I), value all data has been illustrated by a rock example, which is presented in Figure 10. A rock obtained from the field has certain physical properties, such as mass, density, the date that it had been digitized, etc. … The rock could be digitized by means of photogrammetry and stored as a 3D object in a database. The rocks' samples can be classified, and the classification may be stored in database tables. It becomes evident that a single rock has turned into an almost unmanageable volume of independent digital artefacts. (Landre)

OSDU™ Data Platform has been designed for storing, managing, and bringing such volumes of data to the data consumers. OSDU™ Data Platform is split into a content store made up of files, a catalogue store consisting of documents and an elastic search engine. (Landre 2022, November 29), (Landre)

In Figure 10, the cycle from data to wisdom is illustrated. The content and context of a rocks' data is preserved, as information, knowledge and wisdom are drawn out of it. Originally stable content is updated continuously by new information, and new insights are derived. For example, if at a later stage a rock sample is classified as eclogite, this knowledge is applied and implemented as wisdom into a newly updated model. OSDU™ Data Platform can be described as a subsurface knowledge capture platform. (Landre)

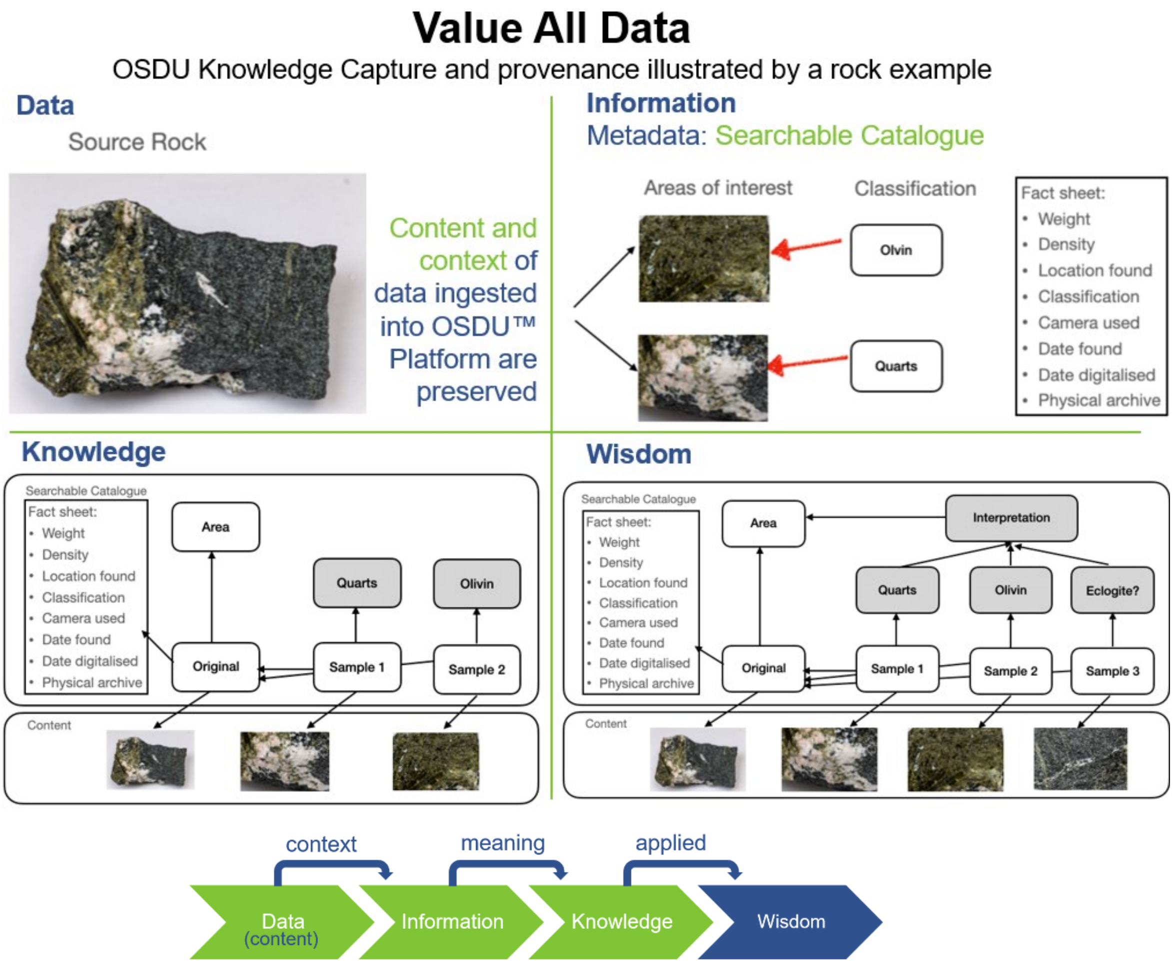


Figure 10: *The relationship of wisdom (applied knowledge), knowledge (information with meaning), information (data in context) and data described as knowledge capture. OSDU™ Data Platform is attempting to extract wisdom from data and making data searchable via a catalogue. Maintaining a version history of data is a concept called provenance and a central part of OSDU™ Data Platform. (Landre 2022, November 29), (Landre)*

When applying the concept of OSDU™ Data Platform catalogue store and content store to the situation of managing, storing, and making well logs retrievable by data consumers (end users), a graph such as in Figure 10 helps to understand. It illustrates how a relationship between a well, wellbore and the obtained well logs, cuttings and cores can be created and how it relates to other data such as seismic.

A searchable catalogue specific to the domain contextualizes well logs, seismic images, and core data. Analyzed data and knowledge obtained from e.g., horizons and makers in turn leads to updated data (insights).

After a seismic campaign, during which seismic surveys are generated, new wellbores are drilled, well logs are generated, new raw data is added and subsequently newly derived knowledge. Now let us assume that another new seismic campaign starts that spans a larger area or uses new advanced technology, such as 3D seismic. Then the old data from the original seismic is kept and enriched / updated by the new advanced seismic data.

Figure 11 illustrates two of the central capabilities of OSDU^{TM} Data Platform:

- Provenance through data lineage
- Data immutability

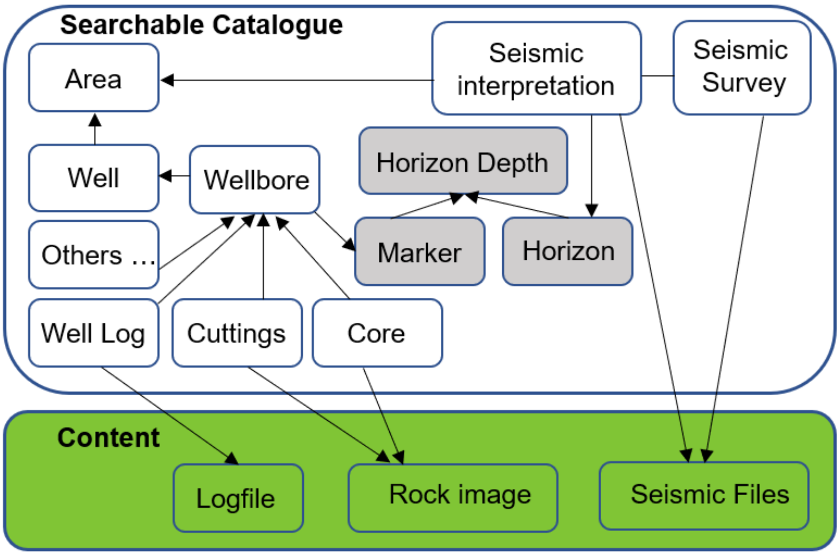


Figure 11: *A catalogue store can be searched and holds well data in a certain area. A content store holds all files such as log files, seismic files, or rock images. Together they form the characteristics of OSDUs' philosophy: classified, discoverable, consumable. (Landre 2022, November 29), (Landre),*

"Lineage enables provenance, basically that we know what data was used as source for a horizon and marker. Provenance is the key to trustworthy information. As new data emerge, old insights are not deleted but replaced by a new instance that is linked to the previous instance. This mean that the platform can hold multiple copies of the horizon and markers […] and capture how the understanding of the underground has evolved over time." (Landre)

User defined "kinds" are representation of data in OSDU^{TM} Data Platform. "kinds" are structured as follows in Table 4:

Kind name	Description
Namespace	Organizational instance
Definition Authority-Source	The organization with authority over definition of data, as different companies use different data model definitions
Group type	Helps classify data by type of data. E.g.: master data, reference data
Individual type	Specific type. E.g.: Well log, 3D seismic survey

Table 4: *Description of different "kinds" (OSDU ™ Data Platform specific term) including some illustrative examples. (Epam a)*

Three operating overlapping zones exist in OSDU™ functional architecture and are presented in Figure 12. These three zones are: Load and Ingest, Process and Enrich, Consume and Deliver. OSDU™ Data Platform code is designed to represent the individual business logic required for different microservices, which are custom-tailored for the users needs. (Epam a)

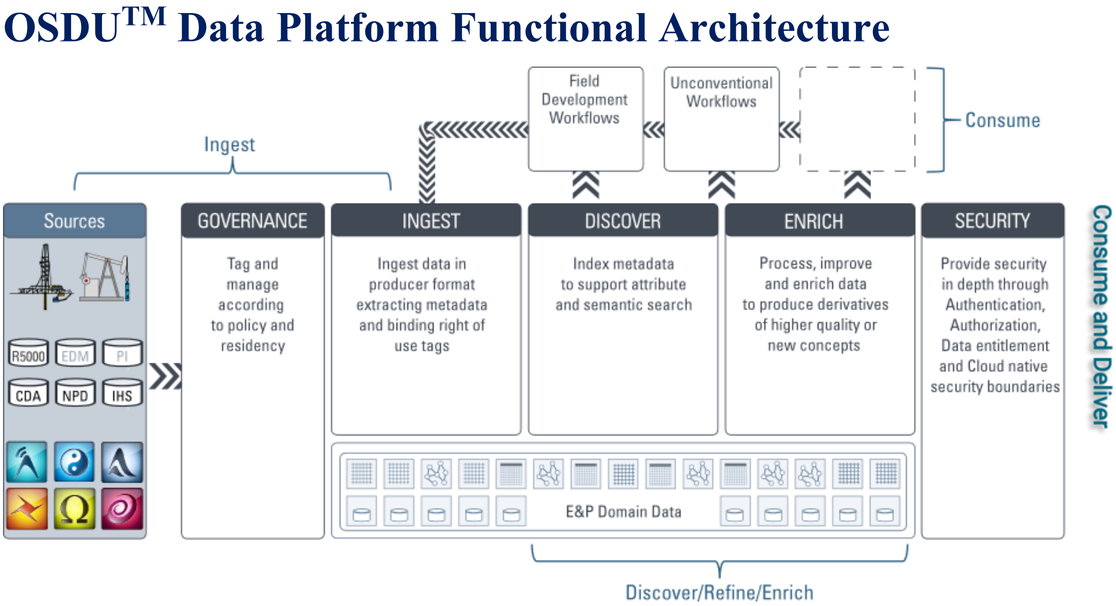


Figure 12: *Functional Architecture of OSDU ™ Data Platform with the three operating overlapping zones of: Load and Ingest, Process and Enrich, Consume and Deliver. On the left-hand side in green are the data sources, such as Techlog Studio or Petrel Studio, but also NPD (Norwegian Petroleum Directorate). Green arrows denote the cyclic processes within Process and Enrich, meaning that there is a data flow of new data generated by applications in enriched form, being fed back into the platforms' ingestion section for discovery by other users. (OSDU 2023a)*

Consume and deliver revolves around access to data and exploration of the data.

Ingest: External data can be loaded into the data platform enabled by load and ingest. For data ingestion, two options exist, either by:

a) Ingest by reference: creating a reference to the source data, whereby the source data is located outside OSDU™ Data Platform, or:
b) Ingest by copy: creating a copy of the source data and ingesting the copied data into OSDU™ Data Platform.

While both are valid options, if possible, option b) is preferred because the data can be directly managed in the platform. Process and enrich allows for discovery, enrichment and delivery of data already existing in the platform. A more extensive explanation on ingestion into OSDU™ Data Platform is given in Chapter 3.1 – New and Backlog Data. There, the concept of schemas (WKS), entities (WKE) and DDMS (Data Domain Services) is introduced and emphasized on in further detail.

Discover: Discovery revolves around searching the metadata catalogue. A precondition for making data searchable is to index documents, which can be either in structured or unstructured form. A separate persistent store exists, into which documents and indices are saved. The reason for the existence of a separate store is due to it being optimized for queries and delivering fast search results.

Enrich: Enrichment takes original data and combines it with other data to produce a more valuable result, in the sense that it can make work easier for users or that it derives new insights which were unavailable originally with data being separated in silos.

Deliver: The single reason for OSDU™ Data Platform to exist is to make data available and usable at the stage of delivery (consumption by the user). Three ways of consuming data in OSDU™ Data Platform are: Content delivery from files, and domain data management service (DDMS). (Epam a), (Epam b) Each come with advantages and limitations, which are listed in Table 5 below:

Data consumption	Advantages	Limitations
Content delivery from files	+ Flexibility toward new data types + Compatibility with applications outside OSDU™ Data Platform	- Applications will not be enriched by OSDU™ Data Platform and vice versa
Consumption zones	+ Information is kept up to date	- Duplicated data - Only read capability
DDMS	+ Read and write possible + Modular, domain driven API, based on microservices	- Complicated to develop

Table 5: *Comparison of advantages and limitations for each data consumption method in OSDU™ Data Platform. (Epam a)*

Data Flow Patterns: Extract – Transform – Load (ETL) vs Extract – Load – Transform (ELT)

When it comes to loading and preparing the data, OSDU™ Data Platform takes a different approach compared to traditional corporate repositories. Figure 13 and Table 6 illustrate these differences both graphically and as a comparison between advantages and limitations of each approach. Two aspects are characteristic for OSDU™ Data Platform:

a) Instead of first preparing the data and then loading it (ETL), OSDU™ Data Platform first loads the data (ELT), then data is prepared for the data store. Data discovery is possible with SQL.

b) Instead of consuming data and then enriching it locally, OSDU™ Data Platform takes the approach of enriching data after discovery and then loading the enriched data back into the data store for further discovery by other users.

ETL: Extract Transform Load (Corporate Repository Approach)

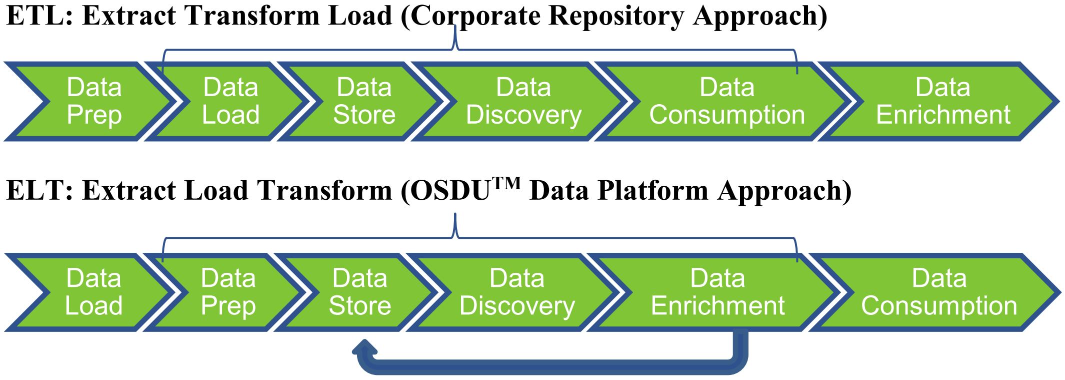


Figure 13: *A comparison between ETL and ELT approach. (Epam a)*

Approach	**Advantages**	**Limitations**	**Example**	**Data discovery**
ETL	System is well embedded in OMV processes and IT landscape	Potential for data enrichment not fully used	Seabed database	Enabled with SQL
ELT	Information is preserved Improvements are shared	Implementation takes time and resources	OSDU™ Data Platform	Enabled with Python

Table 6: *Comparison between ETL and ELT approach with individual advantages and limitations of each approach including an example. (Epam a)*

ELT approach is closely linked to the use of data lakes, in which all data is first loaded, indiscriminant of file format or whether it is structured or unstructured data. An important note is that both ELT as well as ETL work with OSDU™ Data Platform, but the main intended use of OSDU™ Data Platform revolves around ELT. It must be investigated inn more detail whether OMV has the resources to establish an ETL workflow opposed to the ELT approach which is typically used in conjunction with OSDU™ Data Platform.

OSDU™ Data Flow Services

When data is brought to OSDU™ Data Platform, it undergoes a series of stages to make it fit for the purpose of data ingestion, e.g., indexing the data to make it searchable. (Figure 14).

As data flows through OSDU™ Data Platform, it undergoes the following stages:

1) Data is brought into the system from various data sources such as raw files and a schema is created for referencing the data
2) Ingestion framework is initialized, metadata is saved
3) Access entitlements are defined
4) Storage of content data in DDMS
5) As data is stored, an event is triggered leading to a notification
6) Indexer service is called
7) Normalization of reference frame via the unit and Coordinate Reference System (CRS) core services. This step makes data searchable
8) A user searches for specific data
9) Client applications (SaaS or PaaS) allow for data consumption (Epam a), (Epam b)

Figure 14 shows all the steps that data undergoes from source to consumption, e.g., indexing and storage.

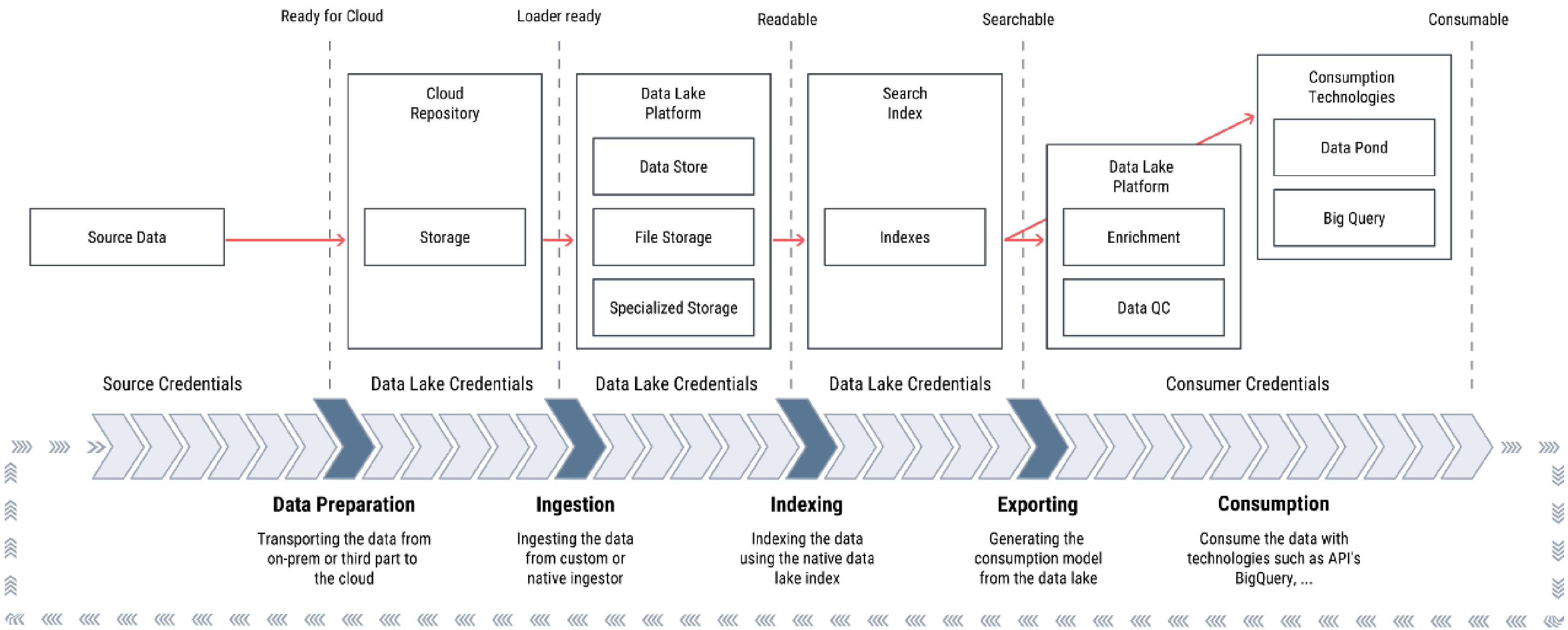
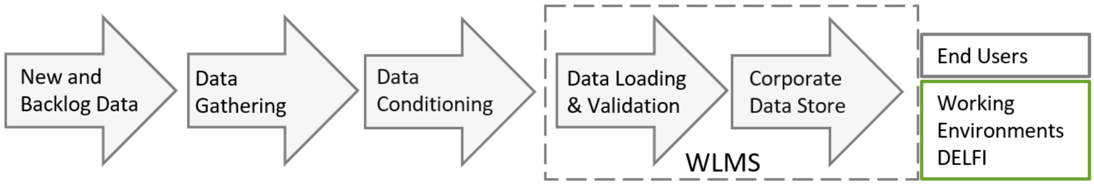


Figure 14: *An overview of the data flow stages from ingestion to consumption by users. Note how each step serves the purpose of delivering enriched, searchable, and normalized data to users with well-defined access rights. (OSDU 2023)*

2.5 DELFI Fundamentals



*The DELFI **cognitive** E&P environment is a **secure, cloud-based** space built on key premises […]. It harnesses data, scientific knowledge, and domain expertise to fundamentally change the way of working in every part of the E&P value chain. The DELFI environment makes applications and workflows **accessible** to all users and enables team members to build **common workspaces** for data, models, and interpretations while respecting proprietary information boundaries. (SLB a)*

DELFI is essentially defined by 4 cornerstones, which are:

I) **Data Security:** As DELFI is running on a cloud environment, its' security is inherently higher than conventional applications' data security. DELFI has obtained the certificate of Service Organization Controls 2 type 2 (SOC2) accreditation, acting as proof of cloud providers' security measures meeting a certain standard, relevant for the customer.

II) **Openness and collaboration:** Like OSDU^{TM} Data Platform, DELFI uses open-source code. APIs can be used to access the solutions of DELFI environment. Various data sources can be integrated and shared. It enables cross-disciplinary collaboration on a shared data platform.

III) **Cognitive:** Supported by artificial intelligence and analytics, users can make better decisions faster, with tools such as automated tasks or intelligent search engines. Supported by DELFI AI, seismic processing, well planning, fault interpretation and field planning can be accelerated extensively compared to non-AI supported work.

IV) **Live:** Users can rely on DELFI providing them with the latest, up-to date version of data in their live-projects. This boosts the users' trust in data by ensuring business continuity and reduction of duplicates. Users can work on the same project at the same time, therefore strengthened collaboration is possible. (SLB a)

Other important aspects of DELFI are that the environment is scalable and compatible with relevant software used in the industry such as Petrel and Techlog. The subscription model is SaaS based. DELFI offers cloud-native applications for each stage in field development, from planning to operation (e.g.: ExplorePlan, ProdOps, …), which are visualized in Figure 15. (SLB b)

DELFI Platform Architecture

OSDU™ Data Platform is an integral part of DELFI as can be seen below in Figure 15. The foundation of this combination lies within the OSDU™ core services which were explained in section 2.4, Figure 14.

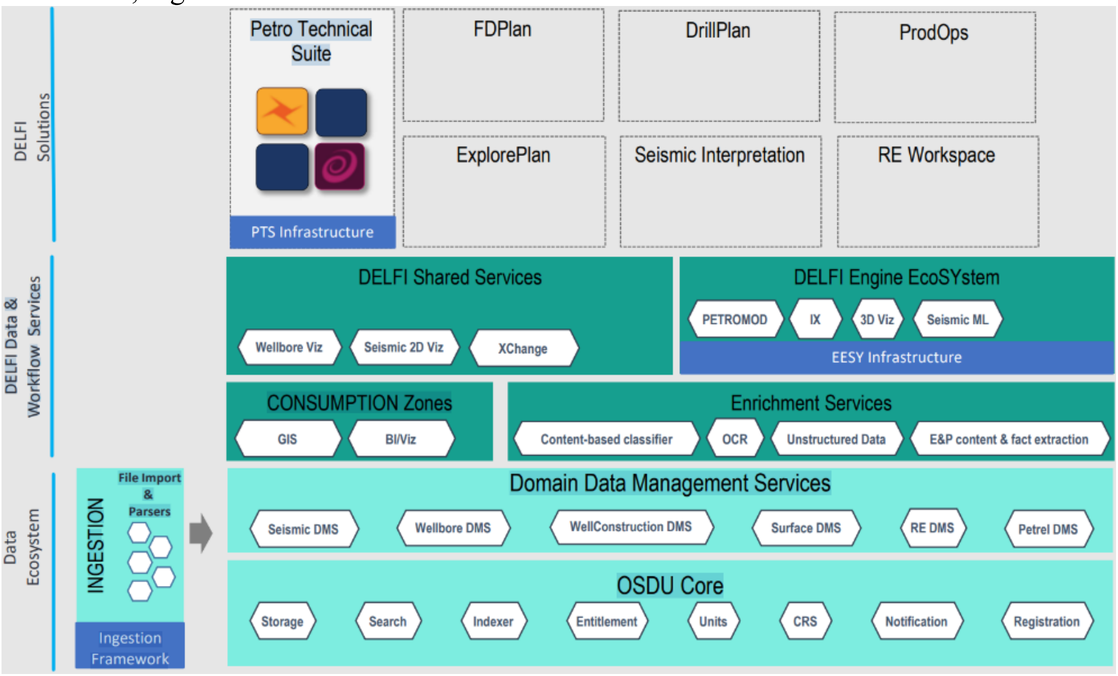


Figure 15: *An overview of DELFI platforms' architecture. The three segments of DELFI are presented from top to bottom: DELFI Solutions: Petro Technical Suite (PTS), cloud native solutions (DrillOps, ...), DELFI Data & Workflow Services: DELFI Shared Services, Consumption Zones, ..., as well as the Data Ecosystem; Containing OSDU ™ Core Services, Ingestion layer and DDMS. (OMV Energy)*

DELFI platform architecture can be split into 3 segments:

- **DELFI Solutions:** SLB software solutions contained in the Petro Technical Suite (PTS), such as Techlog and Petrel, are embedded within the DELFI environment together with cloud-native solutions such as DrillOps.
- **DELFI Data & Workflow Services:** Composed of four sections: DELFI Shared Services, DELFI Engine EcoSystem, Consumption Zones, and Enrichment Services.
- **Data Ecosystem:** DELFIs' data ecosystem is OSDU™ Data Platform, which is made up of microservices, relying on 3 integral parts: an Ingestion Framework with file import and parsers, the Domain Data Management Services (DDMS) as well as OSDU™ Core Services, such as storage, search, and normalization services. (OMV Energy), (SLB b)

Domain Data Management Services:

In the DELFI Data Ecosystem, the Domain Data Management Services are grouped as follows: Seismic Data Management Service (DMS), Wellbore DMS, Petrel DMS, Well Construction DMS, Reservoir Engineering DMS and Production DMS.

As the scope of this thesis is focused on wellbore centric data, more specifically well logs, mudlog and core data, this section on DELFI fundamentals will magnify mainly on the Wellbore Data Management Service section.

Wellbore DMS is a cloud-native solution powered by Google Datastore and Google Cloud Storage Service for data storage and transition to Azure cloud is in progress. (OMV Energy)

DELFI Shared Services:

An important example of DELFI Shared Services is 3D Visualization "3DViz":

As different data types are obtained from OSDU^{TM} Data Platform, such as well logs, 3DViz will find an appropriate application for visualizing the data based on ingested data type. (OMV Energy)

Techlog Data Liberation:

A central feature of DELFI environment is that data stored in Techlog can be liberated into the Data Ecosystem. Figure 16 gives a conceptual overview of Techlog data liberation:

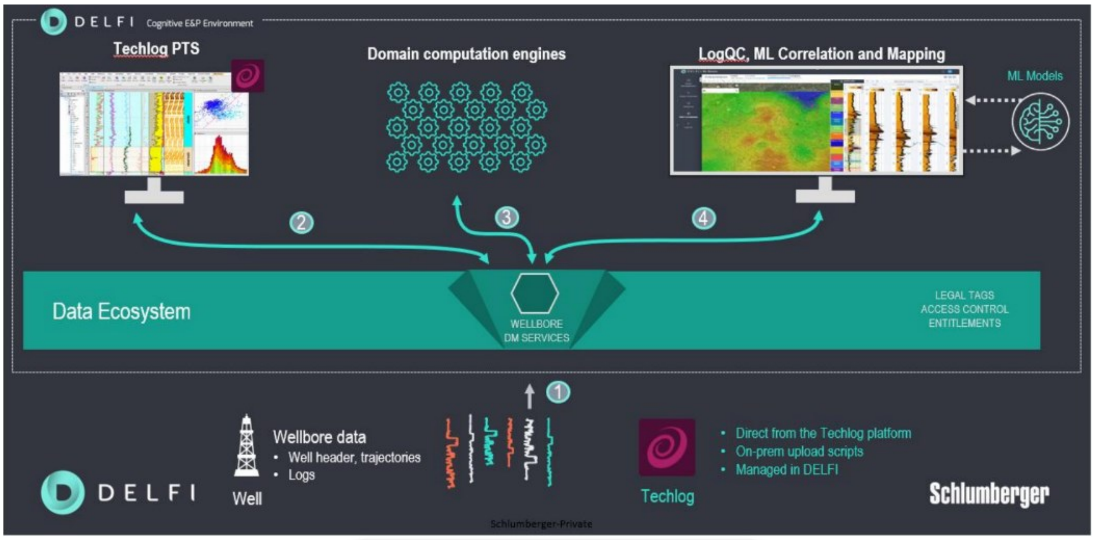


Figure 16: *Techlog data liberation: Wellbore data (header, trajectories, logs) is liberated from Techlog to the Data Ecosystem. The Data Ecosystem is interconnected with Techlog Petro Technical Suite, domain computation engines and DELFI native apps such as log quality control (LogQC) as well as machine learning (ML) correlation and mapping, all of which can access wellbore data from the Data Ecosystem. (OMV Energy)*

Techlog data liberation can be performed in three steps:

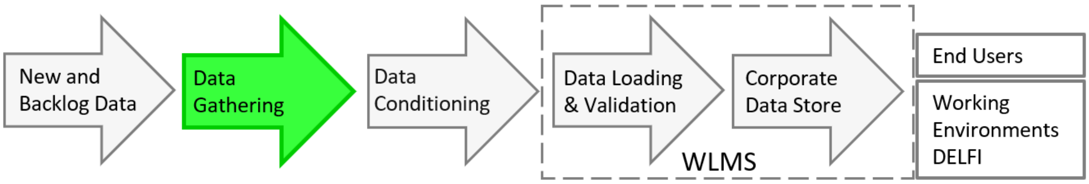
I) Storage in Wellbore DMS can be done either by:
   - Directly saving Techlog project in Wellbore DMS
   - Data Ecosystem ingestion pipeline storage, such as saving LAS or DLIS files in Wellbore DMS

II) Retrieval and saving of wellbore data directly from Wellbore DMS is possible for Techlog, which his embedded in Petro Technical Suite

III) Data that has been liberated to Wellbore DMS is consumable by:
   - DELFI computational engine
   - DELFI native apps such as LogQC (OMV Energy)

DELFI Wellbore DDMS API in the DELFI developer portal allows for storage and manipulation of wellbore log data and related objects, namely: Logs, LogSets and Wellbores. It is possible to create Wellbores, LogSets and Logs based on JSON objects, to get a Wellbore / LogSet / Log by ID, to delete a Wellbore / LogSet / Log by ID and to update a Wellbore / LogSet / Log from its ID with a new JSON object. (OMV Energy)

Chapter 3 Data Gathering and Audit



WLMS as a Data Foundation has been presented in Chapter 2 extensively and an overview was given in Figure 4. The data lifecycle in WLMS consists of different stages, from data acquisition stage to data consumption stage. Data consumption takes place not only in the working environments (Petrel, Techlog Studio), but also during field development steps such as well planning, operations, plugging and abandonment or modelling. The following steps are involved in launching a successful data gathering and audit campaign:

1. *"Aggregate data from disparate sources into flows of high-quality information that can be relied upon for machine-based analysis and decision-making."*
2. *"Manage these automated analyses of data and the classification and ranking of results in accordance with pre-defined methods and rules."*
3. *"Marshal the flow of resulting decision-support information and actions amongst pre-targeted experts."*
4. *"Support activity monitoring for audit and performance management purposes."* (Daum et al. 2010)

Two major challenges arise in data gathering and audit:

1) *"Experience has shown that although basic […] data processing and analysis tasks are reasonably straightforward to automate with resulting efficiency gains and benefits for quality assurance, the real challenge for workflow automation is to support tasks that require collaborative expert intervention in workflow patterns that can vary on a case-by-case basis and change unpredictably from task to task.*" (Daum et al. 2010)
2) Ensure that the workflows defined are followed by users via an audit system that functions based on six sigma principle.

The success of data gathering and - entry as a step within the data lifecycle relies heavily on the fulfillment of following requirements:

- Correct Data shall enter the system.
- Data shall enter the system through a single point of entry.
- Data shall enter the system as a complete, fit for purpose, Final Delivery Package, containing metadata describing the data (e.g., what, why and how it was generated).
- Data shall enter the system with maximum: automatization, standardization, and quality control to enable maximum trust in data.
- Data shall enter the system with minimum: manual effort, inconsistencies, duplications and ambiguity about the origin and context of the data.

Mastering the process step of data gathering and data entry holds such a high significance, because it subsequently affects all the users who consume the data at a later stage. Figure 17 gives an idea of how data is transformed through the data lifecycle, starting in a rather unstructured form, ending up as structured high company value data in the working environments (subset of all data available!).

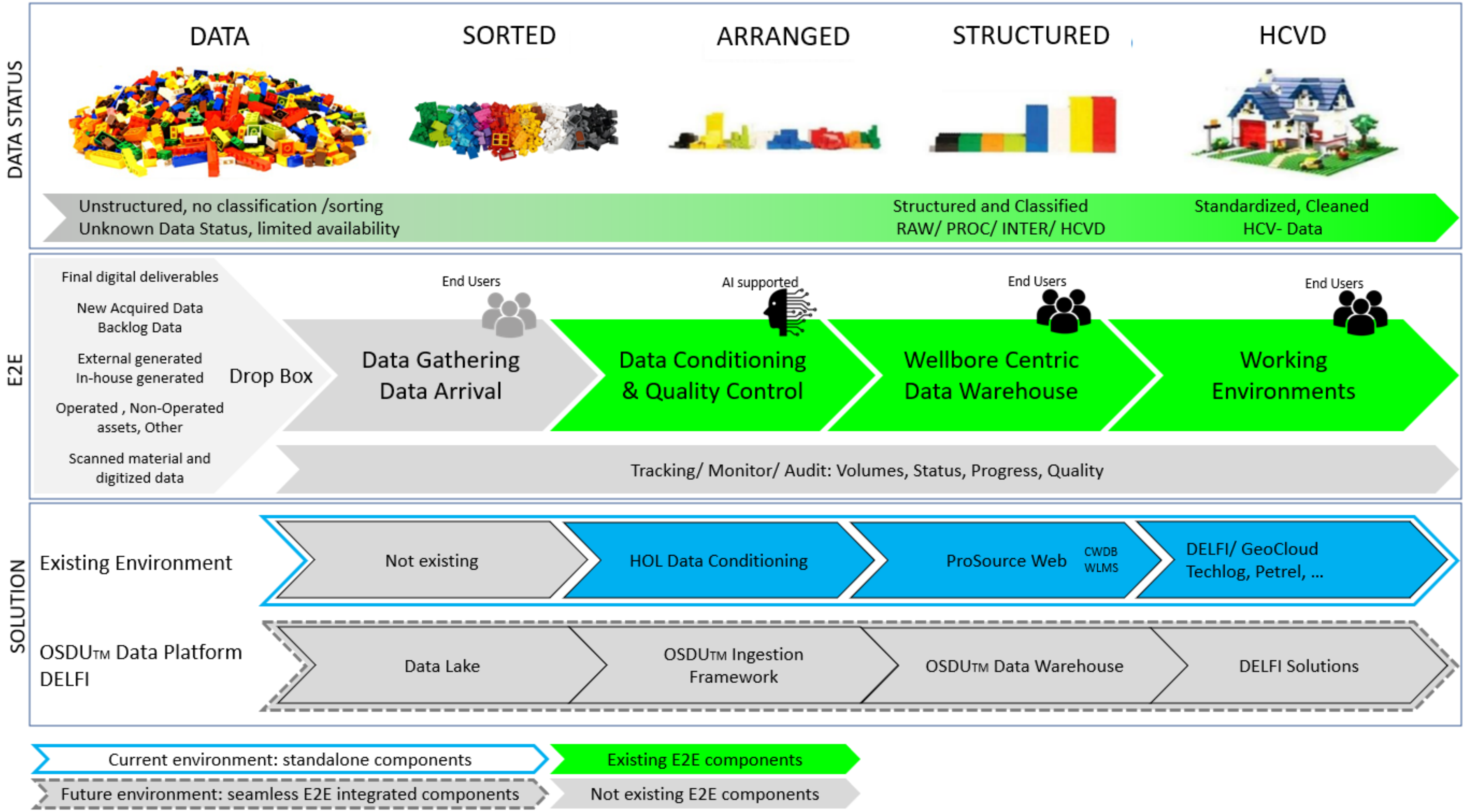
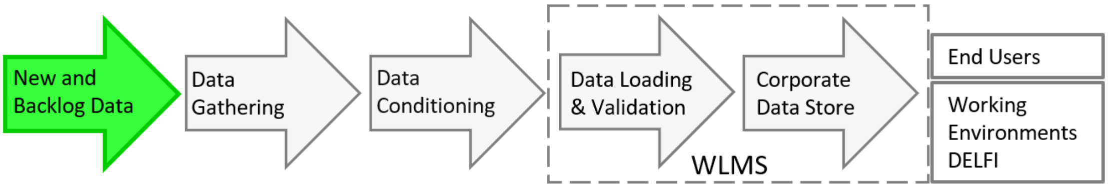


Figure 17: *From unstructured and unclassified data of unknown source (top left) to standardized, structured, and cleaned high company value data (HCVD, top right). A potential element in Data Gathering and Data Arrival stage is implementation of a Data Lake to host unstructured data. (Stanford)*

3.1 New and Backlog Data: Single Point of Entry



Some pain points on new and backlog data are listed below:

- Strenuous effort for delivery: data does not always arrive in WLMS and there is a high effort behind making sure it does arrive finally. This applies not only for non-operated boreholes but also for OMV and OMV Petrom operated boreholes.
- Wrong order: Data is sometimes loaded first into the working environments causing that relevant data are never ending up in the corporate leading system. However, it would be correct to (1) load the data into the relevant leading system (e.g., WLMS) and distributed (2) into the working environments in a controlled and transparent manner.
- Incomplete data: Delivered data from service contractors are sometimes incomplete due to a lack of or no internal QC, inconsistent QC, or lack of expertise for proper QC. The Final Delivery Package is often only partly or not at all submitted.
- Little compliance: BOs fail to see the immediate benefit of delivering data in a way that is fit for purpose and on time. One reason is the lack of personnel in BOs with relevant domain knowledge who are able to ensure proper complete service contractor deliverables and can condition/ prepare the data for loading to WLMS.

For OMV, there are three roles of data submitters who deliver log, core or mudlog data:

- Head Office (HO) Data Submitter (supporting BOs on request)
- Branch Office (BO) Data Submitter
- Service Provider (Contractor) Data Submitter

Ideally, new – or backlog data (Log -, Core - or Mudlog data) shall enter the system through a Single Point of Data Entry. Within this chapter, two forms for a Single Point of Data Entry will be presented, evaluated, and discussed: First generically as Hybrid Data Lake, then as a use case with OSDU™ Data Platform as the Hybrid Data Lake (i.e., Data Platform, - used interchangeably).

While there are already existing regulations on how new data must enter the system, probably one of the most challenging steps in the data lifecycle is backlog data gathering. It is estimated that about 800 000 LAS, LIS and DLIS files are currently existing as backlog data, excluding support documents like log plots (e.g., scanned logs), hardcopy logs and other support documents (e.g., reports) in TIF, PDF format or other formats. Figure 17 illustrates how backlog data loading is currently done.

Data Lake vs Data Warehouse vs Data Platform (i.e., Hybrid Data Lake: Data Lake + Data Warehouse, e.g.: OSDU™ Data Platform)

A data lake is defined by Amazon AWS as: *"[…] a centralized repository that allows you to store all your structured and unstructured data at any scale. You can store your data as-is, without having to first structure the data, and run different types of analytics—from dashboards and visualizations to big data processing, real-time analytics, and machine learning to guide better decisions."* (Amazon)[16]

This definition however gives the impression that information can be simply "**dumped**" into the data lake. It is not advised to do so! It is often the case that data lakes end up as data swamps.

A data warehouse on the other hand is solely made for storing structured data and comes with its' advantages but also limitations and shortcomings. Three inherent limitations of a data warehouse as a standalone repository are:

- Semi structured and unstructured data can only be stored with limitations in the data warehouse.
- Application architectures are nowadays more microservice oriented, which makes them rather incompatible with traditional data warehouses.
- More and more users require direct access to the raw data through analytic tools. A data warehouse can only give that kind of access to users in a complicated way and with limitations. (Zburivsky und Partner 2021)

A **combination between a data lake and a data warehouse in series**, acting as a **Data Platform** or also known as **Hybrid Data Lake**, might be the solution to the challenges connected with data variety and volume that come in the form of well logging, mudlog and core data through a single point of data entry.

"[…] a data lake is almost always coupled with a data warehouse. The data warehouse serves as the primary governed data consumption point for business users, while direct user access to the largely ungoverned data in a data lake is typically reserved for data exploration either by advanced users, such as data scientists, or other systems." (Zburivsky und Partner 2021)

Figure 18 shows the structure of a data platform, with an ingestion, storage, processing and serving layer. Also, two modes for processing are presented.

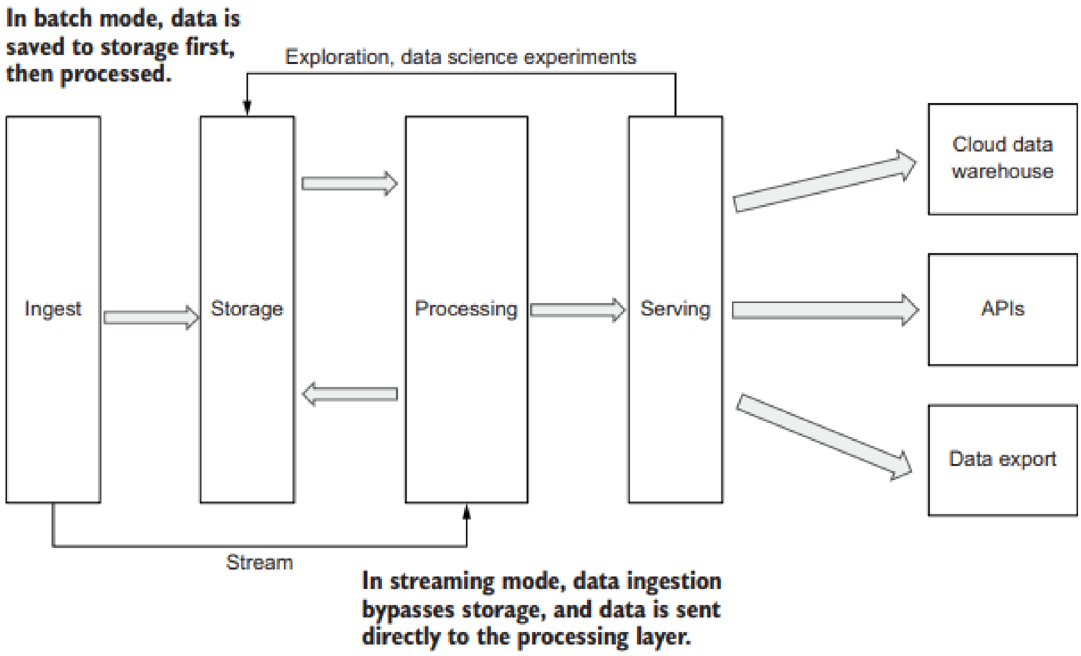


Figure 18: *The four foundational building blocks involved in a data platform are: ingest, - storage -, processing - and serving layers. Two modes for processing exist: batch and stream. (Zburivsky und Partner 2021)*

Ingestion layer: Data is brought into the data platform. A variety of data sources is connected to the ingestion layer, be it relational or NoSQL databases, data storage, APIs etc. … The ingestion layer must provide High flexibility for ingestion of different data types and formats. (Zburivsky und Partner 2021)

Storage layer: Data Lake storage of the raw data. (Zburivsky und Partner 2021)

Processing layer: Data is processed via software such as Apache Spark and then brought back to storage in batch mode (Figure 18). (Zburivsky und Partner 2021)

Serving layer: Data is prepared for consumption. (Zburivsky und Partner 2021)

Ingestion layer is explained further in this section, as it is the main building block for a single point of data entry:

Keeping data "as is", making no modifications to the incoming raw data and therefore enabling data lineage. Data ingestion is compared for two cases in Table 7: 1) Data ingestion into Azure Synapse (data warehouse) and 2) Data ingestion into Azure data platform : (Zburivsky und Partner 2021)

Data architecture	**Azure synapse (Data Warehouse only)**	**Hybrid Data Lake (Data Lake + Data Warehouse)**
Upfront output schema specification required?	Yes	No
Resilience to source (input) schema changes?	No	Yes
Destination (linked service)	Azure SQL data warehouse	Azure blob storage
Data source	Relational database	Multiple data sources with structured & unstructured data
Applicable if level of control is	Control over source data and a process exists to manage changes in schema	No full control over various data sources

Table 7: *A comparison between Azure synapse and Azure data platform in terms of output schemata, resilience against change and linked service.* *(Zburivsky und Partner 2021)*

It is important to note the practicability of not having to define output schemata specifications upfront. This has two reasons:

- It is very difficult to change output schemata,
- Lots of manual effort for defining output schemata.

Figure 19 gives a conceptual proposal of how a data platform (Hybrid Data Lake) can look like, - from ingestion to the Data Warehouse.

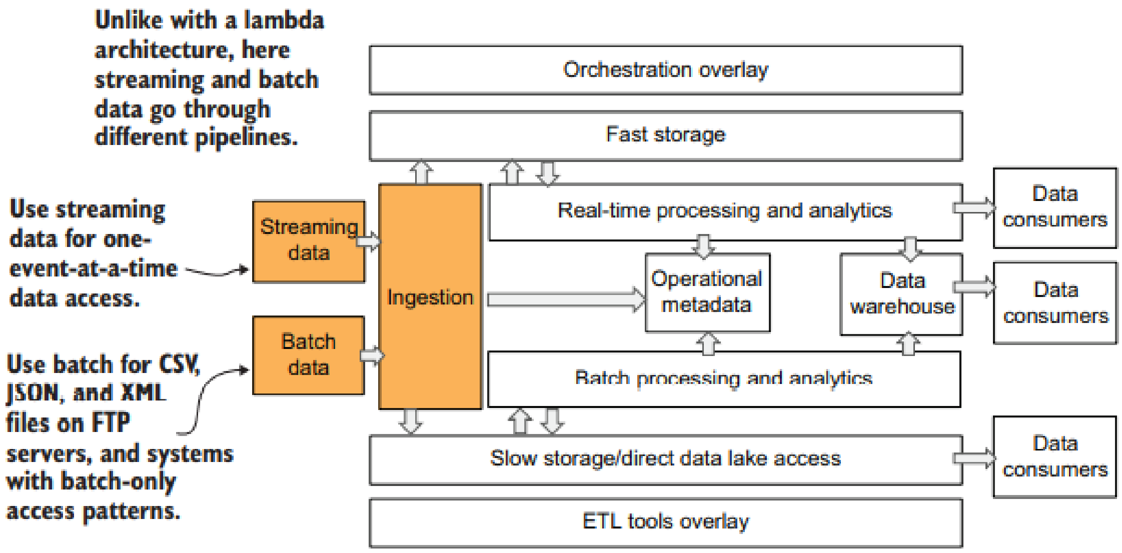


Figure 19: *Data ingestion comparison for streaming data and using batch mode. (Zburivsky und Partner 2021)*

When looking at the oil industry's data, a general distinction can be made according to data type and processing:

- **Data at rest:** processed in batch. E.g.: Historic data
- **Data in motion:** processed live (streaming). E.g.: Drilling data

A Data Lake is structured in a way that allows for processing both data at rest as well as data in motion. The individual purposes of a Data Lake and a Data Warehouse forming a Hybrid Data Lake can be characterized as follows:

- **Data Lake:** Temporary, fast storage. Objective: Making data available for end users
- **Data Warehouse:** Long term, slow storage. Objective: Over time, data is deleted from the short-term Data Lake storage and moved to long term storage.

For example: A well is drilled, and decisions are made live with geo-steering based on LWD technology, with data being streamed directly to the operator. The data is streamed into the Data Lake, where it can be immediately used by decision makers or analysts. Streamed data is not stored, but directly processed, whereas batch data enters the storage of the Data Lake before being processed. Once true vertical depth has been reached, the data is no longer in motion (static data, data at rest) and therefore the streaming and batch data is moved to long term storage of a Data Warehouse. It no longer makes sense to store the existing batch data duplicates in the Data Lake, making it obsolete in the Data Lake, where it is subsequently deleted, and space is made for future incoming batch data.

Data Lake vs Data swamp

The almost unlimited storage capacity of a data lake exposes it to the risk of quickly becoming a data swamp. The following governance rules should be in place to avoid disorganization and thus transformation to a data swamp:

- Consistent file naming conventions.
- Consistent folder hierarchies. (Antonio et al. 2019)
- Exact definition of which attributes must be present for each type of deliverable that the service contractor is uploading to the Data Lake: e.g., FMI curve needs specific curves to be loadable into Techlog.
- Exact definition of which attributes must be entered by the contractor.
- Metadata enrichment: Inclusion of metadata about the files, thus enabling a contextualization and search component. Metadata shall include 5 groups of information: (Camden 2021, June 22)

  General (Logging contractor name, …)

  Asset (Country, field name, well name)

  Context (Discipline, …)

  Control (Data owner, creation time, last update time…)

  Workflow (Approval, Initiation, …)

A kind of system to keep track of inventory inside the data lake is a data catalogue. It should contain information about the file storage location for each information object and a file path as well as the metadata that was defined during data submission at the receiving dock in the metadata entry form. A receiving dock should be considered when designing a data lake, because if it does not exist, there is no buffer zone between data sources and the data lake. Without that receiving dock acting as a buffer, users would be able to upload data directly into the data lake without prior contextualization with metadata, which would pose a risk for the data lake to become a data swamp. A receiving dock will increase the trust in the data. (Antonio et al. 2019)

The dangers/ implications with data swamps are that slowly old and irrelevant data accumulates and at some points users will be unsure of which data is correct and up to date. Also, a data swamp implies the risk of having many duplicates and several slightly different versions of the same data (final versus filed versus working copies). (The Dangers of Data Swamps and how to avoid them)

Generic Hybrid Data Lake as a Single Point of Data Entry

Different sources of information must be centralized into one single repository, and this must be done in a standardized way, otherwise there is a risk of creating a data swamp, as discussed in the previous section. The goal of having a single point of data entry is to: *"aggregate data from disparate sources into flows of high-quality information that can be relied upon for machine-based analysis and decision-making."* (Daum et al. 2010) From data generation to delivery of data to WLMS, in a Hybrid Data Lake data passes the following stages: (Gruijters, S., De Boever, E., & Visser, W 2021, June 22)

- **Acquisition:** variety of data (well logs, mud logs or cores) incoming from different sources, either **operated** (New incoming data from either service contractor or the branch office, backlog data from the branch office) or **non-operated** boreholes (Operator: new incoming data or branch office: backlog data), as well as **other data sources:** Data room visits, purchased data, external databases such as NPD or digitalization and scanning projects.
- **Dropbox:** Data is being brought into the data lake. Metadata enrichment (metadata entry form) takes place with attributes such as country, field, license of the wellbore that has been logged and in which format. Access rights are defined, so that only entitled users may have access to the data.
- **Buffer zone:** To avoid having a data swamp, there should be different stages between upload via Dropbox and storage, such as: Raw / Clean / Ready
- **Storage:** Data Lake storage in containers, folders, and filesystems.
- **Quality Control:** Quality control by domain specialists (Petrophysics): Completeness and correctness must be checked (e.g., new incoming data), followed by data managers quality control.
- **Conditioning: (in OMV HO and OMV Petrom currently done with support of** HOL): According to and honoring the tool/ curve dictionaries and data classification, data must be prepared and conditioned along the data processing chain.

An increasingly popular model takes a traditional Data Lake and combines it with a Data Warehouse: this model is known as a Hybrid Data Lake or other synonyms include: Data Fabric, Lake House.

Figure 20 shows the following architectural elements, which must all function together:

- **Dropbox:** Datasets can be uploaded either via:
  - web service,
  - batch transfer (FTP) from the client to the server,
  - file streaming. (Ocampo et al. 2019)

However, file streaming is nearly impossible with third party data sources and therefore, only the first two options web service and batch transfer are relevant. Different base cases must be considered, where data is coming from different locations and different service providers, branch offices, purchased, digitalized data, etc. … By asking the service contractor, branch office or others to submit the dataset as well as metadata related to the dataset through a single point of entry:

+ A structured, standardized manner of data submission is promoted.
+ Identification of missing data objects is easier. The operator can set/ enforce strict data delivery rules to improve data quality
+ Clear data responsibility is given to the data submitter uploading files to the Data Lake. This data ownership is active until all elements have been delivered to satisfy creation of a Final Delivery Package that can be used by data consumers.
+ Metadata enrichment will upon upload will later help to query the data and support the overall data classification in a data warehouse. Consistent meta data population is the key for future automatization of data flows and interpretation workflows.

- **Cloud synchronization service:** Including audit fields with information on date created and date last updated, in order to only pull the data objects that have been added newly or to pull data objects that have been changed in the data source and must be updated accordingly. CDC (change data capture) processes make sense when there is file streaming in place and frequency of data changes is high. However, when the system is not subjected to highly frequent changes in a web service or batch transfer setup, CDC does not make as much sense to implement and a full refresh daily or several times a day is more practical and economical. A CDC pipeline takes some time and experience to implement. (Microsoft)

- **Quality Control:** If for example the header information in a log is missing, the intake service shall give an error and reject data submission, sending the data back to the service contractor and ask for re-upload with the specific prompt detailing which element or line was missing.

- **Data conditioning:** Chosen service contractors for data conditioning would directly receive the data via the system and data does not have to be further uploaded/ downloaded by OMV personal from/ to an external server any longer. An automated logging summary report (LSR) generation should be implemented in this pipeline.

- **Data Warehouse:** Transfer of the Final Delivery Package to the Data Warehouse for consumption by data users.

- **Final Delivery Package transfer to Data Warehouse:** Seamless transfer from data conditioning stage (HOL) to Data Warehouse (WLMS system) with the focus of getting data first into WLMS and then secondly into the working environments.

- **Data ownership transfer:** Once data is in WLMS system, there is a data ownership transfer from the service contractor to the head office (HO). Optimally, an underlying SLA (surface level agreement) should exist, in which is clearly defined that only when the data has been published in WLMS, the contractor shall receive payment. Therefore, cooperation with legal and payroll is necessary for this system. Please refer to Chapter 3.3 Data Audit Tool on how such an SLA workflow could look like.

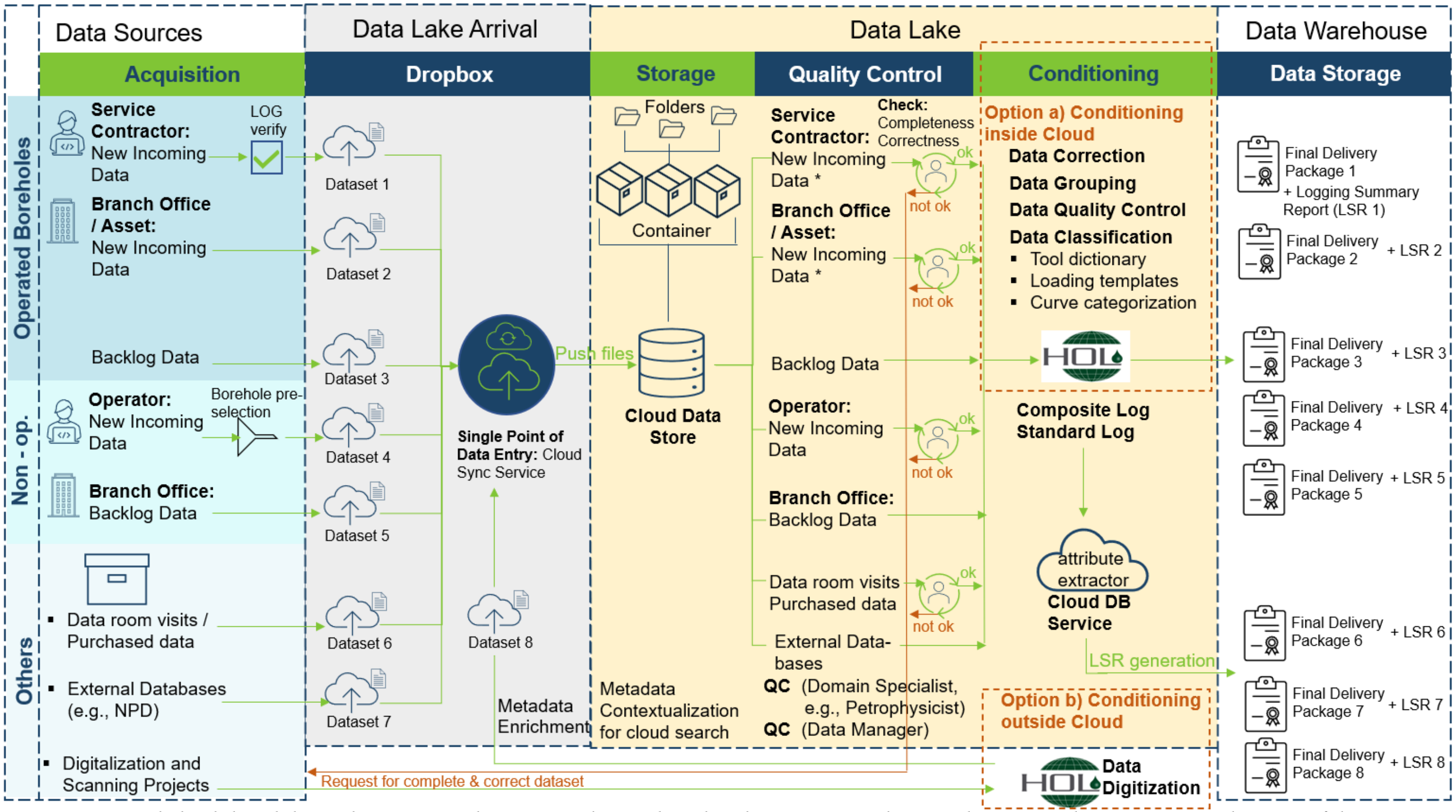


Figure 20: *A hybrid data lake architecture, with a Data Lake combined with a Data Warehouse. The Data Lake acts as a single point of data entry for datasets coming from different data sources, entering the Data Lake through a Dropbox. Left side: Note how log-, core - and mudlog data from service contractors is first verified to be conform and that operator's new incoming data from non-operated boreholes is pre-selected. Right side: Note how data ingested into the Data Lake would undergo another quality control step and if the quality criteria (completeness, correctness) are not fulfilled, datasets are sent back to the respective data source where it came from.*

Use Case: OSDU™ Data Platform as a Single Point of Data Entry

To enable a single point of data entry in OSDU™ Data Platform, different kinds of elements must be in place compared to a Hybrid Data Lake. OSDU™ Data Platform is acting as a reference that is highly customizable. Only OSDU™ core services need to be used every time for ingestion. Other than that, OSDU™ Data Platform is customizable. Two examples are given below to illustrate this customizability of OSDU™ Data Platform:

- OSDU™ Data Platform acts as reference if a highly agreed upon format such as LAS is ingested and stored
- Usage of DDMS (Domain Data Management Service) is not absolutely necessary, OMV can custom-tailor OSDU™ Data Platform and decide on which elements it wants to use.

With Wellbore DDMS, users can ingest and discover (find) data. Wellbore DDMS coordinates the a) consumption and b) ingestion of files in OSDU™ Data Platform. As data goes through different stages (loaded, approved, …), certain attributes are added to data and notifications are generated. For ingestion, there is a set of rules on how data shall be ingested into OSDU™ Data Platform. With the variety of data in E&P, it is impossible to have one single schema for ingesting all the different data formats. This is why OSDU™ Forum has introduced schemas, so called Well-Known Schemas (WKS).

Adoption of OSDU™ Data Platform at OMV: Potential problems and solutions

A number of problems have been addressed by the Open Group as well as by the operators working on OSDU™ Data Platform. Also, some problems that could be expected by OMV when moving toward OSDU™ Data Platform have been summarized below – six problems and six solutions for these problems. The composition of problems presented below have been identified during discussions with the Senior Petrophysicist and Advisor Data Integration of OMV (problem identified during OMV discussion). Some problems were brought up because they would pose generally a problem for Energy companies adopting OSDU™ Data Platform, for example because the specific feature is still under development by OSDU™ Forum (problem identified as general issue). Some solutions were found via internal OMV discussions with the IT Architect ("solution identified during OMV discussion") or via the online OSDU™ Forum documentation / London OSDU™ Forum 2023 documentation ("solution identified via OSDU™ Forum documentation"). The list starts with generic problems and problem 6 refers to a specific issue regarding core/cutting data ingestion.

Regarding the adoption of OSDU™ Data Platform in OMV:

- **Potential generic problems and solutions.** (Problem 1 – Problem 5)
- **Potential problem and solution with ingestion of Core/ Cutting Data into OSDU™ Data Platform.** (Problem 6)

Adoption of OSDU™ Data Platform at OMV: Potential generic problems and solutions

- **Problem 1: Ambiguity of what a dataset is based on OSDU™ Forum definition, and how it relates to WLMS dataset concept, as well as the definition of a work product component. (Problem identified during OMV discussion)**
    - **Solution for problem 1: Dataset in OSDU™ Data Platform contains metadata of digital files, such as file size. OSDU™ dataset does not relate to dataset concept of WLMS. (Solution identified via OSDU™ Forum documentation).**

Work Product Component (WPC): *"This is the smallest independently usable unit of business data content transferred into the data platform. Each Work Product Component points* *to one or more data containers known as Files. For example, SeismicTraceData WPC points to segy file, which contains metadata about the trace data and digital information about seismic traces."* (Amazon) A work-product-component is a type of file in OSDU™ Data Platform. Examples of work-product-components include: "Document", "Data Quality", "WellLog", "WellboreTrajectory", "SeismicTraceData"… For example, the WPC SeismicTraceData is defined as *"a single logical dataset containing seismic samples."* (Work Product Component Definition) The WPC "WellLog" of OSDU™ Data Platform is comparable to the dataset concept in WLMS.

OSDU™ Dataset: *"This provides metadata about digital files and datasets. It doesn't describe business content such as tracedata, log data, etc. that are found within the digital file and dataset, rather it stores information such as file size and checksum. Data containers referred to as Files contains digital business data. Datasets can be defined by a specific file format, such as seismic (segy), drilling (witsml), well log (las), etc., or the file can be of any type, such as file generic."* (Amazon)

- **Problem 2: Can the WLMS dataset concept be honored when loading well log, core and mudlog data into OSDU™ Data Platform and is data classification possible comparable to how it is done in WLMS? (Problem identified during OMV discussion)**
  - **Solution for problem 2: Translation of heterogenous, disparate data from various data sources into homogenous, well-known schemas (WKS) similar to classification in WLMS. OSDU™ datasets group data, but in a different way compared to WLMS dataset. (Solution identified during OMV discussion)**

Figure 21 depicts the process of loading data into OSDU™ Data Platform from data sources, relating the data to schemas (WKS), which could be Petrophysics schema, drilling engineering schema or others. WKE (Well-Known Entities) consist of WKS. An example of a WKE is a well. WKS and WKE are coordinated by DMS.

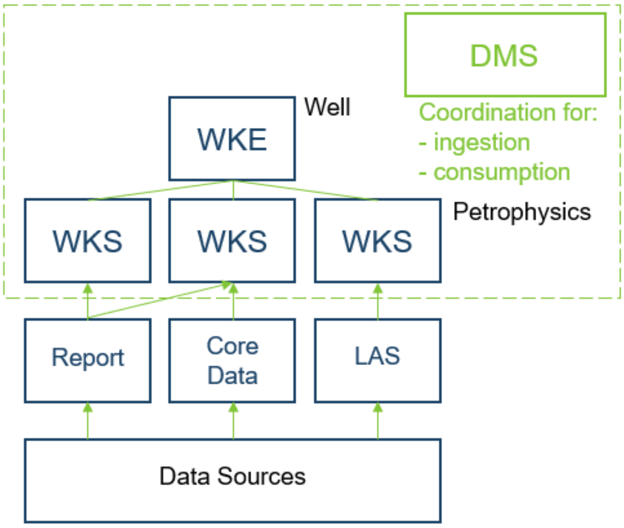


Figure 21: *Data coming from different sources is assigned to Well-Known Schemas (WKS), which in turn are related to Well-Known Entities (WKE). A LAS file is loaded by a Petrophysics schema (WKS), belonging to a well entity (WKE).*

For clarification, please note that WLMS dataset concept is different from OSDU™ dataset. Figure 22 shows how all the WKS schemas (Master-data, Reference-data and Work-product-component) are related to each other and stored in an OSDU™ dataset. The format of the OSDU™ dataset is csv.

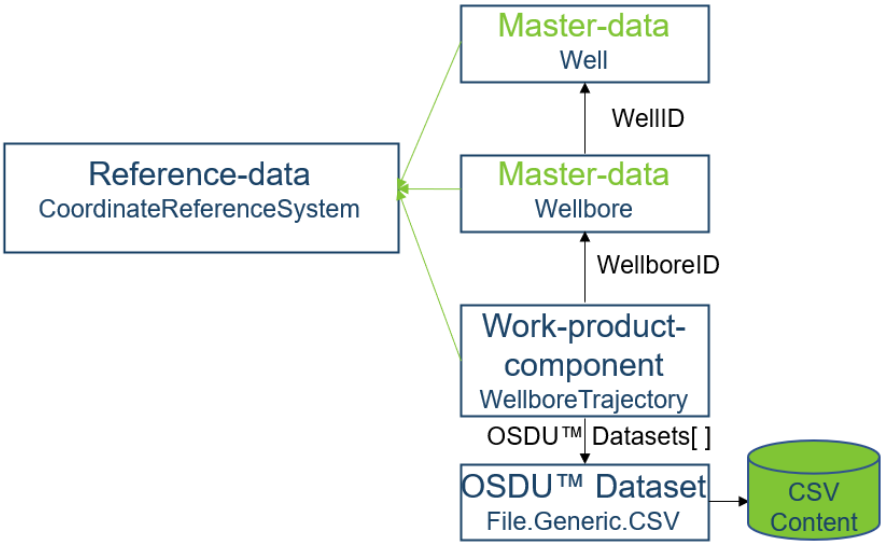


Figure 22: *The relationship between different well-known schemas (WKS) and storage in OSDU™ datasets. Note how the context of the data is not lost, as everything is kept in a OSDU™ dataset belonging to a unique WellboreID (Master-data) and connected via Reference-data. (Kulbrok 2023, April 18)*

Incoming data undergoes the following enrichment transformations:

- **Merging of multiple source entities**
- **Renaming attributes harmonized with standard schemas**
- **Filling data gaps through classification, calculations**
- **Conversion of disparate values to a standard value dictionary with a normalization API.**
- **Frame of reference conversions:** Measurements and Units, Spatial Reference (Feature allowing ingestion in original CRS / spatial reference, Well Elevation Reference: MD to TVD conversions. (OSDU 2023b)

- **Problem 3: How can data duplicates (multiple instances from the same source, e.g.; Field Data (i.e. Intermediate - or Preliminary Data) vs Final Deliverable Data be avoided? (Problem identified during OMV discussion)**
  - **Solution for problem 3: Create an identity for data instances by assigning them to Well-Known Entities (WKE). (Solution identified during OMV discussion)**

Harmonization of different source models into definitive instances (the integration into well-known entitites) is possible in OSDU™ Data Platform. Merging of two sources for the creation of a definitive instance (WKE implementation) between Field WKE and Well WKE is possible. This means that if an oil-and gas field (registered as a unique Field WKE in OSDU™ Data Platform) exists with a number of wells attached to it (Well WKE) and a new wellbore is drilled in this field, the new Well WKE instance is created and linked to the Field WKE. The new Well WKE is added to the existing list of Well WKEs attached to the Field WKE. (OSDU 2023b)

- **Problem 4: How can external data sources such as service contractors or other operators upload data into OSDU™ Data Platform? (Problem identified during OMV discussion)**
  - **Solution for problem 4: EDS (External Data Services) can be used to keep the data library up to date with data coming from external data sources. The OSDU™ API allows direct access of these data sources to the data library. (Solution only partly identified via OSDU™ Forum documentation).**

"EDS enables digital communication between a consumer's OSDU™ platform (i.e., Operator) and OSDU™-aware data sources […]. This allows EDS to transparently pull metadata (master data, work product component, etc.) from the Data Provider into the consumer's OSDU™ platform while leaving bulk datasets, such as files, stored for delivery on demand." (Kulbrok 2023, April 18)

EDS should increase trust in data by ensuring that data consumers work with the most complete, accurate and updated data available. (Kulbrok 2023, April 18) However, according to the quoted statement above, the bulk dataset is stored at source (e.g. Service Contractor) to be pulled by OMV on demand. This would not increase trust in data because data is not automatically synchronized, but rather pulled by demand.

EDS allows to :

- Legally tag data,
- Define master data schemas, such as: Well, wellbore, …
- Define work product component schemas, such as: Well log (e.g. LAS), wellbore trajectory (e.g. directional survey),
- Define reference data: CRS, … (Kulbrok 2023, April 18)

- **Problem 5: It is unclear whether industry well log data formats and support documents (reports or log plots) are supported for ingestion into OSDU™ Data Platform. (Problem identified during OMV discussion)**
    - **Solution for problem 5: Currently OSDU™ Data Platform supports DLIS, LIS, LAS2, LAS3, csv as input files. (Solution only partly identified via OSDU™ Forum documentation).**
      On the one hand, according to OSDU™ documentation, the formats of pdf, csv, txt, docx, xlsx are supported Documents, which can be uploaded to OSDU™ Data Platform as Work Product Components "Document". On the other hand, it is unclear whether support documents can be linked to well logs, in order not to lose the dataset context of a well log and the corresponding support documents. (Chang N., LaRue J., Mosley A. 2023, April 18)

Adoption of OSDU™ Data Platform at OMV: Potential Problems and Solutions with Ingestion of Core/ Cutting Data

Related to the previous problems 1-5 one major challenge of OSDU™ Data Platform arises: the ability to load more unusual data types such as core data.

- **Problem 6: Core data is delivered to OMV in heterogenous formats from different laboratories. It is unclear whether core data loading will be possible in OSDU™ Data Platform (Problem identified during OMV discussion)**
    - **Solution for problem 6: Rock and Fluid Samples DDMS (Solution identified via OSDU™ Forum documentation).**

There is a Rock and Fluid Samples (RAFS) DDMS under development by OSDU™ Forum, which will be donated by ExxonMobil within one of the next OSDU™ releases. (Jackson, J., Clymer, R. 2023, April 18)

Figure 22 shows the structure and functionality of RAFS DDMS. Parquet file format and external parsers are explained.

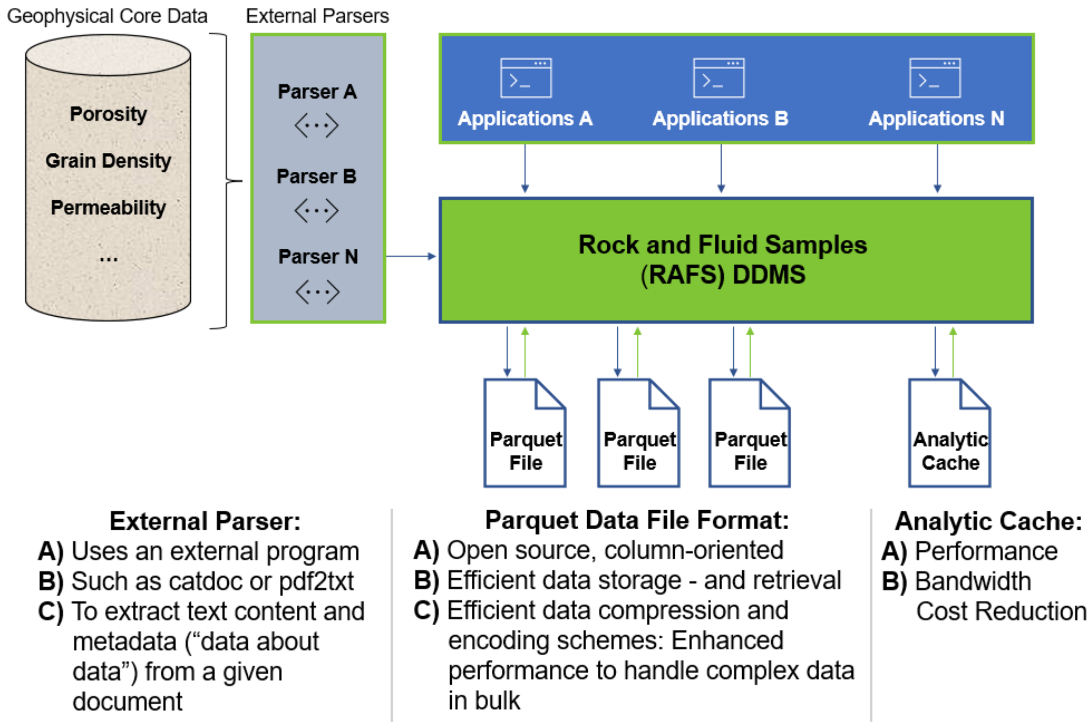


Figure 22: *An overview of Rock and Fluid Samples DDMS (Domain Data Management Service) as an important Data Governance tool in the stages of OSDU™ Data Platform ingestion, curation, and consumption. Geophysical Core data and reports are read by external parsers, which extract text content and metadata. Parquet file format is used for efficient compression, storage, and retrieval of complex data. (Jackson, J., Clymer, R. 2023, April 18; databricks.com; tika.apache.org)*

Parquet file format: *"Open source, column-oriented data file format designed for efficient data storage and retrieval. It provides efficient data compression and encoding schemes with enhanced performance to handle complex data in bulk."* (databricks.com) Parquet is column-based instead of based on rows, which increases storage efficiency in the cloud. CSV data format is based on rows and is therefore not as efficient for cloud storage. Parquet allows for online analytical processing (OLAP) – multidimensional analysis of big data. (databricks.com)

Parser: *"[...] a software component that takes input data (frequently text) and builds a data structure – often some kind of parse tree, abstract syntax tree or other hierarchical structure, giving a structural representation of the input while checking for correct syntax."* (Wikipedia) An external parser is a form of parser which takes an external program to extract text from reports and documents. In the context of OSDU™ core data ingestion, an external parser would help extract text from core reports coming from laboratories. This would occur outside of OSDU™ Data Platform.

Table 8 gives an impression of how core data shall be prepared before it can be ingested into OSDU™ Data Platform.

Preparation before ingestion of Geophysical Core data into OSDU™ Data Platform	
1.) Classification	All Source data is classified. Rich metadata file manifests are also ingested into OSDU™ Data Platform. Core data is classified as such.
2.) Extraction and Transformation	Data is extracted and transformed into pre-defined schemas (WKS) or integrated into data models that were newly created.
3.) Integration and Enrichment	Aggregation of views, infilling with stratigraphy, attributes of core and depth are matched to master data.
4.) Data Quality and Technical Assurance	QC and technical assurance both by experts and automated algorithms before ingestion of prepared data into OSDU™ Data Platform.

Table 8: *Steps in data preparation (curation) that geophysical core data is required to undergo before it can be ingested into OSDU™ Data Platform. (Jackson, J., Clymer, R. 2023, April 18)*

Proposed actions to prepare functionality of core data ingestion into OSDU™ Data Platform at OMV

- Align OMVs' master data for core data with OSDU™ master data for core data: Modify existing OMV workflow for core data ingestion and preparation so that it can be brought to OSDU™ predefined schemas (WKS) for core data.
- Make a use case for visualization of Routine Core Analysis (RCAL) and Special Core Analysis (SCAL) data. A possible feature would be a graphical interface with a comparison between log interpretation results and core analysis results (e.g., porosities, permeabilities) for each well.
- Review possible software options with business as well as IT for integrability. Some options on possible software are listed below:
  - **Consider out of the box software for data ingestion, visualization and discovery by Ikonscience:** Curate Subsurface Knowledge Management Platform has the feature of visualizing wellbore centric data in Well Viewer, data query via Data Explorer and Collaborative Workspace which is a map-based tool for well data.
  - **Software for data ingestion (unstructured data):** Microsoft form recognizer
  - **Software for data visualization:** INT IVAAP Cloud Viewer or PowerBI for Rock and Fluids to visualize OSDU™ Core Data
  - **Software for data query:** Aspen Geolog to search OSDU™ Data Platform for available data. (Jackson, J., Clymer, R. 2023, April 18)

New and Backlog Data: Options for creating a Single Point of Data Entry for Wellbore Centric Data in OMV

To establish a single point of data entry for new and backlog data within OMV and OMV Petrom via OSDU™ Data Platform, the use case with OSDU™ Data Platform is presented. Additionally, a generic way of ingesting data into a Hybrid Data Lake is described.

General procedure: Single Point of Data Entry via e.g.; Microsoft Azure Data Platform (Data Lake combined with a Data Warehouse): Usage of a Data Lake (e.g. MS Azure) for allowing data sources to upload data via a data dropbox, to gather data from all data sources. The combination between Data Lake (structured and unstructured data) and Data Warehouse (data storage system for structured data) would be a Hybrid Data Lake (Data Platform).

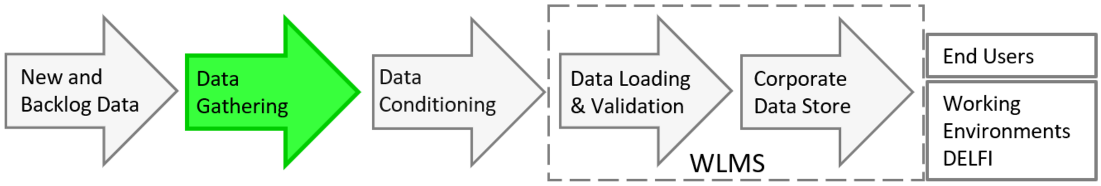
Use Case: Single Point of Data Entry with OSDU™ Data Platform: OSDU™ Data Platform is a specific form of Hybrid Datalake (Data Platform). OSDU™ Data Platform can be used as a single point of data entry, if it is fully adopted by OMV.

In Table 9, the advantages and limitations of a generic Data Platform and OSDU™ Data Platform for the purpose of creating a Single Point of Data Entry are listed.

Data Platform as Single Point of Data Entry	**Advantages**	**Limitations**
Generic Data Platform	+ Simple architecture that is easy to implement + Dropbox data submission is easily feasible for all data sources in the cloud + Easy implementation of buffer stages to avoid a data swamp: E.g.: Raw / Clean / Curated / Ready	- Incompatibility with Techlog - The Hybrid Datalake needs to be managed and curated by someone - Limited data discovery by users - Data Enrichment has to be enforced via governance
OSDU™ Data Platform	+ Easy to implement as PaaS in OMV with MS Azure + High customizability of OSDU™ Data Platform according to OMV business needs + Data loading notification system exists + Internal network effect of OSDU™ Data Platform: One OMV department with OSDU™ expertise can support other departments on their path towards OSDU™ Data Platform	- OSDU™ Data Platform needs to be managed and curated by someone - Contractors / BOs need to gain access and expertise to OSDU™ Data Platform for data submission - People required to maintain and understand OSDU™ Data Platform in OMV & HOL

Table 9: *Advantages and limitations of OSDU™ Data Platform (consisting of a Data Lake and Data Warehouse) for the purpose of creating a Single Point of Data Entry. Additionally, for context, advantages and limitations for generic Data Platforms in E&P companies are given.*

3.2 Standardized Data Gathering



"The most difficult challenge in data quality is completeness of the delivery. Who defines the deliverables? Intuitively, the data users should be the first persons involved in defining logging [service contractor's] *deliverables. [...] . The company must ensure that all basic and auxiliary data are delivered. The similarity with the purchase of a car has often been used. The buyer does not need to specify to the car manufacturer that a clutch should be part of the car, but the car company definitely needs to make sure that this clutch is designed and present in the car. (Theys 2011)*

In OMV, the data users (e.g., Petrophysicists) are involved in defining what data should be acquired and they verify if the data delivered by service contractors are complete.

Some pain points when it comes to standardized data gathering are listed below:

- **Same data/ partly same data are located at different locations** (e.g., BO versus HO).
- **Different versions of data are existing** (e.g., preliminary data versus final data; several final data packages with small differences) and it is not always clear what are the correct data to be used by end users.
- **Lack of resources for gathering correct data in HO and BOs.**
- **High variability in data quality and heterogeneity:** data does not always arrive in the required industry standard format. Data originates from different companies and service contractors with different quality standards and requirements. For some data (e.g., core data) there is no industry standard, and this data type is normally delivered in an unstructured format which complicates the data gathering and preparation process.
- **Data rejection during upload:** data does not always follow industry standards or mistakes in data formatting causes the data to fail during loading (e.g., a file that is not LAS format conform is rejected).
- **Manual generation of data acquisition summary report:** The report is required to verify that all data acquired in the borehole relevant for WLMS are available in the system for end users within a defined time range. Automatization and integration with other processes (e.g., Drill plan) would standardize the process of creating the report and reduce manual efforts.

- **Ambiguity on Contractor data gathering responsibilities:** Contractor has to provide additional information (information not automatic extractable from data itself) during the data submission process so that the data can be properly classified.
- **Manual effort in updating the dictionary:** Tools of service contractors change constantly and therefore, the dictionary must be kept up to date to represent the latest tools and curves generated by the tools.
- **Standardized fully automated data gathering missing.**

The future goal is to share common standards on how well log data shall be delivered with all the contractors who provide services for OMV, OMV Petrom and all its' affiliates. When it comes to LAS and DLIS files that are provided by the contractor, it is important to have well-defined attributes and headers containing all necessary information (e.g., borehole UBHI, service date). However, overcomplicating regulations and standards for service providers should be avoided. This will lead to reluctance of contractors to fill out and populate necessary information.

Currently, standardized classification at OMV is achieved in WLMS via **Data dictionaries** for different service contractors' specific dictionaries were generated including service contractor specific logging tool (including different toll versions) and relevant curves generated by each tool. All dictionaries are embedded in the system and used during the data loading process to ensure consistent data classification and naming conventions. In the **Data dictionaries**, information is available on:

Tool family description: Main logging tool type, number, method combination, method combination mnemonic, raw data category (conventional, advanced, …).

Logging tools: tool description, tool family, OMV name (normalized name), whether it is a cased hole (CH) or open hole tool (OH), tool type (to simplify log name and to have a service contractor independent log naming approach; this simplifies data search for Geologist and other subsurface experts who are not so familiar with the service contractors specific tool naming conventions).

Logging channels per tool: tool specific generated channels/ curves (includes curve naming convention to be used for log curve digitization) and OMV specific curve naming conventions (e.g., composite, standard log, CPI).

Overall, consistent data classification and dictionaries (e.g., tool and curve dictionaries) are the essential basis for consistent data usage and subsequently for automatized data flows and workflows. Classification attributes are used to differentiate High Company Value Data from RAW data. For example, when a standard log is uploaded and classified/ published as such, the standard log is transferred to the working environment.

A data lake is one distributed semi-organized folder structure with data structured around one well, basin, etc. … Gathering data honoring the dataset concept and storing the data while honoring the dataset concept can be facilitated by having an architecture of the data lake structured in the same way as data is structured inside an incoming dataset.

As mentioned in Chapter 2.3 Data Conditioning, currently in OMV, data gathering is designed around fulfilling the following requirements to collect data in a structured, complete way:

Data gathering and structure:

- Borehole (UBHI, based on CWDB);
- Data Type (Log, Mudlog and Core, Image, etc. ...);
- Data grouped per dataset concept (e.g., final delivery package from Service Contractor for a specific logging run, core delivery package based on a specific order, Standardlog dataset, CPI dataset);
- Data quality, correctness and completeness must meet the OMV defined standards. (OMV Energy)

The following questions should be asked by organizations when it comes to gathering data and the amount of trust data consumers have toward the gathered data:

- What data is critical?
- Where did the data come from? Where else is it used?
- Who owns the data? Is there a Data Steward one can hold accountable?
- How "fresh" is the data? When was the last time it was populated and updated?
- Are working environments populated with the latest High Company Value data?

(Brown 2021, April 22)Standardized data gathering enhances the trust in data, by making data discoverable, measurable, and traceable.

Three Critical Factors for Data Trust

"We were around when a lot of the technology transition from analog to digital happened. What we began to recognize pretty rapidly was that the minute something was put on a computer, it was assumed to be correct. And we have lived through that issue for a long time […] just because it moved from analog to digital, it does not mean it has gone through quality data checks. If you have not done quality data checks then you cannot be certain that you are getting high quality data. Just because it has been turned into digital. Just the same is valid for AI: just because AI is making decisions based on data, how do you trust the information that comes out is good?" (Jim Crompton, Tim Coburn, John Jacobs)

Factors 1. - 3. are critical in order to have data consumers trust the data:

1. **Discoverable**

"I'm not even sure what data we have or where it is at. There is no single source of truth that I trust." (Brown 2021, April 22)

- Centralized location to catalog and manage metadata
- Figure 14 and 15 in Chapter 2.1 OSDU™ Data Platform describe how OSDU™ Data Platform indexes incoming data to make it discoverable by query for certain data attributes.
- Data can be tagged according to business domains, e.g., Low Carbon Business (LCB), so that it can be discovered by users in that context.

2. **Measurable**

"We have lots of data. The problem is I don't know how good it is." (Brown 2021, April 22)

- Established metrics for data quality and governance
- Understand if metrics fit your data purpose
- Data quality-tagged assets that span the enterprise. Tagging the data quality of datasets throughout the enterprise gives users information on how much they can trust a specific dataset in terms of correctness, timeliness and completeness (i.e. whether they are working with correct, up-to-date and complete data).
- Data audit metrics custom-tailored for OMV have been defined in Chapter 3.3 Data Audit Tool for evaluation of completeness and timeliness of the data.

3. **Traceable**

"I can find my data and see that it is good. I just don't know where it comes from and goes." (Brown 2021, April 22)

- Track lineage and impact of data
- See Chapter 4.1 E2E Data Flow Diagram for data lineage in OMVs wellbore centric data workflows.

Data Gathering powered by Workflow Engines and Data Ownership Roles

In order to make data discoverable, measurable and traceable, a holistic data gathering system is required. This data gathering system must be able to capture data independent of its origin (operated, non-operated boreholes, traded data, …), be aligned with the audit system, OSDU™ Data Platform / WLMS and must honor the dataset concept (see Chapter 2.2).

- Data is coming to OMV due to specific reasons and due to certain processes. It is important that data is captured where it is generated. It must be ensured with proper set up via workflow engines that the data ends up where it should be. For example a borehole is planned and with that the data acquisition begins. This information is captured in Drill Plan. Here it would make sense to enforce that via relevant workflows/ processes that all relevant data are properly entering OMV and OMV Petrom, that data is quality controlled again by the domain expert in charge, loaded to the leading system and transferred to the working environments if relevant.
- For a workover job this looks different and must be as well handled different via another process/ workflow. The important thing is that data are not forgotten and that the process ensures that data are loaded to the correct leading system. Figure 23 shows a concept ("4 – Eyes – 4 Grids: Holistic Data Gathering powered by Workflow Engines") on how such a system could look like in the future.
- Note that the data gathering approach has to look different depending on where the data is coming from, depending during what process/ workflow the data was generated,whether it is an operated or non-operated borehole, etc.. The data sources in Figure 23 align with the data sources presented in Figure 20 and represent global data sources of OMV, OMV Petrom and its affiliates.
- It should be also noted that Workflow Engines for data gathering are designed to be complimented by a Data Domain Specialist who is physically located in the respective Branch Office (BO), where the data has been generated. Ideally, this is the person that was also involved in the data acquisition process, e.g. the Petrophysicist present on site during a well logging job done by a contractor.

The roles in OMV might differ slightly from the official Data Management Association (DAMA) role definitions. DAMA has set up definitions of the following two roles: Data Steward and Domain Data Owner.

Data Steward responsibilities:

- Responsible for data handling processes and relavent regulations
- Ensurance of quality and use of organization's data assets,
- Creation and documentation of business quality control rules and data standards,
- Ensure that processes and regulations are followed
- Tracking and identification of data gaps in ongoing projects as well as re-use of data assets across projects,
- Responsible and managing relevant system adaptions to fulfill current and future business needs (e.g., system optimizations, automatization)
- Management of data quality issues and standardization rules. (Sykora 2017)

Domain Data Owner responsibilities:

- Accountable for data, higher seniority than Data Steward,
- Business professional who is accountable for a data subset in their data domain,
- Cooperation with Data Stewards to define data quality and business rules,
- Authoritty for approval of decisions regarding the data in their data domain,
- Ensures data for subject area is fit for purpose (Sykora 2017).

Table 10 is complementary to Figure 23 and definitely assigns the different ways of gathering data ("grids") to the different data sources. Also, the different ways of making sure that data is loaded in a complete and timely manner ("frames") are listed for each data source.

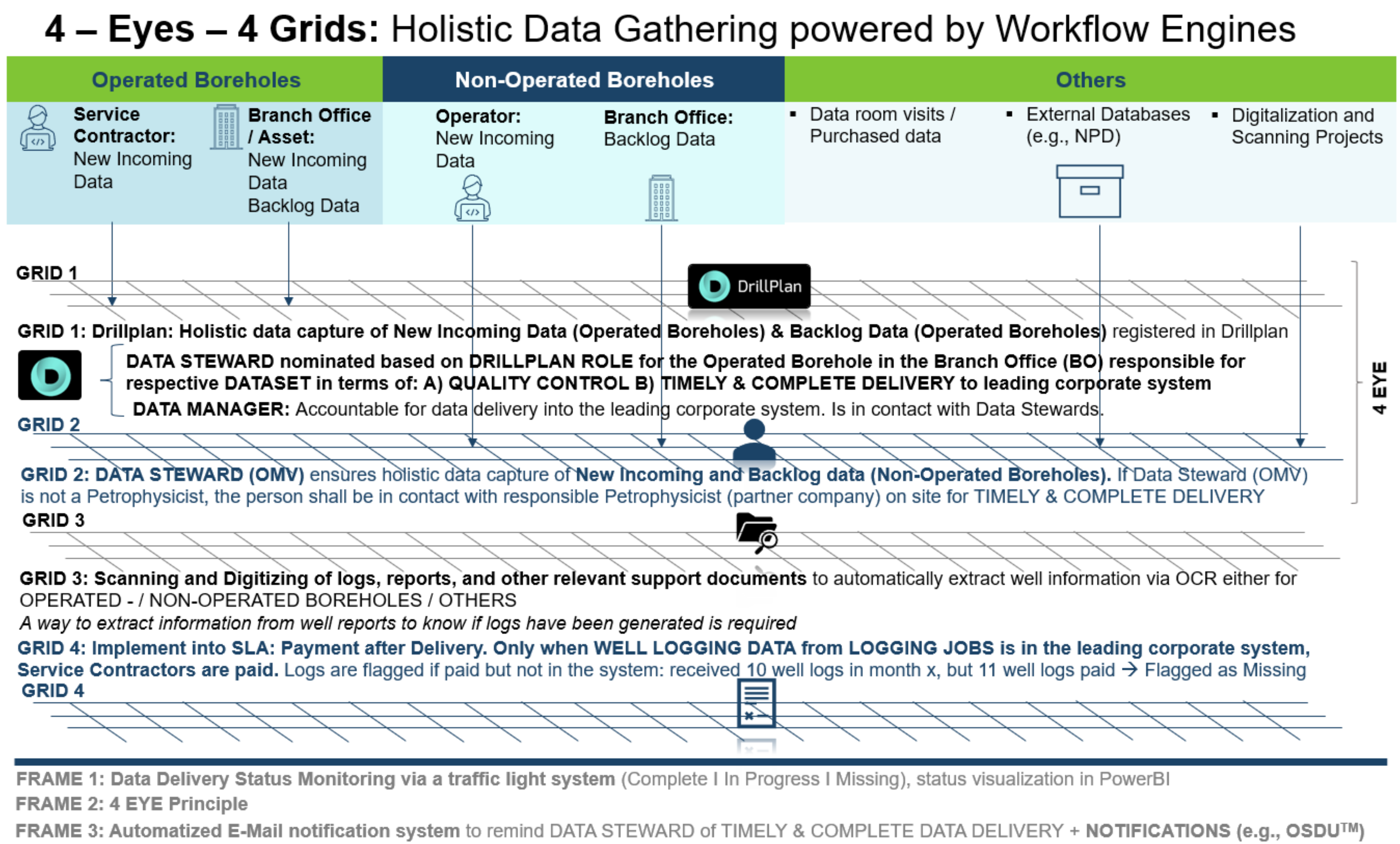


Figure 23: *4 – Eyes – 4 Grids System for holistic data gathering. Data is captured from all sources via four grids. Grid 1 is based on DELFI Drillplan, Grid 2 relies on cooperation with a Data Steward on site in the respective Branch Office (BO), Grid 3 relies on OCR scanning of reports and optional Grid 4 aims to combine subsurface data with financial data in order to identify missing data. These four grids are supported by three frames. Frame 1: Status monitoring via Drillplan. Frame 2: 4 eye principle via Data Steward and Data Owner. Frame 3: Automatized E-mail notification system combined with OSDU notifications.*

Generic Type	Affiliate	Data Type	Applicable Grids	Applicable Frames
Operated Boreholes	Service Contractor	New Incoming Data	Grid 1 Grid 3 Grid 4	Frame 1, 2, 3
	Branch Office / Asset			Frame 1, 2, 3
		Backlog Data	Grid 2 Grid 3	Frame 1, 2, 3
Non-Operated Boreholes	Operator	New Incoming Data	Grid 2 Grid 3	Frame 2, 3
	Branch Office	Backlog Data	Grid 2 Grid 3	Frame 2, 3
Others	Data Room Visits / Purchased Data		Grid 2	Frame 3
	External Databases (e.g., NPD)		Grid 2	Frame 3
	Digitalization and Scanning Projects		Grid 2	Frame 3

Grid	Grid Description (Holistic Data Capture and Gathering)	Frame	Frame Description (Timely and Complete Delivery of Data)
Grid 1	Drillplan: Holistic Data Capture & Automatic Role Assignment	Frame 1	Status Monitoring Drillplan: Traffic Light System
Grid 2	Nominated Data Steward	Frame 2	4 Eye Principle
Grid 3	Digitizinig logs, reports for Holistic Data Capture & OCR	Frame 3	Automatized E-Mail & Notification System
Grid 4	SLA in Logging Contracts: Pay only for data delivery		

Table 10: *Different categories of affiliates and data sources are presented above, with the respective grids / frames that will be most effective / relevant in gathering the data listed. This classification can be helpful in setting up a way forward to plan resources and assign roles. SLA from Grid 5 is detailed in Chapter 3.3 Data Audit Tool.*

The objective of the four grids in Figure 23 is to make sure that all data from disparate data sources is gathered and no data assets are overlooked. The objective of the three frames is to ensure that data arrives complete and within the defined time range. Data gathering activities must be monitored / auditable. The data flow must be traceable, discoverable and automatized as much as possible.

Complementary to this data gathering system is the folder structure of the datasets ("Dataset concept"). Figure 24 shows how this folder structure looks like.

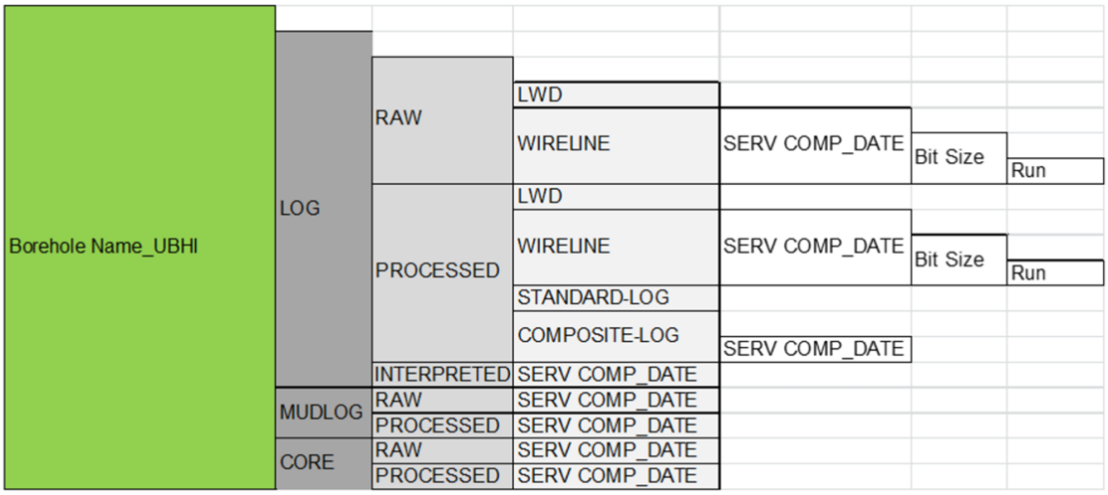


Figure 24: *The Dataset Concept is honored by structuring data according to borehole, subsurface data type, as well as logging tool, bit size and run.* *(OMV Energy)*

The advantage of pre-structuring the dataset in folder-structure for all boreholes belonging to a field is that it becomes easier to:

a) Identify the right WKS and WKE when ingesting the Dataset into OSDU™ Data Platform (already conform with OSDU™ structure around boreholes and fields),

b) Automatically identify any missing data (e.g., reports, LAS, DLIS files or plots) in the respective folders, but needs to be checked as well by Domain Specialist as data deliverables can vary based on actual data acquisition agreements.

c) Set a specific status (complete / in progress / missing) for a certain folder.

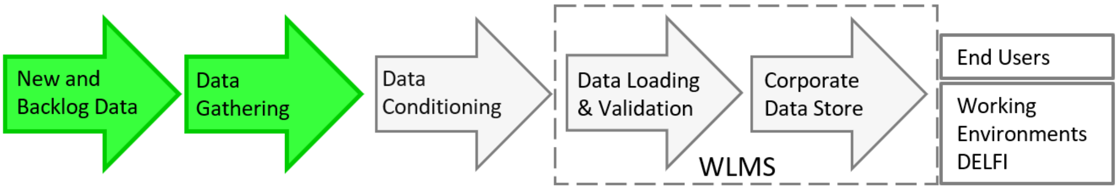
d) A folder pre-structure could be generated based on the planned data/ actual data acquisition program reflecting data which should come in as minimum.

Example case for automatization in data gathering: A LAS file has been reported missing (Borehole 3), during execution of a wireline logging run.

- Domain Data Specialist must have a look over the data in any case as the person has to check the data completeness and data quality considering the actual data acquisition requirements. For this example case, let us assume that even though the Domain Data Specialist had checked data for completeness, one LAS file has gone missing.
- Automatized missing data identification: The Dataset folder structure (Figure 24) helps automatically identify that in the folder "Borehole 3", in the "Wireline" folder, no LAS file was found for a certain run in Borehole 3,
- Automatized role identification via Grid 1 (Drillplan): From the HO, identify the borehole drilled and via Drillplan, the responsible Data Steward can be contacted,
- Automatized borehole identification via Grid 3 (OCR scanning of reports): Identify the name of the borehole as well as if -, or which data have been acquired (e.g., logs)

One limitation of the 4-Eyes - 4 Grids holistic data gathering powered by Workflow Engines is that workover data cannot be easily captured. One idea for the future to achieve a tracking and gathering of well logs conducted during workover would be to automatize the creation of workover programs. This workover automatization would enable digital capturing and monitoring of workover programs and therefore tracking of any well log exectued within workover programs. This way it could be possible in the future to identify well logging runs that have been executed and check, whether the outputs from the workover well logging program are reflected in the well logging data uploaded to the leading corporate system.

3.3 Data Audit Tool



Audit pain points are listed below:

- **Trust in data** is not yet fully developed.
- **No contractual data delivery obligations:** Contractors are paid, even if the data has not (yet) been submitted WLMS system. A better workflow could look like this: Data obtained by contractor from operated wells → QC by Petrophysicist → publishing of data in WLMS → Payment of contractor only once data has arrived in WLMS.
- **Unestablished completeness reporting responsibility by branch office:** It should be the responsibility of BOs to confirm that data (final service contractor deliverables) has been uploaded to the system within defined timelines.
- **Interpretation results (end of well products, final study output results) are not uploaded to the system.** Completeness reporting for High Company Value data (inhouse generated by specialists) is missing and upload is not enforced.
- **An email follow-up system is missing**, in which escalation takes place if data is not delivered in a timely and complete manner to the system.

The central question of this chapter is: How to enforce data delivery on domain specialists (e.g., Petrophysicists) at OMV and OMV Petrom as soon as a new data has been acquired/ generated? Via Audit? Systems? Management? Contracts/ Service Level Agreement (SLA)?

Probably a combination of audit, systems and top-down enforcement from management side would help. If OMV has the size to influence the contractors: putting into the contract what data (e.g., logs) need to be delivered in format x, then a lot of preprocessing is handed over to the contractor to save time and money. Beforehand, strict definitions on how the data must be delivered need to be established. When the contractor is required to deliver data in unique formats, the contractors' capabilities will probably be insufficient to do so, but it can be expected for long term contractor partners to deliver data the desired way. This would enforce the Service Contractor to pay better attention to the deliverables and deliver better quality-controlled deliverables. Giving monetary incentives to contractors will most likely not lead to constructive results. OMV and OMV Petrom personnel must be very diligent in QC during data acquisition or final data delivery.

Operated, New Boreholes: Event-driven, Automated Auditing for Data Gathering

An example for an event-driven, automated auditing for operated, new drilled boreholes to ensure all defined data end up in the corporate system and working environments within defined timelines is proposed below in Figure 25.

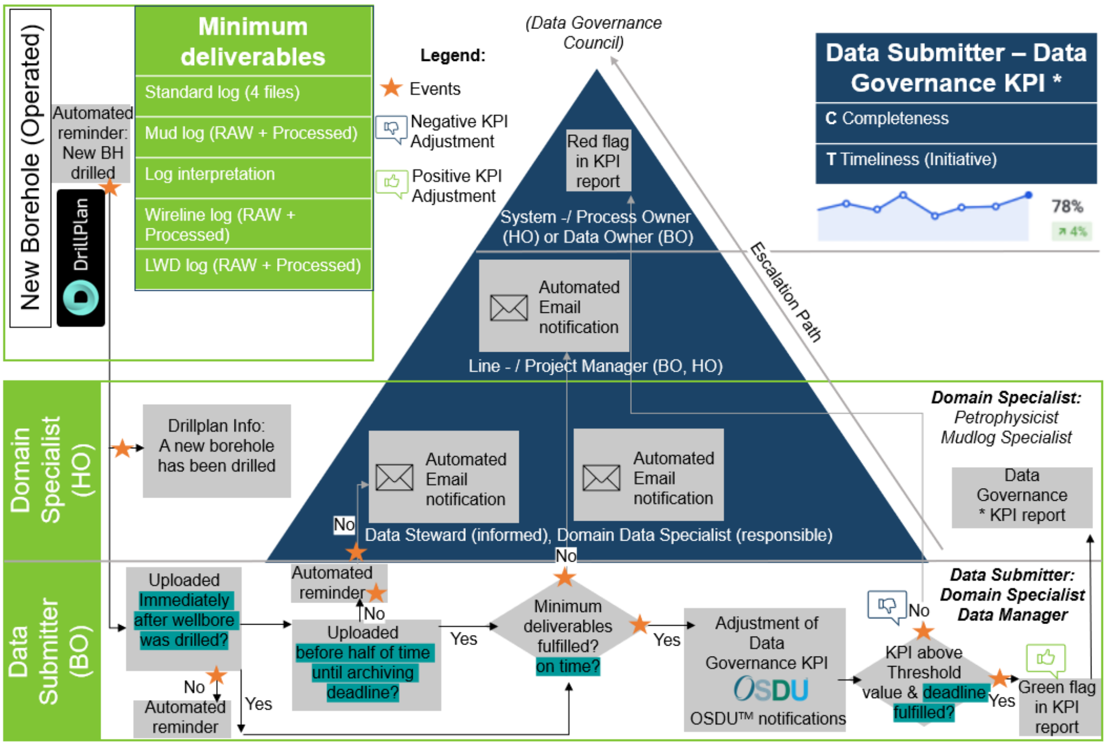


Figure 25: An event-driven, automated audit system for Operated, new drilled boreholes for the delivery of well log and mudlog data. The system combines a) Data audit framework b) Automated email notifications and c) Management involvement. The audit system is also aligned with the data gathering system mentioned in the previous Chapter 3.2. Timelines for data archiving can be looked up in Figure 28.

As Figure 25 contains many roles, please find below definitions of OMV roles in the data domain:

- **Data Owner** is located in the Branch Office (General Manager in BO): This person has to ensure that data are uploaded timely to the leading system.
- **Domain Data Owner:** This is the person with the technical authority. This person defines how data must be handled within the organization, data quality aspects and generated relevant processes and regulations.
- **Domain Specialist:** This person is involved in the data acquisition definition and QC and has to ensure that the data received by OMV is complete and correct for specific activities.

- **Data Steward:** This is a new role in OMV. The following questions are to be answered: What kind of skills does this person need to have? What level of Domain knowledge does this person need to have? Please refer to Chapter 3.2, where the DAMA definition of a Data Steward is given.

Regarding figure 25 and figure 26: First, there is an automated reminder for the Data Submitter (Domain Specialist or Data Manager) to deliver the data as soon as a new borehole has been drilled.

If the data is not submitted in a defined timeline (e.g., half the time before deadline for data archiving), the **following escalation path** is valid for figure 25, figure 26 and figure 27:

- Reminder to Data Submitter e.g., half the time before deadline for data archiving. If the data is not uploaded, there is a further escalation to the Line – or Project manager.
- If, however the data is not delivered beyond the deadline for archiving, this violation is flagged in a KPI report (data has not been delivered at all, i.e. within the timeframe when it should have been delivered).

While managing data of operated wells is rather straightforward, non-operated wells place a challenge because operations are not managed by OMV and its affiliates, but rather by the relevant operating company. OMV has a set of standards for operated boreholes regarding data capture custom-tailored to its business rules. In non-operated boreholes, the operator is still required to make data available to the partner, - OMV, but in a format chosen by the operator which could differ from OMV standards.

Figure 25 (well log and mudlog data delivery) and figure 26 (core data delivery) are depicting the minimum deliverables as well as final minimum deliverables for timely and complete delivery of data from operated boreholes. Figure 26 also contains final minimum deliverables for non-operated boreholes (core data delivery). For core data, timelines for delivery are slightly different, as core analysis programs sometimes take several months and up to years.

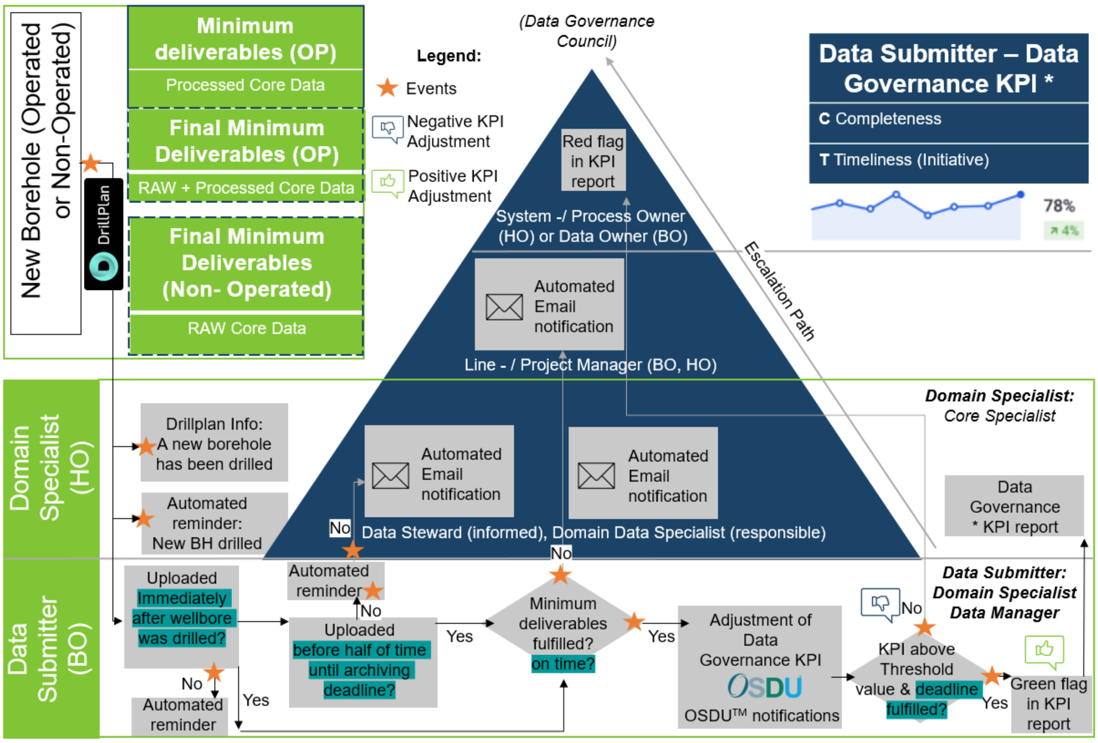


***Figure 26:** An event-driven, automated audit system for operated and non-operated, new boreholes for the delivery of core data. The system combines a) Data audit framework b) Automated email notifications and c) Management involvement. The audit system is also aligned with the data gathering system mentioned in the previous Chapter 3.2. Timelines for data archiving can be looked up in Figure 28.*

Non-Operated and Operated Boreholes - Backlog data: Event-driven, Automated Auditing for Data Gathering

Figure 27 focuses on the audit system and relation between Data Steward (Data Submitter in the Branch Office, BO) and Domain Data Owner (Domain Specialist in the Head Office, HO) for non-operated and operated boreholes regarding backlog data.

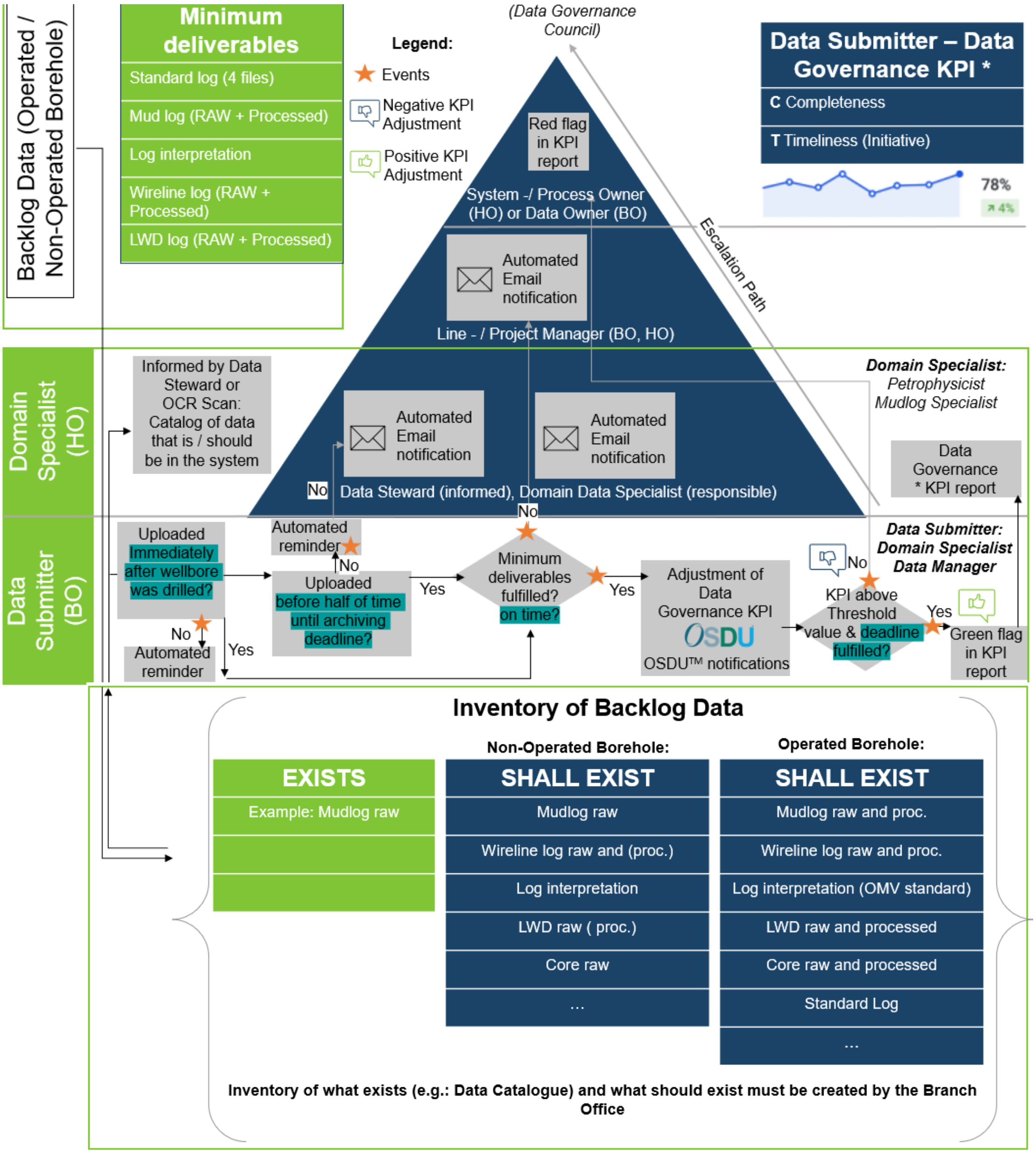
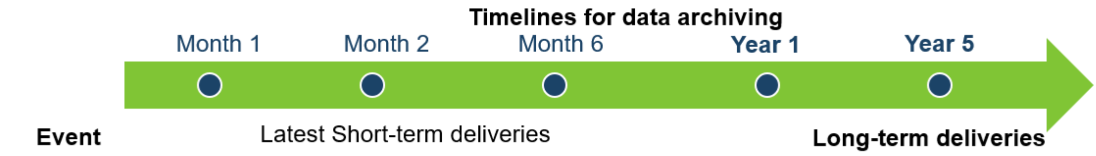


Figure 27: *An inventory of backlog data ("Exists") is created by the Data Steward. The inventory of existing backlog data is then compared with the data that "Should Exist" regarding non-operated as well as operated boreholes. Determining which data should exist is done in different ways for the two cases highlighted above. Timelines for data archiving can be looked up in Figure 28.*

Figure 28 is a summary graph on the different timelines of when the data coming from disparate data sources (laboratories, well logging contractors, …) shall be archived in the leading corporate system. The timelines for data archiving from Figure 28 are closely related to the data audit triangles (Figure 25, 26 and 27). In the data audit triangles, management is informed via KPI reports (timeliness, completeness) if data is not archived according to the timelines presented in Figure 28.



Event	Month 1	Month 2	Month 6	Year 1	Year 5
Borehole reached TD			Raw Log and Mudlog data		
OMV/OMVP processing / Data Retrieval	Processed Log and Mudlog data		Core/Cuttings data		
Receiving / completing final data package	Interpreted Log and Mudlog data	Digitized Log and Mudlog data			
Data room visit / Trade			Log and Mudlog data (Low volume)	Log and Mudlog data (High volume)	
Final Lab report package delivery		Core/Cuttings data			
WLMS Go-live					Historical Mudlog, - Log- and Core/Cuttings data (Backlog- , Digital available-, Hardcopy data from physical archives)

Figure 28: *Timelines for data archiving, according to WLMS regulations. On the left-hand side are different events and in the center of the graph the type of data (orange) and the exact time horizon (1 month / 2 months / 6 months / 1 year / 5 years) when the type of data must be archived is denoted at the top by a green arrow. (OMV REAL Regulation)*

3.4 Filling Resource Gaps for E2E Workflow

In addition to having the 4-eye principle between Domain Specialist (BO, HO) and Data Steward (HO or BO), the integration of a Wellbore Centric Data Integration Group is important for a number of activities associated, such as Data Gathering, OSDU™ Data Platform implementation and maintenance relevant for specific functions for WLMS data types, supporting Branch Offices with all WLMS activities and data conditioning projects or data loading plan (short - and long-term plan). Figure 29 shows the proposed associated roles and their activities. A Data Steward supports the QC process only and does spot checks. QC of new incoming data must be done in the BO by the relevant Domain Specialist. Together with the Data Manager the data should be loaded and approved in the system. There must be as well a person in the BO who is the WLMS key user. This person plans together with the HO Data Steward all WLMS related activities (data loading plan, budget and resource planning, …)

Please note that WLMS currently has following roles:

- Data submitter
- Data Domain Specialist
- Data Manager
- WLMS Key User in branch office
- Data Owner (General Manager in BO responsible that WLMS regulation is followed
- Data Domain Owner (Functional Technical authority)
- Data Steward

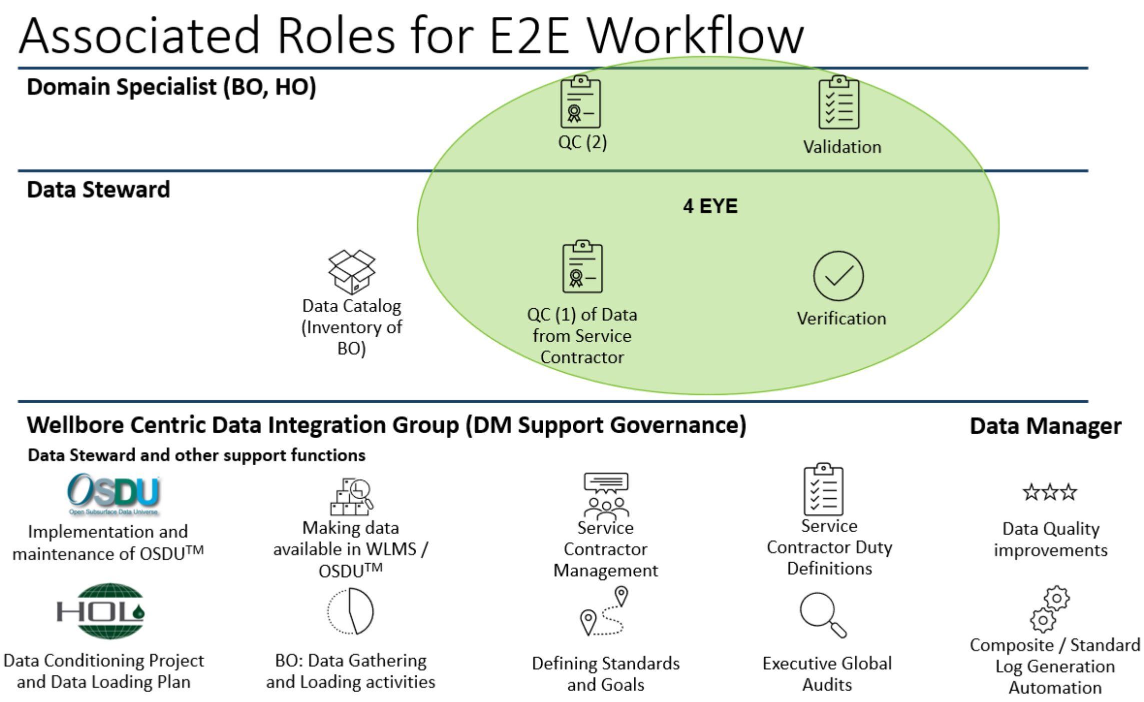


Figure 29: *Associated roles with the data gathering system that are also partly involved in quality controlling the data from the service contractor, as well as service contractor management. Supporting this service contractor management with service level agreements (SLAs) can give a strong position when it comes to making sure that data is delivered in the expected data format and structure.*

Wellbore Centric Data Integration Group should exist of a key team with clearly defined responsibilities and skills, including programming skills and data science skills. Please find below a non-exhaustive list of activities in which the Wellbore Centric Data Integration Group must be involved:

- Content management of existing WLMS system, e.g., quality control on the database level, dictionary management,
- Providing High Company Value Data to end users in the working environments,
- Branch office support for relevant questions,
- Execution of global audits and reporting to ensure that relevant processes and regulations are followed,
- Definition of new standards and quality control rules,
- Permanent data quality improvements on existing data in the database,
- Further automatization (e.g., Composite - and Standard Log generation),
- Further development and optimization of data conditioning workflows,

- Definition of duties of HO versus BOs,
- Definition of roles of HO and BO,
- Definition of duties of Service Contractor/s; set up and execution of data conditioning and loading projects,
- Execute audits on global level (inputs from all BOs),
- Lead and execute data conditioning projects,
- Service Contractor Management: SOW, data deliverables quality control, project follow up, data exchange, etc.,
- And others…

The following questions are relevant regarding the roles of the HO and BO in filling these E2E workflow resource gaps:

- What kind of work can be executed by the branch office personal considering skills, available time and budget?
- How does the individual interaction look like between the relevant branch office and HO?
- How will the charging of services to BOs work? For example, charging services to BOs by distributing the costs of the HO team based on a key to relevant BOs needs to be considered.
- Who is responsible for all data in HO? Who is responsible for all data in HO from countries where OMV is currently not active anymore?

As minimum requirement towards the BO, data managers must be made available for backlog data collection (based on data collection concept) and providing this data to HO for conditioning and loading. A short- and long-term loading plan shall be worked out between HO and BO to ensure that data are uploaded to WLMS in timely manner and according to the WLMS regulation. For new acquired data the responsible Data Domain Specialist and Data Manager from the branch office are responsible that data are uploaded to WLMS within given time limits.

The Way Forward: Actions for implementation of the new Data Gathering and Audit System

Some recommended actions for holistic, standardized data gathering and audit that minimizes frictions are listed below to reach the goal of a holistic data gathering and audit system:

- **The final goal** is then to combine the new a) Data Gathering System with b) The Audit System, considering c) the timelines when the relevant data shall be delivered in the leading corporate system.

- **Most difficult aspect for non-operated boreholes** is implementing a backlog data gathering system and finding a Data Steward (Grid 2), because it is:
  a) Hard to identify the Petrophysicist who was involved in acquisition of certain logs of a non-operated borehole, as well as finding a suitable person from OMV who will act as a Data Steward at a non-operated borehole,
  b) Time consuming to identify all the backlog data of the non-operated borehole.

- **Required information technology & software implementations:**
  a) DrillPlan: Automatic extraction of roles data for OMV operated wells and important non-operated boreholes ("Who will be responsible for the generated data in the role of Data Steward"?).
  b) Also, there must be a responsible person for other non-operated boreholes which are not necessarily captured in DrillPlan,
  c) Transfer of role data to a database system where the KPIs of timeliness and completeness are stored as benchmarks of how well the Data Stewards are performing in their roles, stored together with the contact e-mail address of the individual Data Steward,
  d) Forwarding of this database system information to e-mail system, in which automatic notification emails are sent based on violations against timeliness and completeness KPIs, as well as KPI reports generated automatically that are sent to management and automatic notifications to management by email if data was incomplete or not delivered on time,
  e) OCR scanning software: Could be potentially done with HOL, as they already have the necessary in-house experience with OCR technology,
  f) Automatic extraction of payroll data and automatic comparison with existing data in WLMS/ OSDU™ Data Platform. ("What should be existing in the leading corporate system vs.: What is already existing in the leading corporate system."),
  g) Automatic E-Mail notification system and OSDU™ notifications.

- **Six sigma approach applied to data gathering and data delivery audit** revolves around the following elements:
  - Define and plan – what is the problem at hand?
  - Measure – measure current state using the metrics defined
  - Analyze – identify causes, feasibility …and plan
  - Improve – implement and measure again
  - Control – identify a business owner for the ongoing process of monitoring (Wikipedia)

Six sigma approach is depicted in Figure 30. It is split into three dimensions, dimension 1: Define & Plan, dimension 2: Measure & Analyze, dimension 3: Improve & Control.

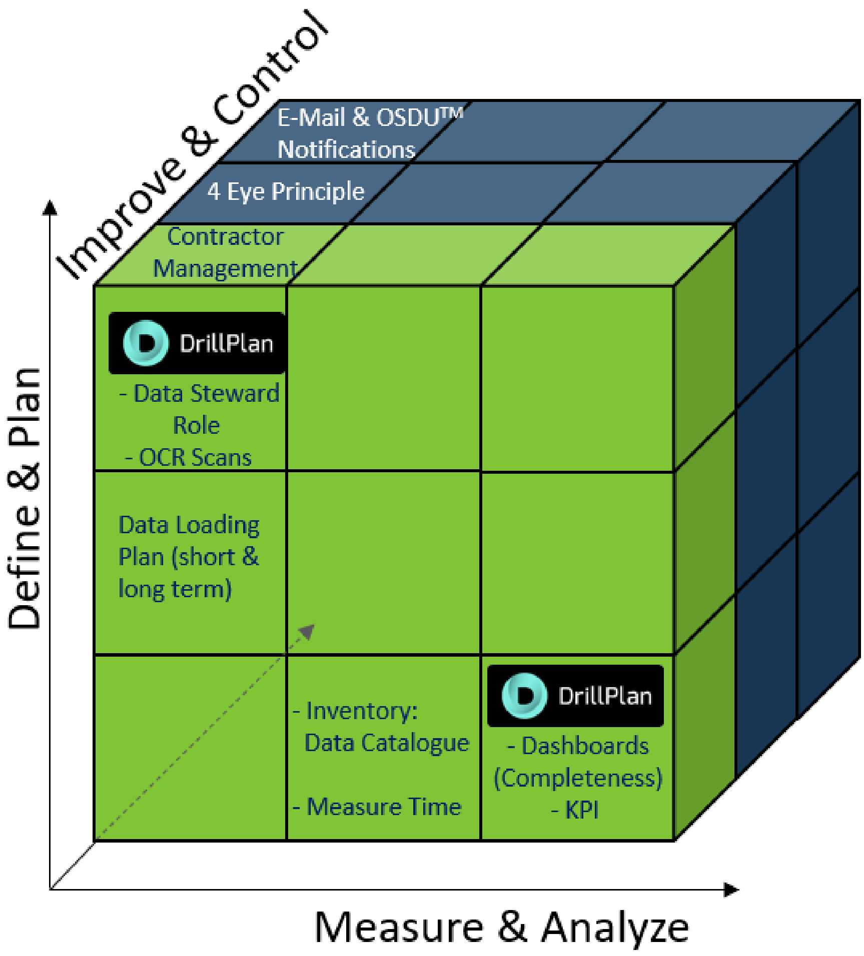
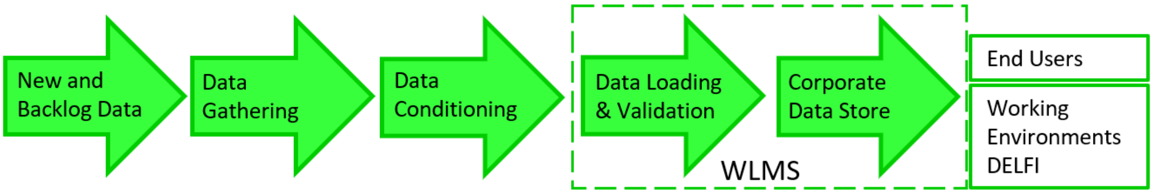


Figure 30: *Six sigma approach for data gathering, - delivery and - audit. Note how each axis measures different dimensions and how the dimensions are linked to each other.*

Chapter 4 Data Lineage and Information Architecture



4.1 Data Lineage - E2E Data Flow Diagram

The data flow of wellbore centric data in OMV and OMV Petrom and its affiliates can be accurately mapped from end-to-end for all relevant data sources (new and backlog data) throughout the organization to the end users and working environments in DELFI. The E2E data workflow diagram (data lineage) demonstrates the complex data flows existing currently in OMV. These data flows exist for fulfilling actual business needs.

The data flow diagram has been split into different WLMS relevant business workflows. Depending on the business need, data may enter only one workflow or data may undergo several workflows in different combinations. The E2E diagram will give OMV the ability to translate the workflows existing in the current IT architecture to workflows in an OSDU™ Data Platform / DELFI ecosystem. The established concept can support an OMV company-wide adoption of OSDU™ Data Platform as a cornerstone and can be directly handed over to IT as implementation concept.

Please refer to Chapter 2.3 Data Conditioning Fundamentals to find the general data conditioning workflow relevant for this data lineage diagram, as well as the definitions and distinction between the meaning of digitizing and digitalization.

The data flow diagram delivers two key value components:

a) Breakdown into clear wellbore centric data workflows

b) Possibility to attribute individual workflows to specific DELFI Solutions / OSDU™ Data Ecosystem

Workflows 1 – 12 in Appendix 2 can be categorized according to the OSDU™ Data Platform Ecosystem / DELFI Solutions that they will be assigned to and are named as follows:

OSDU ™ Data Platform (Data Ecosystem)

Data Lake:

- Workflow 1: Hardcopy Evaluation
- Workflow 6: Digital Data Evaluation

OSDU ™ Ingestion Framework:

- Workflow 2: Scanning
- Workflow 3: Hardcopy Processing
- Workflow 4: Historic Digitized Data QC
- Workflow 5: Digitizing
- Workflow 7: Digital Data Handling
- Workflow 8: Industry Standard Conformity

Data Warehouse: OSDU ™ Core Storage:

- Workflow 9: WLMS Loading

OSDU ™ Consumption Zone:

- Workflow 11: Data Consumption

DELFI Solutions

Petrotechnical Suite / OSDU ™ Ingestion Framework:

- Workflow 10: Composite - / Standard Log Creation
- Workflow 12: Transfer to Working Environments

Please refer to Appendix 2 – E2E Data Flow Diagram for the full version of the data flow diagram.

4.2 Information Architecture

"The way we see it, it is like trying to help a team of mountaineers to reach the top of a mountain safely without knowing which peak they are climbing, and without having a joint base camp to assemble the team for the climb. The consequence is that every team member is building their own base camp at a different location and targeting a different peak to climb, which increases, not decreases, the risk of failure. […] We need to think about where we are now, and where we would like to end up before we start. […]

There needs to be a shared vision on which to base a strategy to become a digital company. It is not enough to add an app to the existing products and throw the results over the wall to Operations to run and maintain them." (Ramsay et al. 2021)

IT architecture at OMV should be focused on the following goals:

- Avoid monolith applications, in which workflows (orchestration layers) and data are linked. Rather focus on establishment of a data platform where data and workflows are detached, and the data platform only takes care of data.
- Workflows (orchestration layers) that incorporate business rules, compliance checks and audit trail.
- Orchestration layers structured around different disciplines, e.g. Well Planning and Delivery discipline with an orchestration layer (workflows) and a data platform facilitating the discipline.
- Avoid linking apps and data platforms via extensions and rather have an IT architecture that facilitates the integration of quality controlled, approved, and unambiguous workflows.

4.3 Components Integration

Upgrading from a legacy system (PS Web/Seabed) to a new data platform or working environment (DELFI/ OSDU™ Data Platform) is a complex task and should be done in a seamless way that will integrate components such as Data Conditioning by HOL while at the same time not interrupting daily business. Data Conditioning by HOL has already been characterized in the Data Conditioning Fundamentals section, Chapter 2.3. The currently existing IT architecture in OMV as of PS Web 2023.1 allows data liberation into DELFI/PTS. PS Web-TL connector operating in the DELFI environment helps OMV transfer standardized data (High Company Value data) from PS into the Techlog environment. Additionally, the PS Web interface will be available in DELFI/ PTS for data browse, search, and download. When migrating towards an IT architecture that embraces OSDU™ Data Platform, HOL component/ functionality integration must be considered.

Two central concepts of getting high company value data into WLMS are Data Gathering (not existing in a systematic way) and Data Conditioning (standalone and not integrated seamlessly in the overall E2E flow). When transitioning to OSDU™ Data Platform, three options are considered when it comes to Data Conditioning, - the first two options have already been presented in Chapter 2.4, Figure 13:

I) **ELT: Extract Load Transform**
Ingestion of data into OSDU™ Data Platform and subsequently Data Conditioning (QC) directly in OSDU™ Data Platform.

II) **ETL: Extract Transform Load (currently done in WLMS solution based on PS Web)**
Data Conditioning (QC) in cooperation with the service provider HOL outside OSDU™ Data Platform and afterwards ingestion of quality-controlled data into OSDU™ Data Platform.

III) **EDS: External Data Source**

The goal is the reduction of manual efforts related to exchanging well logs with HOL, i.e. sending well logs to be conditioned to HOL and receiving High Company Value data from HOL. In addition, it would be great benefit if the functionalities HOL is offering are embedded in the overall E2E data flow as it would offer several benefits:

- It can be avoided that high volume of data are sent back and forwards between OMV, OMV Petrom and external service companies.

- Optimization of data quality control process , keeping OMV quality control efforts to a minimum.
- Tracking of progress and auditing if delivery deadlines can be done in an OMV controlled way, supported by automatization. If promised delivery deadlines are not kept, relevant action points can be set.
- HOL tools and functionalities are available for internal usage within the OMV and OMV Petrom environment. Several people are digitizing data in OMV Petrom. To ensure that defined OMV and OMV Petrom standards are followed and make the internal process more efficient, people would benefit to have the functionalities available in their working environment.
- Possibility to integrate HOL functionalities with other modern tools available in the OSDU™ Data Platform / DELFI environment.
- Possibility to integrate and use data from other domains/ systems to the improvement of the data quality of WLMS data types or complete missing mandatory information.
- Other service contractor than HOL can be used working directly with the E2E workflow setup within the OMV environment to ensure that all data conditioning processes and standards are followed. This ensures that data end up in the system in a consistent manner.

Some questions to ask beforehand and considerations to be made involve:

- Referring to I) ELT Extract Load Transform, to which extent is data conditioning possible inside OSDU™ environment?
    - How much effort (man days and cost) would it take to build in HOL data conditioning functionalities (directly or with new approach like DataIKU workflows) into DELFI/ OSDU™ Data Platform?
    - How long will it take to implement those functionalities considering that OMV did already conditioning and loading of more than 230 000 files with existing tools?
    - How long will it take to be able to do the correction of non-industry standard conform files, semi-automatic data classification, log digitization, formatting of unstructured data and digital transformation directly in an OSDU™ Data Platform user environment?
- The goal is a seamless, fully integrated, transparent workflow: Only one single environment, in which the a) manual effort and b) complexity for data gathering and

data conditioning will be minimized. This way, the number of errors and missing files can be reduced,

- The implementation of a seamless business workflow driven E2E workflow and maintenance in operational phase based on acceptable cost must be the goal to ensure a long-term sustainable environment for an acceptable price. This point is very important, considering the growing LCB which is a different world, regarding cash generation, compared to the oil and gas business.

Four basic, realistic, short-term options on how to integrate the service contractor for data conditioning (currently HOL) into the OSDU™ environment at OMV, OMV Petrom and its affiliates exist:

- **Option 1 - SaaS (Software as a Service i.e. Data as a Service):** Host APPS on the environment of HOL,
- **Option 2 - PaaS (Platform as a Service):** Host APPS on Azure in a Cloud2Cloud setting,
- **Option 3 - EDS (External Data Source):** HOL is treated as an EDS in OSDU™ Data Platform. On request, data can be brought back and forth.
- **Option 4 - Rebuild HOL APPS in DELFI/ OSDU™ Data Platform:** A full integration of HOL functionalities directly in DELFI/ OSDU™ Data Platform.

Option 5 is not listed here, - it should however be considered as well: Option 5 would involve first implementing Option 1 (Software as a Service) and then gradually moving to Option 2 (Platform as a Service). Option 1 - SaaS (Software as a Service) is represented in Figure 31, in which case data is transferred from OMV to HOL (data to be conditioned), being read and written in the SaaS environment of HOL. The respective advantages and disadvantages of Option 1 are listed in Table 11.

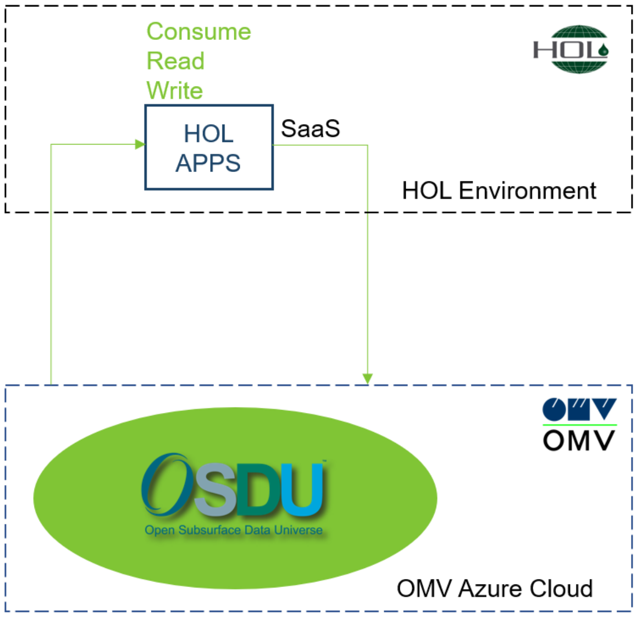


Figure 31: *Option 1 - SaaS (Software as a Service)**:** The HOL APPS are hosted in a software as a service environment of HOL. This environment communicates with OSDU™ Data Platform via Azure Cloud.*

Option 2 - PaaS (Platform as a Service) would involve a Cloud-to-Cloud setting, in which consumption, reading and writing of data would occur in Azure Cloud of HOL, functioning as a platform as a service and communicating with Azure Cloud of OMV. Figure 32 shows Option 2.

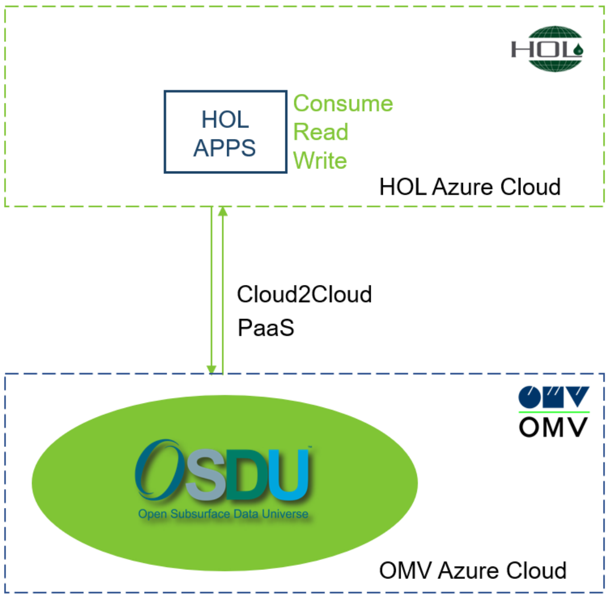


Figure 32: *Option 2 - PaaS (Platform as a Service)**:** Cloud2Cloud involves two Azure Cloud systems communicating with each other.*

Finally, Option 3 - EDS (External Data Source), involves two OSDU™ systems interacting with each other. HOL is treated as an OSDU™ – external data source. Both are running an OSDU™ instance. Figure 33 shows option 3.

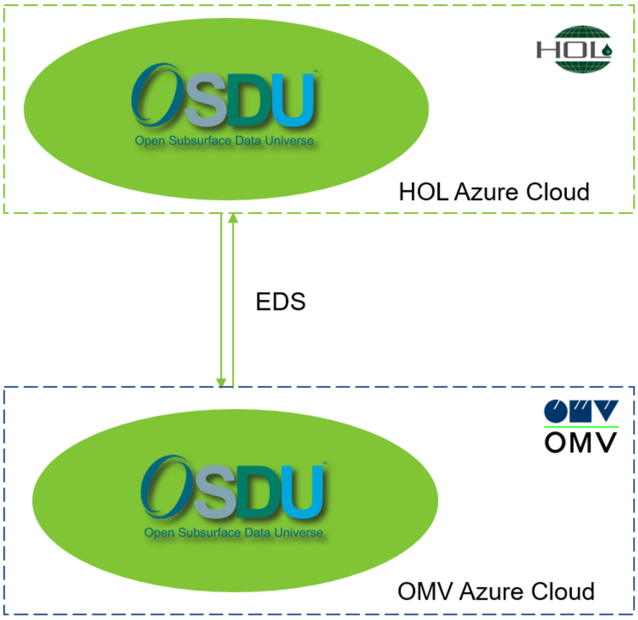


Figure 33: *Option 3 - EDS (External Data Source): Both OMV as well as HOL would require Azure Cloud to run OSDU™ environments individually, which would communicate with each other.*

Table 11 shows a comparison between advantages and disadvantages of each option (Option 1 – Option 4).

Option	Advantages	Limitations
Option 1 - SaaS	+ Easy implementation	- Dependency on Service Contractor - OMV and OMV Petrom can use the tools and functionalities only by connecting into the HOL environment. - Big volumes of data have to be send back and forwards between OMV, OMV Petrom and HOL
Option 2 - PaaS	+ Independence from Service Contractor + Platform can be easily used by OMV and OMV Petrom personnel	- Requires clear idea of how to deploy updates, version management, … - not integrated in overall E2E workflow
Option 3 - EDS	+ Already proven to work in a joint venture between 3 E&P companies + Most efficient way to exchange data between companies using OSDU™ Data Platform	- Increase of cost for data conditioning to be expected - On-demand requests required every time to transfer data back & forth
Option 4 - Rebuild HOL APPS in DELFI/ OSDU™ Data Platform	+ Fully integrated in E2E workflow + Workflows access control easy (for internal & external user) + External service contractor work in OMV environment + No data transfers in and out from OMV environment + Always up to date dictionaries + Integrability with other domains + DataIku functionalities (AI, …) + Independent from HOL	- Cost - Repeat all work was done until now - Extensive work to set up all workflows in OMV environment - Limitation of existing software tools to execute certain data condition workflows

Table 11: *Four options on how to achieve service contractor's components integration into OSDU™ Data Platform and their respective advantages and disadvantages are compared.*

For OMV, OSDU™ Data Platform can be deployed via Microsoft Azure. The feasibility of establishing MS Azure at HOL or any other service contractor is uncertain and must be evaluated beforehand.

Currently, for OMV the most feasible option out of all four, considering a OMV and OMV Petrom transition to OSDU, would be Option 3, in which HOL or another service contractor would be treated as an EDS (External Data Source). However, beforehand the feasibility must be thoroughly analyzed in an evaluation of feasibility for establishing a service contractor as an OSDU™ EDS.

The impact of cost is a huge factor to consider, as currently the cost of data conditioning by HOL is optimized. Establishing an OSDU™ instance both for HOL as well as for OMV would involve considerable costs.

4.4 Evaluation of OSDU™ Data Platform - HOL Compatibility

The following criteria are important when transitioning to a novel IT architecture (Option 3 – EDS (External Data Source) in regard to integration of HOL APPS and components:

- Creation of a clear list of APPs that are currently being used by HOL. This creates a meaningful link between the E2E data flow diagram (the APPs involved), which was presented in chapter 4.1 and link it with the future IT architecture. That way, the applications ("APPS") from the service contractor HOL can be seamlessly integrated in the overall information architecture, be it with or without OSDU™ Data Platform. A clear list of APPs which are currently involved in data conditioning and scanning has been crafted based on the following aspects:

a) **Software** used by each APP,

b) **Programming language** used for each APP. It is hard to migrate code written in programming languages such as Fortran or COBOL to the cloud,

c) **Containerization potential** of each APP (dockers): Can certain APPS be run in an isolated scenario?

d) **Filesystem interaction** of APPS: How do APPS interact with filesystems? APPS should be able to interact with cloud-specific mechanisms,

e) **Communication patterns (Industry standards)** characteristic of APPS. Examples of such communication patterns are: REST - API (text based) or gRPC (binary),

f) **Functionality** of each APP including the current functionality, as well as the future vision of that Apps' functionality,

g) **In – and output** of each APP can be listed as pairs according to the six sigma tool SIPOC. Inputs may be: other APPs, persons, suppliers, … . Outputs may be: Customers, …

h) **Name** of each APP,

i) **Status** of each APP (under development / finished). For example, the HOL APP "DLIS Writer" is still under development.

The document containing the latest information on the aspects a) - i) of HOL apps is available on the EDSA-1 OMV SharePoint ("HOL_App_List_V2").

Figure 34 shows a possible approach on how to move within OMV toward OSDU™ Data Platform readiness, with the respective deliverables listed for each department. It is often not easy to separate the deliverables of each department clearly, so the aim of Figure 34 is to assign deliverables to each department, which when fulfilled move OMV one step closer toward OSDU™ Data Platform readiness.

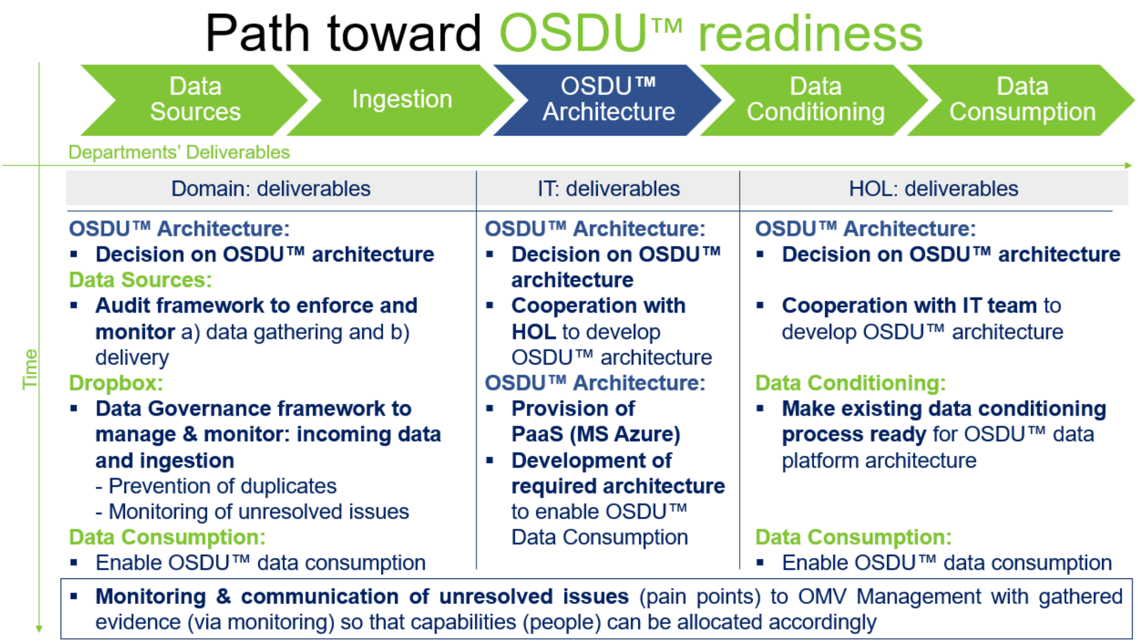


Figure 34: *Deliverables to achieve OSDU™ Data Platform readiness. Deliverables are split between three departments: Domain, IT and HOL as well as categorized by workflow elements from data source to data consumption.*

Chapter 5 Economic Evaluation

To complement the previous chapters, this chapter focuses on the economic impact.

Why Well Log Management matters

Modern Energy companies find themselves competing in a Digital Economy, which requires proper data governance in order to discover and measure the value of data and data management to keep the quality of data high with the goal of maximizing operational efficiency and reducing the expenditure of time. A metric such as the Return of Data can help track the value of data as a strategic corporate asset which moves vertically through Energy companies through all four technology stacks.

Digital Economy

To take advantage of the benefits that AI and other advanced technologies have to offer, organizations must have a consistent, thorough way of managing data, especially subsurface data such as well log, core and mudlog data. Historic hardcopy data must be digitized, and historic digital data must be integrated, while volumes of new incoming digital data from branch offices / contractors are growing. Purchased data acquired by OMV, OMV Petrom and its affiliates must be integrated. Qualified experts need to be involved in the data take over and acquisition process to optimize the data flow and have data as fast as possible available user environments and ready to be used by experts to make efficient decisions, these volumes of historic as well as new incoming data need to be properly managed. Direct loading of data to end user environments will not promote a digital economy, but rather deteriorate quality of data and trust in data.

Data Governance to power data discovery, - measure and – valuation

There are extensive regulations to govern financial data, but not in Energy data. Still, errors in the data within the Energy domain can lead to errors in decisions with potentially enormous impact. Part of data discovery is analyzing the data in a corporation and looking at how it is being used. Is it more geologic-oriented, engineering-oriented, or finance-oriented data? What is the critical data required for the decisions in the company? Tracking and measuring is essential as well: If one does not measure the quality of data as well as how much and if data assets are actually used, how shall one know if the data management initiative has been effective?

Operational Efficiency and the Return on Data

"If you don't invest in data - and information management, the cost of poor decisions i.e., poor investments is going to dwarf the cost of taking care of data in an ongoing manner." (Jim Crompton, Tim Coburn, John Jacobs),

Having dozens of people which have to go through the data over and over again would be an example of operational inefficiency. Data should be regarded as a strategic asset, because a lot of money was paid beforehand to acquire the data. Data management adds to the operational efficiency by structuring that data. The focus should be on serving the data and making it fit for purpose (consumption) before the data is needed by the end user. Only delivering data on an ad-hoc basis when it is needed is not sufficient.

- Data is a strategic asset spanning across an entire organization and should be perceived as such.
- While in the financial domain, the notion of Return On Investment (ROI) is well known and acknowledged, it might be wise to consider using a similar notion in the data world.
- Return on Data: It is not the company with the most data volumes at its hand that makes the best decision, but rather the company with the most accurate, complete, and current data.
- Unfortunately, most E&P companies do not measure the value of data by how much it returns (Return on Data), but rather only by how much it cost to acquire it.
- The Return on Data is a metric that can be measured within the data management initiative, e.g.: the number of users n actively using a certain dataset over a defined timespan t, leading to x amount of successfully finished projects as a result. (Jim Crompton, Tim Coburn, John Jacobs)

The Expenditure of Time

A Geoscientist has spent significant time looking for a particular data set she needs to finish her interpretational work. In an interview, she has stated that 80% of her time she is gathering, formatting, and preparing data. Only 20% of her time she can focus on her actual job of seismic interpretation. (CDA)

When looking at a full-time position in a company and multiplying the salary of that position times 0.8 times the amount of similar positions times the years in the company, one can get an idea on the expenditure of time as well as cost of not having the ready-to-use data in the leading corporate system, greatly reducing operational efficiency. The expenditure of time relates to two of the Four Technology Stacks (Data and Integration): The time required to find the good data, as well as to integrate this data into the leading corporate systems.

Business – Data – Technology: Debunking the Technology Myth

The technology myth is described by Prof. Crompton in his podcast, where he explains the pitfall of many organizations: They first invest in expensive technology in the hope of magically fixing the underlying data quality issues that exist within the organization. "*Nowadays there is the belief that in a way, tech can help us with all the problems* [we have in data]*, we can simply skip all the hard work. All you have to do is take the* [technology application] *license and the technology will fix it. But that is simply not true. Afterwards we still face the old data quality issues we had in the first place and the salesman comes back and you can go from version 3.2 to 5.3 and there will always be an amazing new product. However, it will not resolve the data quality issues existing at the root."* (Jim Crompton, Tim Coburn, John Jacobs)

WLMS program has not fallen into the trap of to the technology myth. Please note that WLMS was set up as a holistic implementation taking most critical components into consideration to make it functioning in the organization. This includes many perspectives:

- WLMS related business processes
- WLMS related business regulations (data delivery requirements, standards, etc)
- Systems design to support business defined processes and defined minimum standards
- Software development to support data condition in an efficient way, but still fulfilling business needs.
- Continuous data quality improvements
- Defining data standards and minimum quality criteria (e.g. High Company Value Data)
- Data conditioning and loading initiatives to get business critical data into the system (singly data source!)
- Execution of audits to ensure that regulation is followed (could be improved as it is a lot of manual efforts)
- All new acquired data must be loaded into the system (operated wells)
- Actual/ future business needs
- Historical data handling mistakes made,
- Global standardized approach for WLMS data types,
- Globally auditable process,

- System limitation of available products on the market (align business needs with most cost-efficient system implementation without sacrificing the "must haves!" from business side),
- Ensure that WLMS is at one point of time the single data source for WLMS related data types.
- Etc.

Key achievements and pain points of WLMS are listed in Chapter 2.2 Data Foundations by the Example of WLMS.

First things first – Three Central Questions to answer how a Data Management Initiative can serve the business needs, - for example of OMV:

1. What will OMV's business look like at different time horizons (1 / 5 / 10 years)?

- OMV will continue to explore for oil, gas and LCB (Low Carbon Business) in the future,
- Volumes of Data will continue to grow (e.g.: 40 vs 80 000 channel recordings in seismic),
- Need to enhance value creation by leveraging data assets across the company.

2.What data is needed to achieve OMV's business goals?

- High Company Value well log -, mudlog - and core data available in working environments or easy to transfer into relevant environments with minimum efforts and fast way (Data only a click away!).
- Trustful source of all RAW and other relevant PROCESSED/ INTERPRETED data must be accessible in an efficient way.

3.What changes to the technology landscape need to happen at OMV?

- A good data foundation is needed for scaling digitalization initiatives at pace to derive the business value detailed in Chapter 5.1 Business Value Proposition.

Data – Integration – Insight – Decisions across all Four Technology Stacks

As introduced in the first chapters, data flows vertically through all levels of Energy companies during field development, from exploration & appraisal to production. Figure 35 displays the four technology stacks through which data flows typically in upstream business of oil and gas companies.

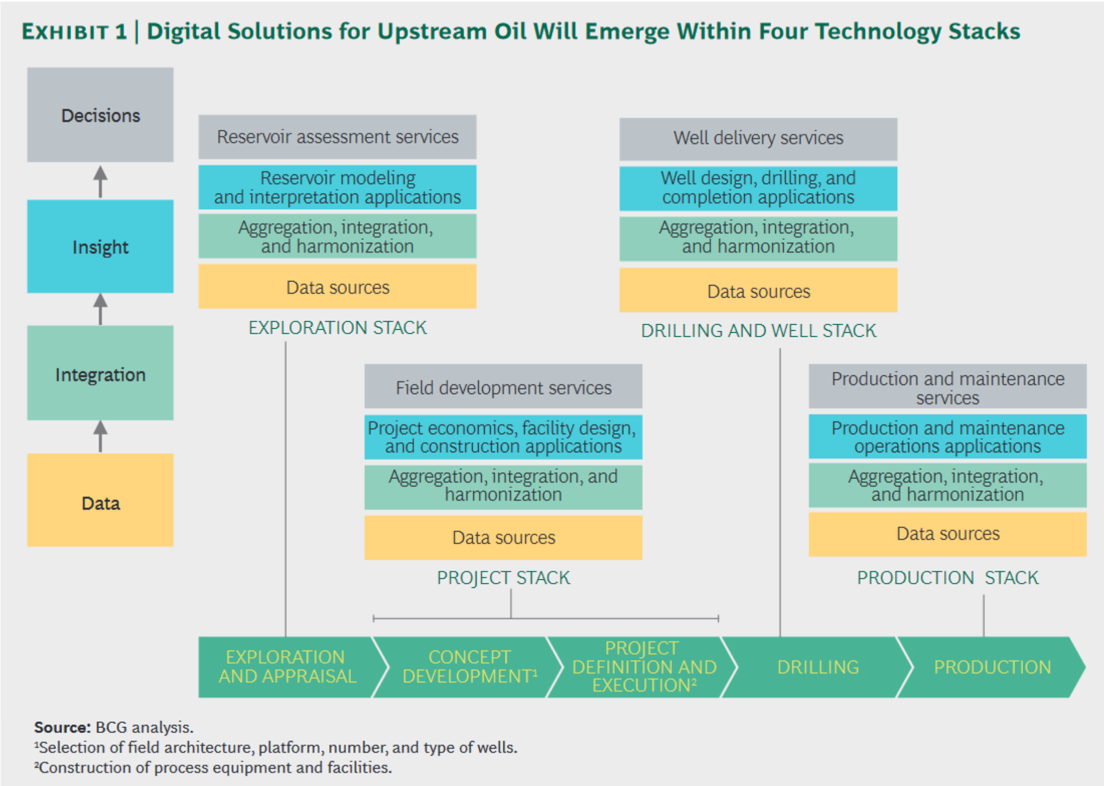


Figure 35: *The Four Technology Stacks (Exploration -, Project -, Drilling and Well -, Production Stack) revolve around data, integration, insight, and decisions. Data flows across these four technology stacks and each stack relies heavily on a standardized way of getting the correct, complete data on time from data sources. (Sylvain Santamarta, Rohit Singh, Peter Forbes 2017)*

Having a common portal for ingesting data from disparate data sources through all stages of field development would save a lot of time and money. If each stack has 10 data sources, times 4 equals 40 data sources. When there is a way to semi-automatically aggregate, integrate and harmonize these data sources inside OSDU^{TM} Data Platform and WLMS for each data source, and we consider the time that this task would require to be 20 hours per data source, we can come up with a number amounting to 800 hours times e.g., a 60 € gross hourly salary for a full-time employee position, amounting to 48 000 €. If this is multiplied by 3 field development projects per year, a total of 144 000 € is spent on getting data from sources and integrating the data in the leading corporate system in one year. Such operational inefficiencies can be reduced by redistributing budget toward financing the company-wide wellbore centric data platform WLMS, or the combination between WLMS and OSDU™ Data Platform.

When it comes to establishing OSDU™ Data Platform in an organization, according to Prof. Crompton, out of the four elements (data → integration → insight → decisions), integration is the element that is the most challenging to achieve. He described the integration of OSDU™ Data Platform as the holy grail. The technology myth that integration with OSDU™ Data Platform will work magically without standards leads to inflated expectations and ultimately creates disappointment.

There are three ways regarding how OMV can manage its' data and in a narrower sense how WLMS can handle data:

1. **Data as an asset:** Understand which is the key data, where to find it, data quality, standardization, and governance.
2. **Data as a product:** How will data consumers see the data: Portal, accessibility – how to access the good data quickly.
3. **Data as a service:** How much data management capability does OMV want to retain internally? To what extent does OMV want to outsource data management? Does OMV want to fully rely on technology platforms such as DELFI / OSDU™ Data Platform (by SLB).

Combinations of 1. ,2. and 3. are possible.

In any case, better data in OMV will lead to better decisions across all four technology stacks.

OSDU™ Data Platform: First Movers'(Dis-)advantage and Internal network effect

Having the First Movers' advantage will give OMV an:

+ Advantage of managing its corporate data assets most efficiently and potentially allowing it to make better decisions, creating new opportunities.

+ Internal network effect: Establishing a working use case of OSDU™ Data Platform in OMV within the scope of well log management would have the advantage of being able to replicate that experience and leverage it to help other departments in OMV obtain OSDU™ Data Platform.

Implications of OMV being the first mover involve:

- High initial upfront investments in software engineers, software licenses, data governance, …
- Risk of OMV overextending itself beyond its capabilities. It might be strategically wiser to

wait for other Energy companies with the same size and digital capabilities as OMV to adopt OSDU™ Data Platform.

5.1 Business Value Proposition

In their study on "Digital Powers Value Creation in Oil and Gas" in 2020 BCG found the following insights: *"The upstream oil and gas (O&G) industry is in the midst of a value-creation crisis […]. The demand collapse following the pandemic outbreak is both enhancing the urgency and acting as a catalyst for change. To survive, O&G companies need to transform their legacy business to achieve a step change in performance. And they must innovate their business models to tap new value pools. Leveraging digital technologies is critical on both of these fronts."* (Håvard Holmås, Sverre Lindseth, Sylvain Santamarta, Marie-Helene Ben Samoun, Mauro Castilhos 2020)

"Given that most O&G companies are announcing capex cuts of 10% to 30% and opex cuts of 5% to 20%, the need to use digital to power fundamental transformation is greater than ever. The study generated three major insights: upstream O&G companies are struggling to deliver value from digital; digital maturity correlates with value delivery; and digital's importance will grow during the crisis." (Håvard Holmås, Sverre Lindseth, Sylvain Santamarta, Marie-Helene Ben Samoun, Mauro Castilhos 2020)

Figure 36 shows the potential behind unlocking the value from E&P companies' digital assets. It illustrates that up to 70% of engineering hours can be reduced in the field development technology stack, with a symbiosis between human elements (engineers, internal and external stakeholders) and technology elements (smart, integrated models and digital twin technology). In Exploration technology stack, more than 25% increase in discovered barrels is possible.

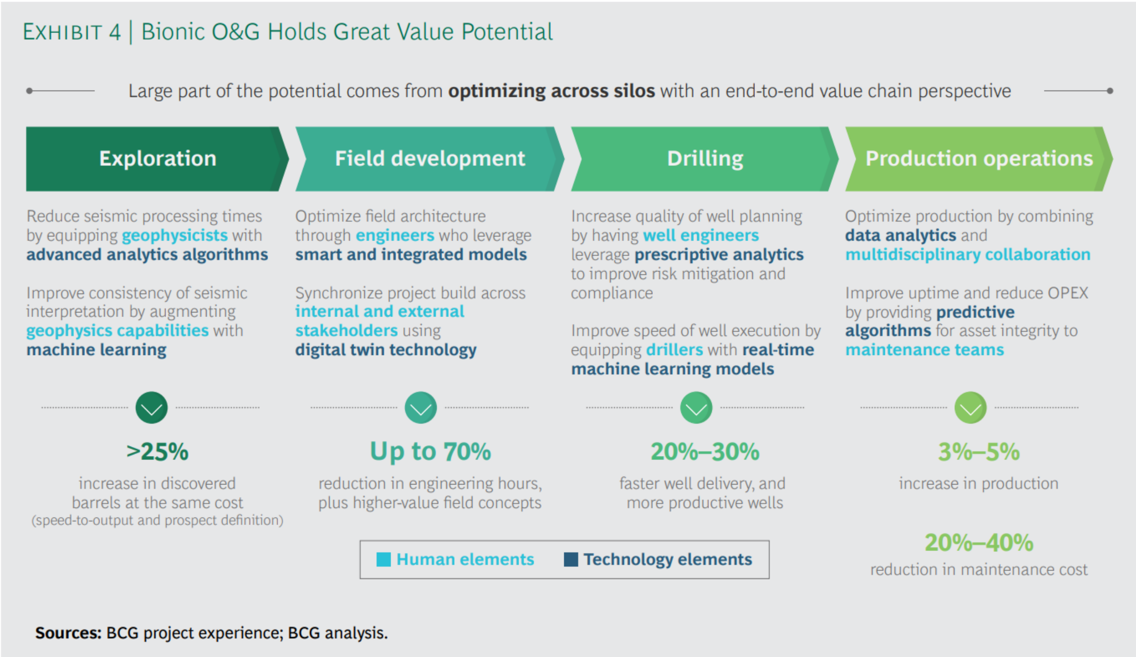


Figure 36: *Unfolding the value of E&P data holds a great potential from exploration to production operations. The potential for savings is depicted at the bottom. Human elements and technology elements complement each other, and their symbiosis brings value potential. (Håvard Holmås, Sverre Lindseth, Sylvain Santamarta, Marie-Helene Ben Samoun, Mauro Castilhos 2020)*

The potential savings and time reduction values presented in Figure 36 above may not be completely applicable to OMV. However, they give a good idea of approximate numbers that can be achieved if OMV combines the human and technology elements presented above.

Figure 37 depicts the four technology stacks (Exploration, Field development, Drilling and Production operations) and shows how WLMS data types are involved in each stack.

- The figure also shows how profit (Profit = Revenue – Cost) is generated.
- Revenue is generated via production and selling of oil (volume of barrel oil equivalent), gas (quantity of kWh) and geothermal (quantity of kWh).
- Cost is composed of OPEX (Operational Expenditures) and CAPEX (Capital Expenditures).
- WLMS is involved in all the three technology stacks (Exploration, Field Development, Drilling) that are necessary to achieve economic production.
- WLMS also supports production by being the collaboration basis for workover -, production - , casing integrity - and timelapse logs.

- Figure 37 also shows that WLMS actively contributes toward OPEX reduction throughout all four technology stacks by increasing efficiency of Full Time Employees (FTEs) leading to less required FTEs for the same workload.

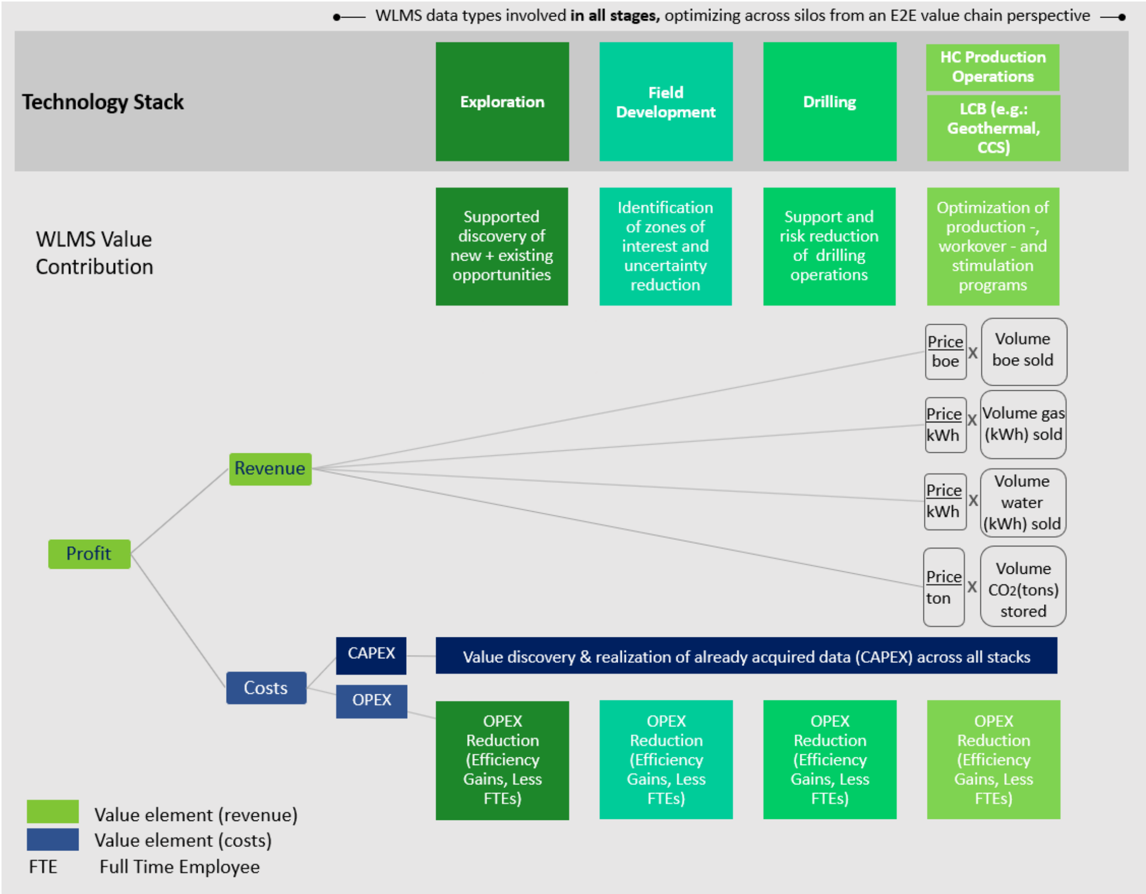


Figure 37: *WLMS data types are involved in all stages of Energy. The top of the graph shows all four technology stacks from exploration to production and below the value derived from WLMS. The split of profit into revenue and costs (which is split into OPEX and CAPEX) shows that WLMS contributes toward increasing revenue and decreasing Costs (OPEX). The cost of acquiring data during well logging campaigns (CAPEX) is leveraged by WLMS. WLMS is therefore massively helping to increase profits in a sustainable way.*

Outlook: AI Use Case - WLMS as a lever for transforming OMV into an AI-powered company

As illustrated in Figure 36, digital transformation holds a huge potential to leverage machine learning and advanced analytics algorithms. However, in order to make use of this advanced technology, the data assets must be first unified from disparate data sources, aggregated, harmonized and integrated. The high variety and volume, as well as the different grades of digitization of data in E&P pose a significant challenge for directly applying advanced AI technologies on data without first preparing the data to make it fit for purpose. These AI technologies require data to be ingested in a uniform, standardized way. WLMS bridges this gap and can act as a lever for transforming OMV into an AI-powered company. **Analytical AI** (artificial intelligence) can be split into two major branches: (A DATE WITH DATA)

- **Predictive AI** for assessing the future by demand Forecasting, predictive maintenance, …,
- **Generative AI** (Large Language Models LLM) for creating content, transforming content.

Figure 38 gives an example of how generative AI could be combined with WLMS in order to allow users to query for information they need for their projects.

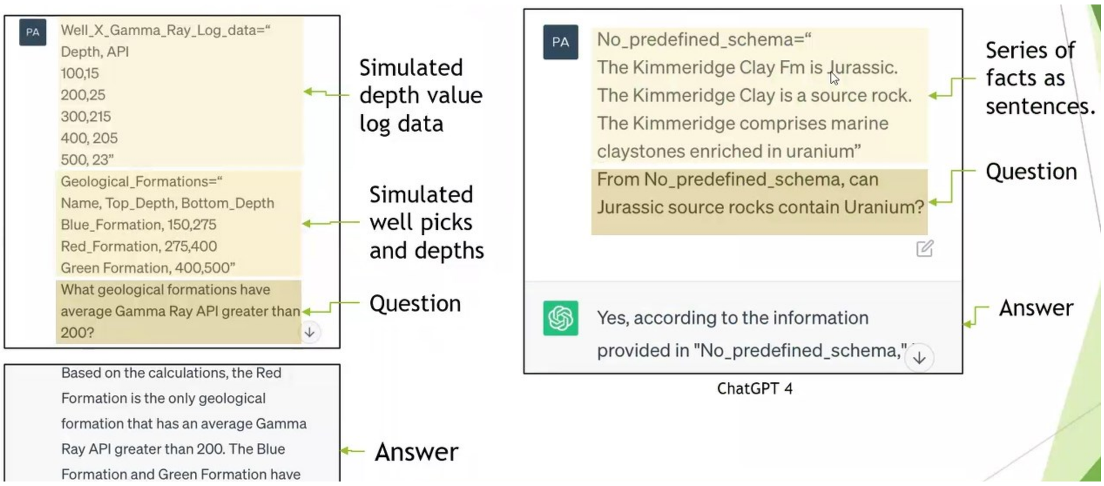


Figure 38: *Querying for information in the subsurface domain is being changed by AI. Generative AI holds a high potential for new ways to interrogate and query for information. On the top left, simulated depth values and gamma ray API values are ingested into the large language model (LLM). Geological formations and simulated well picks are ingested to the model. Based on this information, a question is asked: "What geological formation have average Gamma Ray API greater than 200?" and the model provides the correct answer on the bottom left side of the graph. A similar case is presented on the right side of the graph, where the model is asked whether Jurassic source rocks can contain Uranium.* (Cleverly)

Chapter 6 Conclusion

Conclusion

6.1 Summary

The fundamentals on well logging have been described in a single graph in a very clear form. Well logging data formats have been concisely summarized and compared in a way that is not yet existing in the scientific E&P community. WLMS as a data foundation is presented. The way how data conditioning helps create high company value data is explained. Fundamentals on DELFI and OSDU^{TM} Data Platform, as well as their relevance in the digital economy are emphasized on.

An approach on how to establish a single point of data entry has been laid out. A new proposal for standardized data gathering and audit for timely and complete data delivery via automatized emails, management involvement and data governance has been presented. An E2E data flow diagram for holistic data lineage has been included, with all relevant HOL Apps as well as workflows. Roles of Wellbore Centric Data Integration Group have been laid out and defined according to business needs to achieve the E2E workflows and support those during operations. A concept on data tracking and monitoring has been developed. An information architecture matrix has been drawn out, including ways to incorporate and integrate different components.

A business value proposition has been made. An outlook on how WLMS can drive OMV forward in becoming an AI-powered company has been presented.

In the section below, a series of recommendations has been given on how to implement the solutions in this thesis as well as a proposed action plan on a possible way forward.

6.2 Recommendations

The key recommendations for OMV are split up, referring to four recommendation categories. The elements 1. – 6. which are required to navigate the digital roadmap presented in Figure 39 are addressed by the recommendations and four recommendation categories A) – D):

A) Key recommendations for a timely, efficient gathering and delivery of data into the leading corporate system of WLMS, acting as a trusted, single source of truth for high company value data to end users. (I – VI)

B) Key recommendations for E2E IT architecture, Components Integration (VII – X)

C) Key recommendations on OSDU™ Data Platform (XI – XVI)

D) Key recommendations on Business Value proposition (XVII – XX)

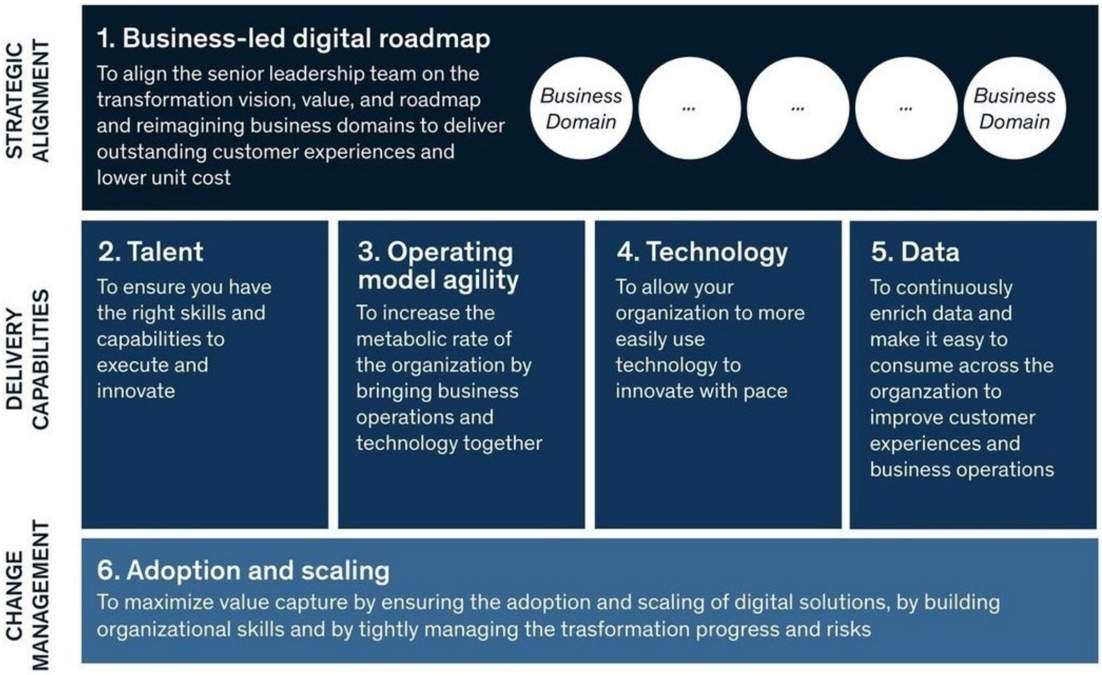


Figure 39: *For OMV to navigate the AI and digital roadmap requires different elements to work together, namely strategic alignment, delivery capabilities and change management. (Smaje 2023)*

Strategic Alignment (1. Business-led digital roadmap) is addressed by recommendation B) and D),

Delivery Capabilities (2.Talent) are addressed by recommendation III),

Delivery Capabilities (3. Agility) are addressed by recommendation VIII),

Delivery Capabilities (4. Technology) are addressed by recommendation categories B), C),

Delivery Capabilities (5. Data) are addressed by recommendation category A),

Change Management (6. Adoption and scaling) are addressed by recommendation categories B), C).

A) Key recommendations for a timely, efficient gathering and delivery of data into the leading corporate system of WLMS, acting as a trusted, single source of truth for high company value data to end users. (I – VI)

I) Realize that every company is different in its company-specific goals, challenges, history and therefore every company has a different information management strategy and different business-led digital roadmap:

Just because an approach is feasible and relevant for Company X does not mean that it is also feasible for Company Y. OMV must take its own path in terms of wellbore centric data management, specific to its own characteristics and capabilities.

- **Consider creating an Information Management Maturity - / Data Governance Maturity Assessment in OMV.** This is a form of tool or methodology that quantifies the maturity of the information management and data governance initiative in the organization.
- **Data Quality assessment** would focus on aspects such as Data Quality, comparing where OMV is right now (as-is) in terms of Data Quality and where the company would like to be (target). (Sykora 2017)
- **Data Governance assessment** would an evaluation of OMV's status in terms of Data Governance maturity levels (level 1: Initial, level 2: Repeatable, level 3: Defined, level 4: Managed, level 5: Optimized). (Sykora 2017)
- Justification of certain decisions or promotion of ideas can be made based on those assessments.
- Consider that Data Culture and Governance in OMV are different from other companies, who might invest in OSDU™ Data Platform because the platform is the right fit for them.

II) Proactively prepare-before-demand all final, ready-to-use data, instead of sending individual datasets via deliver-on-request basis for projects:

All data must exist in the leading system of WLMS, where it serves projects as single source of truth. It is more efficient to have the data proactively prepared before demand rather than having to send individual datasets on a single request basis. The reason is that it takes time to condition data and bring it into the system, especially High Company Value Data. Certain data conditioning workflows need specific domain specialist with deep understanding of the data itself.

III) **Delivery Capabilities: Wellbore Centric Data Integration Group will be required to fill the resource gap for E2E workflows.** Wellbore Centric Data Integration Group should exist of a key team with clearly defined responsibilities and skills, including programming skills and data science skills. It was seen during the last four years, since WLMS Go Live, that branch offices are not necessarily following WLMS processes (OGMS) and relevant regulations (REAL) due to resource, budget and required domain knowledge limitations. A Wellbore Centric Data Integration group will be able to support BOs and close the resource gap.

- **Which skills are required in the BO** and how is the interaction between HO and BO? How will the charging of services to BOs work? For example, charging services to BOs by distributing the costs of the HO team based on a key to relevant BOs needs to be considered.

IV) **Deploy a) an audit system that makes use of the b) IT system to operate based on automated, event-driven email notifications, involving c) management:** A combination between audit with a system (IT) and management can help lead to an improved data delivery process. Implementation of data completeness checks based on DELFI Drillplan, as well as coupling the data delivery audit system with the roles defined in Drillplan. 80 - 20 Rule: 80 % of the gathering, grouping, and structuring of the data according to the Dataset Concept shall and can be done with the greatest impact when the data is generated, and the context is still known. This shall be done by a service contractor (Service Contractor Management), or it could also be done by a Data Steward in the respective Branch Office for new - and backlog data. The further downstream (the later in time) the gathering and grouping of the data, the more context is lost. 20% of the gathering and structuring can be done by the Domain Data Owner, a person with high functional expertise who defines how data should be handled and QC'd in the organization and who makes sure in the end that all data is there and complete but does not always necessarily know the exact context of data generated. The Data Owner is the person who has to ensure that the WLMS regulation is executed and that data go in the leading system based on definitions.

- **Consider including OMV personal specific goals in the yearly Management by Objectives (MBO) goals.** Clear, measurable goals with enough weighting ensuring that the regulation is followed.

V) **Put data delivery requirements into the SLAs for Service Contractors:** Where as soon as the subsurface specialist (well log, core and mudlog data) cooperation with the Service Contractor has led to the successful upload of high company value datasets into the leading corporate system of WLMS, the money is transferred to the Service Contractor. The money is withheld and only transferred once the Service Contractor has fulfilled his obligation for data delivery in the leading corporate system. This legal requirement implemented internally in OMV processes should enforce timely, complete data delivery.

- **OMV and OMV Petrom personnel must be very diligent in QC during data acquisition or final data delivery.** The service contractor must be actively made aware of problems and pushed to strictly follow rules. If in addition the jobs are paid without quality checking, the service will get worse and worse.
- **Automatize quality control more and more.** Automatization requires better integration of workflows and processes and more strict standards.
- **Consider creating a KPI to track and monitor the performance of service contractors:** *"If the Data owner fails to provide the data within the stipulated timeframe, days calculated from the reference date (for this data class, reference date will be the well release date) to the received date […], then it will flag as SLA violation by the Data Owner as mentioned in the KPI"* (Diwakar und Akoum 2019)
- **Data quality control and follow up is time extensive** and OMV must have enough resources and assign properly the resources

VI) **Consider creating data catalogs to understand which data exists where and try to first migrate data to WLMS, then delete old data and legacy data storage systems like O-drive:** If data is still available in other systems / locations, people will not have an incentive to work with the data from WLMS. If however WLMS is the only available, single source of truth, the amount of active users will also increase. Complementary to recommendation II), this way efficiency when working on projects can be increased.

- **Get all business-critical backlog data into the system** (first WLMS (ProSource); OSDU™ Data Platform: data lake and data warehouse) as fast as possible to establish a single, trustful data source and remove relevant data from all other locations (keep back up in location with restricted access.)

B) Key recommendations for E2E IT architecture, Components Integration (VII – X)

VII) The E2E workflow is a key facilitator that the majority of data quality control, classification and loading into WLMS is done directly after data enters the company following strict data flow processes (e.g.: Branch Offices when the data is acquired during well drilling campaigns or when data is purchased). Correct data handling is essential directly after data has been acquired or was handed over, because it is still possible have the greatest impact on correcting the data or asking for missing data.

VIII) Create strategic alignment by promoting the E2E data workflows from the data lineage diagram and incorporate the workflows further into the business. Make sure people understand and follow these workflows. A business-led strategic roadmap would involve: Firstly, business workflows. Secondly, data must be served according to business workflows. Thirdly, technology should serve data to the consumers.

- **OMV needs to establish business workflow thinking** and give personnel an understanding what is done in these workflows. Business workflow thinking will help power the digital transformation in OMV, acting as a key-enabler of elements such as automatized service contractor QC or artificial intelligence.
- **Enforce end users to work according to the workflows defined.**
- **Use the E2E workflow diagram to raise awareness** of it what is behind WLMS (not just a replacement of a storage system!) and use it to help people keep the business objectives in mind and that those cannot be neglected.

IX) The criteria that are currently relevant for the existing IT architecture will most likely be relevant for the future IT architecture as well:

- Data conditioning, data preparation activities must be kept at a low price,
- Systems for data storage must be cost efficient also considering big data,
- Applications to use and analyze data must be cost efficient,
- Get maximum out of data with minimum resources (people), …

X) Regarding WLMS, perform Business Capability mapping as a comparison of the state of WLMS as it is today vs. how WLMS should be in the future

C) Key recommendations on OSDU™ Data Platform (XI – XVI):

XI) OMV may risk overextending itself beyond its capabilities with a full adoption of OSDU™ Data Platform:

- **High initial upfront investments as well as ongoing cost** in software engineers, software licenses, data governance for a full OSDU™ Data Platform adoption are required,
  - **It might be strategically wiser to wait** for other Energy companies with the same size and digital capabilities as OMV to adopt OSDU™ Data Platform.
  - **TotalEnergies, a major E&P company adopted OSDU™ Data Platform for well delivery, however, this required a lot of resources:** *"Using OSDU™* [Data Platform], *well data was ingested and connected to Sismage to evaluate well drilling scenarios. The drilling data was shared directly from the rig, ingested where interpretation and follow ups were completed using Geolog."* (OMV-SLB Tech Day 2023) However, this solution required a lot of software to be engineered around OSDU™ Data Platform, with a number of more than 200 full time employees involved.

XII) Considering that OSDU™ Data Platform is still in the initial development phase, it is important that next transition steps towards OSDU™ Data Platform are planned with caution under consideration of costs and resources (long-term sustainable and cost-efficient solution). Bringing WLMS workflows into OSDU™ - /DELFI environment will enable seamless integrated workflows and support further integration of different functions across the company. Knowhow on the OSDU™ Data Platform can be shared across departments within OMV and facilitate a companywide adoption of OSDU™ Data Platform. As data have been acquired over around 100 years, it is clear that there is a big data variability. This variability raises the need of several different data conditioning workflows. Many functionalities are not considered in an integrated way and or do not exist on the marked. How to embed the needed functionalities and tools into a seamless E2E workflow. Most likely it will not be possible to execute all identified workflows in a fully automatized manner due to data complexity - and variability.

XIII) **Get a better understanding of OSDU™ Data Platform, participate in OSDU™ Forum and close follow up.** Direct contact with OSDU™ Forum members to exchange experience and knowledge as there are many things unclear and not clearly defined. Maybe establish user groups focused on OSDU™ Forum, for example within the Wellbore Centric Data Integration Group.

- There are still many elements around OSDU™ Data Platform that are not fully understood – it is a complex topic and requires time and effort for training personnel.

XIV) **OSDU™ Data Platform should not be enforced no matter what. Rather focus on the main functionalities/ advantages that make OSDU™ Data Platform interesting and find ways to achieve these via non-OSDU™ solutions or adopt a slimmed-down version of OSDU™ Data Platform instead of full adoption.**

- **Main OSDU™ functionalities include:**
  Interoperability: Hassle-free data sharing,
  Openness: Adaption of open data standards,
  Scalability: Compatibility with the cloud,
  Extensibility: A data foundation that can be extended and adapted to business rules in a sustainable way.
- **Focus on repurposing these main OSDU™ functionalities, integrating these functionalities into multidisciplinary workflows** and making sure that OMV personnel works according to these workflows.
- **Integrating wellbore centric data from disparate sources in variable industry data formats into OSDU™ Data Platform is the most challenging aspect.**
  Consider that data integration is already working very well with WLMS system.

XV) **If OMV decides to fully adopt OSDU™ Data Platform, consider the necessary steps toward OSDU™ Readiness from each department (IT, Wellbore Centric Data Integration Group, HOL) to build a Minimum Viable Product.** Consider that OSDU™ Data Platform requires ongoing maintenance efforts. Also review the considerations regarding how OSDU™ Data Platform handles data duplicates, the process of ingesting data into OSDU™ Data Platform, loading core data, the supported well log data types as well as relevant suggested software. Integration is the most challenging aspect of establishing OSDU™ Data Platform in OMV and any other E&P company.

- Review problems and proposed solutions for common questions regarding OSDU™ Data Platform (Chapter 3.1 New and Backlog Data).
- Review proposed actions to prepare functionality of core data ingestion into OSDU™ Data Platform at OMV (Chapter 3.1 New and Backlog Data).

XVI) **Regarding Components Integration into OSDU™ Data Platform, Option 3 – treating the service contractor for data conditioning (e.g. HOL) as an EDS (External Data Source) would be the most sustainable option** according to the IT Architect of OMV. However, it must be evaluated to which extent the service contractor can adopt OSDU™ Data Platform in their environment, as well as the cost involved in establishing an external data source in OSDU™ Data Platform. Consider that EDS does not imply that a fully-fledged OSDU™ Data Platform is required for the service contractor.

D) Key recommendations on Business Value proposition (XVII – XX):

XVII) **Digital transformation should not occur with an approach of "Does not matter what it costs."** What direct business value is generated with the new technology which is implemented? What will all things cost until the full workflows are implemented? What will it cost to maintain all this new technology and workflows? Is the solution sustainable? Does OMV and OMV Petrom get all necessary personnel for efficient cost with relevant skill sets?

XVIII) **No matter what the future OMV IT Architecture will look like, WLMS forms the center of gravity for serving wellbore centric data to users.**

- Better decisions leading to better investments,
- Taking all data into consideration and make full use of all acquired data to reduce uncertainties,
- Faster project execution across all Four Technology Stacks (e.g., subsurface studies),
- More efficient usage of subsurface specialists,

- Reduced investments in data acquisition and operational cost due to efficient usage of offset data and analogs.
- In some instances, data is not accessed directly from DELFI, but from WLMS. Examples include core laboratories, workover teams on the field, etc. …

XIX) Convince management to assign resources and budget to WLMS, in order to start initiatives to get the data into the system by:

- Demonstrating how WLMS data types are involved in all of the Four Technology Stacks, as shown in Figure 35, 36 and 37.
- Focusing on the value WLMS can bring as an enabler of AI in OMV Energy business.
- Demonstrating how WLMS can not only tackle large volumes of data, but also the huge variety of data formats involved in wellbore centric data types (log-, mudlog- and core data).
- Data in Motion with a high variety of data types is hardest to handle for AI systems.
- Data standardization is key to managing this variety of data: It is recommended to build a metadata glossary to understand and train AI to be able to access the data in a structured way and give meaningful insights.

XX) In order to fully make use of analytical and generative AI, data must first be standardized, which is a challenge considering the variety of data in E&P. WLMS solves this problem.

- When looking at all the dimensions of big data (volume, velocity, variety, …), two dimensions are especially important when it comes to AI: Volume and variety,
- Volumes of data can be managed quite easily for ingestion into systems that use either a) Analytical AI or b) Generative AI.
- Variety of data however is much harder to deal with when it comes to getting data into AI systems.
- Data can be categorized as either a) Data at Rest (SQL databases, legacy data, archives, …) or b) Data in Motion (live incoming sensor data, drilling data incoming via WITSML, …).

6.3 Evaluation

The objectives and key deliverables defined in the introduction section of the thesis have been met. In addition, optional objectives such as the business value proposition have been achieved. The thesis project timelines were fully adhered to.

6.4 Future Work

An adoption of OSDU^{TM} Data Platform at OMV and OMV Petrom or its affiliates would be a next step. Before going forward with an implementation, an evaluation project is recommended to be done to see what functionalities are there and how possible data ingestion workflows and data consumption workflows could look like. The earlier the evaluation is done the better it is because gaps (e.g., missing functionalities) can be identified in an early state and actions can be taken.

It is highly recommended that the as is E2E diagram is reviewed in detail between business and IT so that there is a full understanding of each workflow, why it is needed and how the outcomes are used by end users. After this is fully understood, the mapping of the existing workflows to the future environment can start. For business it is important that the implementation of workflows in an OSDU^{TM} / DELFI environment enables business to work more efficient and automatized and that existing workflows are further optimized taking advantage of all modern tools that OSDU^{TM} Data Platform / DELFI offers. It has to be avoided that the new solution leads to less functionalities compared to the existing solution. Go a step forward and not backward!

Four options on service contractor software component integration have been presented within this thesis. A discussion on the most efficient option must be initiated before the decision on which option will be finally chosen can be made. Depending on which option out of the four options is chosen in the end, the next steps must be initiated in cooperation with the service contractor and OMV IT department for a full seamless integration of service contractor applications into OSDU^{TM} Data Platform/ DELFI.

A limitation of this thesis project is that no software implementations have been done, but rather a series of concepts for semi-automatized tools have been thoroughly drafted and presented. Software implementation of automatic data audit with automatized e-mail notifications as well as automatized data gathering would be relevant future work.

Holistic data gathering of well log data acquired during workover can only be implemented properly, once creation of workover programs is fully automatized and can be tracked in the system.

References

A DATE WITH DATA, zuletzt geprüft am 21.06.2023.

Amazon: OSDU Data Platform. Online verfügbar unter https://aws.amazon.com/de/blogs/industries/osdu-data-platform-on-aws-ingestion-series-1-overview-of-data-types-for-osdutm-data-platform/, zuletzt geprüft am 07.07.2023.

Amazon: What is a Datalake? Online verfügbar unter https://aws.amazon.com/de/big-data/datalakes-and-analytics/what-is-a-data-lake/, zuletzt geprüft am 09.03.2023.

Andy McDonald: Loading Well Log Data from Dlis using Python. Online verfügbar unter https://towardsdatascience.com/loading-well-log-data-from-dlis-using-python-9d48df9a23e2, zuletzt geprüft am 02.03.2023.

Antonio, Daniel; Sadlier, Andreas; Winston, Joseph (2019): Leveraging Oil and Gas Data Lakes to Enable Data Science Factories. In: Day 2 Wed, October 30, 2019. SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition. Bali, Indonesia, 29.10.2019 - 31.10.2019: SPE.

Brown, J. (2021, April 22): Data Trust: Gain Understanding and Confidence in Your Data. SPDM Conference 2021. SPDM, 2021, April 22.

Camden, D. (2021, June 22): Contextualization of Structured and Unstructured Data. SPDM Conference 2022, 2021, June 22.

CDA: The Business Value Case for Data Management a Study - Results. Online verfügbar unter https://oeuk.org.uk/wp-content/uploads/2020/09/CDA-The-business-value-case-for-data-management-a-study-Results-1.pdf, zuletzt geprüft am 14.06.2023.

Chang N., LaRue J., Mosley A. (2023, April 18): Overhauling Point-to-Point Data Integration Leveraging the OSDU Data Platform. OSDU. London, 2023, April 18.

Cleveland, Cutler J. (Hg.) (2004): Encyclopedia of energy. Amsterdam: Elsevier.

cwls.org: LAS 1.2 Standards. Online verfügbar unter https://www.cwls.org/wp-content/uploads/2014/09/LAS12_Standards.txt, zuletzt geprüft am 12.04.2023.

cwls.org: LAS 3: File Structure. Online verfügbar unter https://www.cwls.org/wp-content/uploads/2014/09/LAS_3_File_Structure.pdf, zuletzt geprüft am 12.04.2023.

databricks.com: What is Parquet. Online verfügbar unter https://www.databricks.com/glossary/what-is-parquet, zuletzt geprüft am 05.05.2023.

Daum, Steve; Getty, Robin; van Berg, Frans; Clinton, Robert (2010): A Guided Workflows Approach to Oil Field Management. In: All Days. SPE Intelligent Energy Conference and Exhibition. Utrecht, The Netherlands, 23.03.2010 - 25.03.2010: SPE.

Diwakar, Amitesh; Akoum, Mohamed (2019): A Soft but Effective Strategy to Ensure Data Completeness and Data Quality. In: Day 2 Tue, October 22, 2019. SPE Gas & Oil Technology Showcase and Conference. Dubai, UAE, 21.10.2019 - 23.10.2019: SPE.

Dreyer, J. (2019, September 17): JSON Well Log Format, 2019, September 17.

Epam a: Data Platform Overview (OSDU Data Platform Application Developer Training). Online verfügbar unter https://osdu.projects.opengroup.org/subcommittees/business-model-outreach/projects/app-dev-training/work-products/supporting-docs/, zuletzt geprüft am 17.02.2023.

Epam b: Data Platform Overview (OSDU Data Platform Application Developer Training). Online verfügbar unter https://osduforum.org/application-development/osdu-application-developer-training/, zuletzt geprüft am 17.02.2023.

Gruijters, S., De Boever, E., & Visser, W (2021, June 22): Our Journey from Data Standards Towards Open Information Platforms. SPDM Conference 2021. Geological Survey of the Netherlands, 2021, June 22.

Håvard Holmås, Sverre Lindseth, Sylvain Santamarta, Marie-Helene Ben Samoun, Mauro Castilhos (2020): Digital Powers Value Creation in Oil and Gas. BCG. Online verfügbar unter https://web-assets.bcg.com/61/25/c684ac3f4fb4b9abdac9ca4f15f4/bcg-digital-powers-value-creation-in-oil-and-gas-oct-2020.pdf, zuletzt geprüft am 24.05.2023.

Jackson, J., Clymer, R. (2023, April 18): Rock and Fluid Sample DDMS. ExxonMobil accelerating OSDU Platform Value with Partners. OSDU. London, 2023, April 18.

Jim Crompton, Bill Eustes (2023a, March 6): Data Visualization. Petroleum Data Analytics. Colorado School of Mines. Golden, CO, 2023, March 6.

Jim Crompton, Bill Eustes (2023b, March 6): The Data Foundation. Petroleum Data Analytics. Colorado School of Mines. Golden, CO, 2023, March 6.

Jim Crompton, Tim Coburn, John Jacobs: Three Data Wise Guys + One. Vorgelesen von Jim Crompton, Tim Coburn, John Jacobs: PPDM.

Kansas Geological Survey: LAS. Online verfügbar unter https://www.kgs.ku.edu/software/SS/HELP/las/index.html, zuletzt geprüft am 12.04.2023.

Kulbrok, F. (2023, April 18): CSP and Vendor Perspective Presentation: TGS Well Data OSDU Integrations. OSDU. London, 2023, April 18.

Landre, Einar: Thoughts on software architecture and design in the digital age. Online verfügbar unter https://emlandre.com/author/emlandre/, zuletzt geprüft am 23.02.2023.

Landre, Einar (2022, November 29): Equinor OSDU Adaption Experience. SPDM Conference, 2022, November 29.

Microsoft: Azure Data Lake Change Feeds. Online verfügbar unter https://learn.microsoft.com/en-us/dynamics365/fin-ops-core/dev-itpro/data-entities/azure-data-lake-change-feeds, zuletzt geprüft am 17.03.2023.

Ocampo, Alexis; Farah, Gabriel; Gutiérrez, Ernesto (2019): Towards a Self-Serving Big Data Analytics Platform for Oil and Gas. In: Day 2 Tue, November 12, 2019. Abu Dhabi International Petroleum Exhibition & Conference. Abu Dhabi, UAE, 11.11.2019 - 14.11.2019: SPE.

OMV Energy: OMV Internal Document.

OMV Energy (2021a): Data Conditioning Software Development Agreement.

OMV Energy (2021b): Workflow for WLMS related Data. Log Digitization & Digitalization and Digital Transformation Principles.

OMV REAL Regulation. Online verfügbar unter https://omv.sharepoint.com/sites/CoPPetroleumEngineering/Petrophysics/WLMS/SitePages/OGMS-&-REAL.aspx?csf=1&web=1&e=xBRWrShttp://, zuletzt geprüft am 14.03.2023.

OMV-SLB Tech Day (2023). SLB, 14.03.2023.

OSDU: Who We Are. Online verfügbar unter https://osduforum.org/about-us/who-we-are/osdu-mission-vision/, zuletzt geprüft am 22.02.2023.

OSDU (2023a): Functional Architecture. Online verfügbar unter https://community.opengroup.org/osdu/documentation/-/wikis/OSDU-(C)/Reference-Architecture/Functional-Architecture, zuletzt aktualisiert am 14.03.2023.

OSDU (2023b): Demystifying Well-Known-Schemas, Well-Known-Entities, Enrichment-Pipelines. Online verfügbar unter https://community.opengroup.org/osdu/documentation/-/wikis/OSDU-(C)/Design-and-Implementation//Entity-and-Schemas/Demystifying-Well-Known-Schemas,-Well-Known-Entities,-Enrichment-Pipelines, zuletzt aktualisiert am 04.04.2023.

Ramsay, Stephanie; van Brink, Kees den; Marie, Sylvain (2021): The turning point: a novel about agile architects building a digital foundation. First edition, first impression. 's-Hertogenbosch: Van Haren Publishing (The Open Group Series).

Seera, Oberto (2008): Well Logging Handbook.

SLB: Core definition. Online verfügbar unter https://glossary.slb.com/en/terms/c/core, zuletzt geprüft am 04.07.2023.

SLB: Measurements While Drilling. Online verfügbar unter https://glossary.slb.com/en/terms/m/measurements-while-drilling, zuletzt geprüft am 12.04.2023.

SLB: Well log definition. Online verfügbar unter https://glossary.slb.com/en/terms/w/well_log, zuletzt geprüft am 04.07.2023.

SLB a: What is DELFI? Online verfügbar unter https://www.software.slb.com/delfi/what-is-delfi, zuletzt geprüft am 27.02.2023.

SLB b: New Ways of Working with the Delfi Cognitive Environment. Online verfügbar unter https://www.software.slb.com/videos/new-ways-of-working-with-the-delfi-cognitive-e-and-p-environment, zuletzt geprüft am 27.02.2023.

specserve.redguard.com: MWD and LWD - A comparison. Online verfügbar unter https://specserve.redguard.com/mwd-and-lwd-a-comparison/, zuletzt geprüft am 12.04.2023.

Sykora, Elena (2017): DAMA-DMBOK. Data management body of knowledge. 2nd edition. Hg. v. Deborah Henderson, Susan Earley und Laura Sebastian-Coleman. Bradley Beach, New Jersey: Technics Publications; Data Management Association.

Sylvain Santamarta, Rohit Singh, Peter Forbes (2017): How Digital Will Transform the Upstream Oil Ecosystem. BCG. Online verfügbar unter https://boston-consulting-group-brightspot.s3.amazonaws.com/img-src/BCG-How-Digital-Will-Transform-the-Upstream-Oil-Ecosystem-Aug-2017_tcm9-167570.pdf, zuletzt geprüft am 24.05.2023.

The Dangers of Data Swamps and how to avoid them. Online verfügbar unter https://www.dataversity.net/the-dangers-of-a-data-swamp-and-how-to-avoid-them/, zuletzt geprüft am 05.07.2023.

Theys, Philippe P. (2011): Quest for quality data. Paris: Editions Technip.

tika.apache.org: Parser package summary. Online verfügbar unter https://tika.apache.org/1.1/api/org/apache/tika/parser/external/package-summary.html, zuletzt geprüft am 05.05.2023.

Wikipedia: Parsing. Online verfügbar unter https://de.wikipedia.org/wiki/Parser, zuletzt geprüft am 10.07.2023.

Wikipedia: Six Sigma. Online verfügbar unter https://de.wikipedia.org/wiki/Six_Sigma, zuletzt geprüft am 23.03.2023.

Work Product Component Definition. Online verfügbar unter https://community.opengroup.org/osdu/data/data-definitions/-/tree/d85ab877101c007e33cc579e4bd679751d38e6c1/E-R/work-product-component, zuletzt geprüft am 10.07.2023.

Zburivsky, Danil; Partner, Lynda (2021): Designing Cloud Platforms.

Appendices

Appendix 1 – WLMS Data Fundamentals

WLMS is the leading corporate system for well log -, core - and mudlog data. Please find below some relevant definitions of the petrophysical techniques used to generate these data types:

- Well logging can be defined as an evaluation method that: *“[…] senses and measures electrical, radioactive, and acoustic (sound) properties of the rocks.” (Cleveland 2004)*
- Well logging is used to determine petrophysical rock properties. *“The petrophysical properties of a rock are essentially linked to its porosity, its water saturation and its permeability. These parameters control the economic potential of a reservoir.”* (Seera 2008, S. 82)
- A well log is defined as: *“The measurement versus depth or time, or both, of one or more physical quantities in or around a well.”* (SLB)
- Mud logs are described as a type of logs in the following way: *“Mud logs that describe samples of drilled cuttings are taken and recorded on surface.”* (SLB)
- A core is defined as *“A cylindrical sample of geologic formation, usually reservoir rock, taken during or after drilling a well.”* (SLB),
  Core data poses a challenge in WLMS, as there is no industry standard, and the data is highly variable. That is why loading templates for unstructured data exist in WLMS.

For obtaining the petrophysical properties recorded in well logs, acquisition can be done either:

- **During drilling,** resulting in immediate data acquisition (Logging while drilling, LWD)
- **Immediately after drilling,** resulting in a delayed data acquisition (Wireline, slickline, etc.)
- **After well completion / extended production period,** examples:
  - Production logs (e.g., reservoir saturation logs)
  - Workover logs (e.g., packer settings, perforation correlations)
  - Casing integrity logs (e.g., corrosion logs, cement bond logs)
  - Timelapse logs (e.g., fluid level change monitoring)

Wireline logging

Figure 40 shows a wireline truck that is used for running logging tools (sonde) into a wellbore via the wireline drum, connected with two sheaves and a load cell. Wireline uses a cable that transmits data from the subsurface to the surface.

Essential wireline logging equipment components include wireline truck, cable, downhole logging tool, memorizer.

Figure 40 shows a **wireline truck** consisting of the following components:

- Main winch (E) with a pulling capacity of several tons and cable length capacity of up to 9000 meters.
- Auxiliary winch (G)
- Control panel of the winchman (C)
- Surface panel (A)
- Electric generator (H)
- Depth measurement (F)

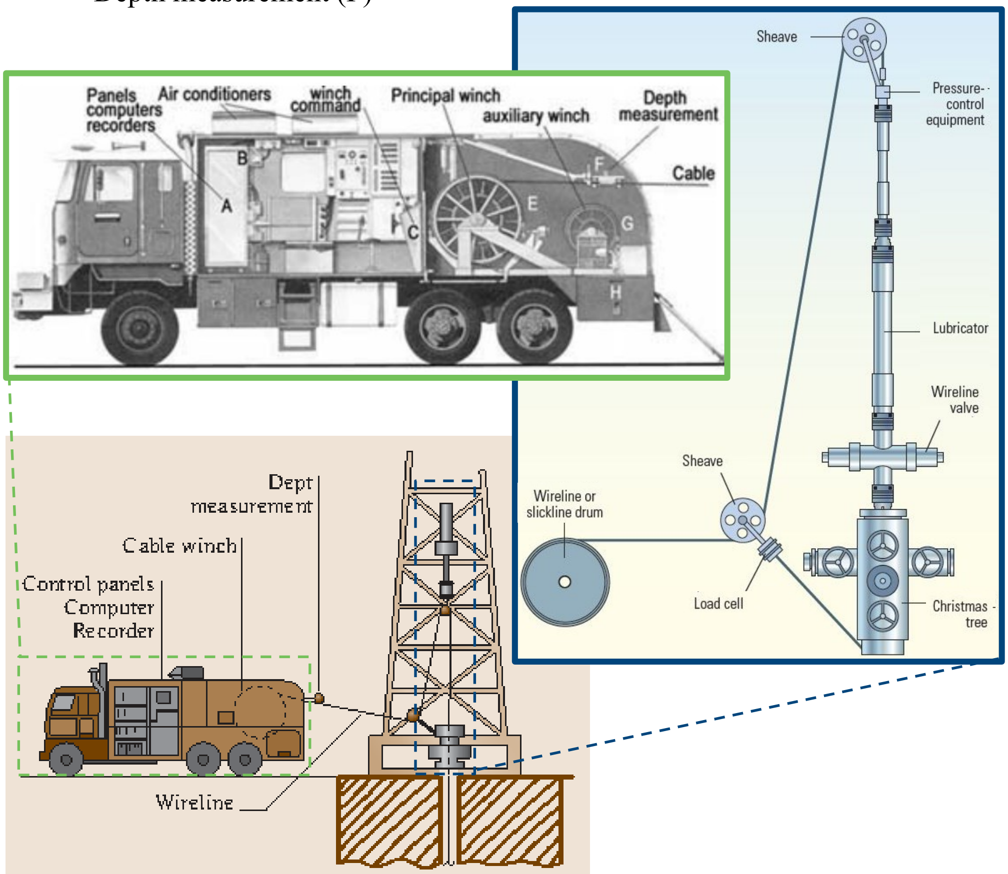


Figure 40: *Top left: Wireline truck with basic components such as depth measurement (F), principal - (E) and auxiliary winch (G), winchman's control panel (C), electric generator (H) and computer panels for data acquisition (A). Right hand side: Wireline rig up with basic components: Wireline or slickline drum, sheave with load cell, pressure control equipment, lubricator, wireline valve and X-MAS tree. (SLB; Hsu, Chang & Robinson, Paul 2017; Seera 2008)*

Opposed to wireline, where the cable is transmitting data to the surface, a slickline does not transmit data and is simply used for installing or removing downhole equipment in the wellbore. Sometimes slickline is also used to enter memory tools into the borehole for recording (e.g., pressure gauges). The **cable** used in wireline logging fulfills the following functions:

- Running - in and pulling out the sonde (low stretch coefficient of the cable is a requirement).
- Electric power supply and data transmission between subsurface and the surface data acquisition equipment.
- Measurement of depth.

The **downhole wireline logging tool** consists of:

- Sonde, made of detector, sensor, emitter, and receiver.
- Cartridge for power supply, signal filtering.
- Telemetry system, which takes a signal that is originally a function of time and converts it to a signal as a function of depth.
- Recording equipment (tapes, disks).

Measurement while drilling (MWD)

- Acquisition of basic physical drilling parameters (wellbore trajectory, borehole pressure, temperature, mud pump volume) in three-dimensional space, during the process of drilling a well. MWD data is collected downhole in real-time and/ or memory mode and then transmitted to the surface.
- Data transmission devices in MWD are similar to LWD data transmission devices, as described in the following section. (SLB)

Logging while drilling (LWD)

Logging while drilling is a great tool for reservoir evaluation, however, the key advantage of LWD compared to wireline logging is the ability for geological steering (geo-steering) and enabling real time decisions that can be made during drilling based on the LWD readings. LWD requires the components listed below:

- **Surface unit,**
- **Rotary steerable devices:** Help to orient the bit,
- **Sonde: sensors are** located in the different drill collars. With the different sensors various data can be measured and acquired for geo-steering and reservoir characterization,
- **Data transmission devices:** For LWD there are several options in terms of transmitting data to the surface:

I) Mud pulse telemetry: Positive pulse, negative pulse or siren telemetry.

II) Electric wave transmission

III) Modern telemetry

IV) Wired drill pipe

Comparison between MWD and LWD

In Table 1, a comparison between MWD and LWD is made.

Technology	Examples	Data Transmission Speed	Detail
MWD	**Measurement while drilling:** Drilling parameters; Directional data	Real time	Low
LWD	**Logging while drilling:** Gamma ray, resistivity, porosity, acoustic, formation pressure etc. …; oriented or non-oriented; 360° images, or sector, or borehole average data	Limited real time data transfer (specific data frames based on operational decisions). Memory data recorded during drilling downhole, memory data, retrievable after tool reached surface	High

Table 1: *Comparison between MWD and LWD technology, giving examples, the speed of data transmission and the level of detail of each technology. (specserve.redguard.com)*

Well logging data formats

Three main (industry-standard) well logging data format types are commonly used:

- LAS
- DLIS
- LIS

with each of them having inherent, unique advantages and limitations. Another option includes converting these formats to JSON for practicability reasons.

- **LAS (Log ASCII Standard):** An industry standard file written in ASCII, in pure text format published by Canadian Well Logging Society around 1990 for storing and transferring E&P data sets (well logs, denoted as curves which exist in the form of gamma ray -, resistivity logs etc.) belonging to a single well. LAS files are pure text files compared to binary DLIS and LIS formats. LAS is intended as a simple file format to exchange well data between operators and contractors in a standardized form. The

LAS file structure is flat, meaning that it cannot easily store array data in the form of a single array. LAS is designed for optical curves only and is structured in up to 6 sections:
VERSION (version number), WELL INFORMATION acting as a header (metadata about the well and file: location and name of the well, data start and stop depth), CURVE INFORMATION (curve units, mnemonics), OTHER (comments) and ASCII LOG DATA (first column: depth values and consequent columns: e.g. spontaneous potential (SP) logging curve values, etc. …). (Seera 2008; Andy McDonald)

Comparison between the features of different LAS versions:

Table 2 gives an overview on the different versions of LAS.

Version	Features
LAS 1.2 (since 1989)	- Basic well log format to store log data on floppy disks - File must be written in ASCII, not binary or using a compression mode - Data types: well log data only
LAS 2.0 (since 1992)	- Dominant version until today - Data types: well log data only - Available sections: log definition section - Supported format type: floating point number
LAS 3.0 (since 1999)	- Data types: Well log data, core data, drilling data - Supports 1D, 2D and 3D arrays - Supports multiple runs - Available sections: log definition -, tops definition and perforation definition section - Supported format types: Floating point, Integer, Exponential, String, Date and Time, deg/min/sec - Adding user defined data is possible and feasible

Table 2: *Comparison of different LAS versions: LAS 1.2, LAS 2.0 and LAS 3.0. The basic and added features are described for each version. (cwls.org; Kansas Geological Survey; cwls.org)*

- **DLIS (Digital Log Interchange Standard):** structured binary file format for storing and transferring E&P information of wells and well logs. *"With the introduction of more complex logging tools with a wide variety of data types (waveforms, arrays...) and record length (from few bits to several thousand bits), the existing formats had difficulty handling variable sampling rates, which could be recorded during the same logging run. To address these limitations, API imposed, as an industry standard, a* *general format known as the DLIS format."* (Seera 2008, S. 118) DLIS was published by American Petroleum Institute (API) in 1991. DLIS is the most common well log file format, but there are not many applications that can read it. *"DLIS files contain large amounts of metadata associated with the well and data. These sections do not contain the well data, these are stored within Frames, of which there can be many representing different logging passes/runs or processing stages (e.g., Raw or Interpreted). Frames are table objects which contain the well log data, where each column represents a logging curve, and that data is indexed by time or depth. Each logging curve within the frame is referred to as a channel. The channels can be a single dimension (one value per depth) or multi-dimensional."* (Andy McDonald)[14]

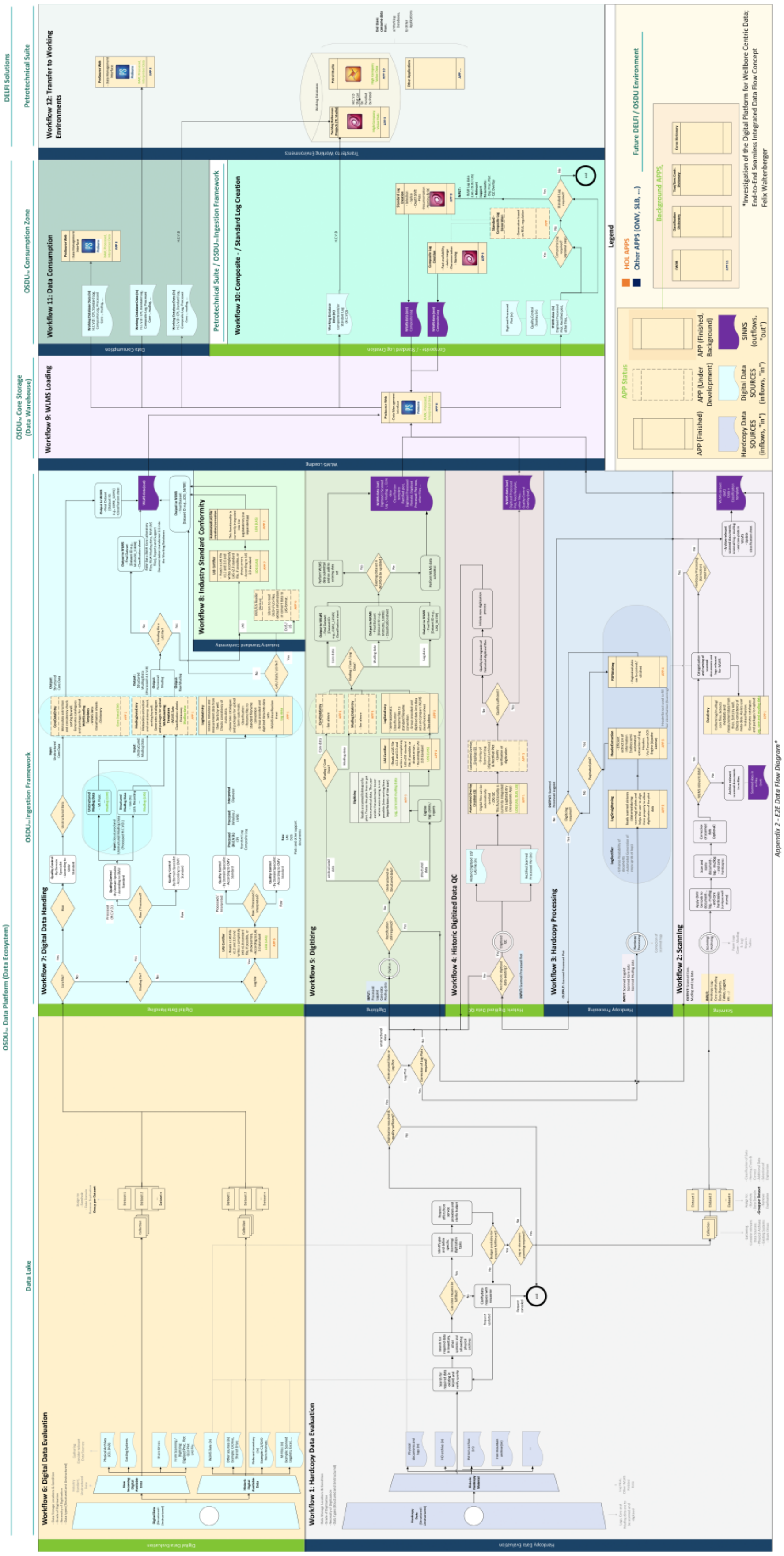
- **LIS (Log Information Standard):** structured binary file format for storing and transferring E&P information of wells and well logs. LIS was published by Schlumberger in 1979 as the predecessor of DLIS. LIS relies on three types of data:
  - I) Data frame: Indexed sensor measurements. The first index value is assigned at the depth where the logging is started and then data frames are continued starting from the primary index value, stored in the field tape.
  - II) Transient information: Comments, messages, etc.
  - III) Static information: Description of the logical structure of the real or disk file and how data frames are formatted. (Seera 2008)

An overview about these standards is given below in Table 3:

Data format	LAS	DLIS	LIS
Full name	Log ASCII Standard	Digital Log Interchange Standard	Log Information Standard
Storage Medium	Past: Disk Present: CD/DVD, Realtime systems (e.g. Interact), FTP, etc. …	Past: Tape Present: CD/DVD	
Data structure	Pure text	Binary	
Advantages	+ Compact + Human-readable	+ Most commonly used + Allowance to contain arrays & textual strings + Variable length frames & types + Variable sampling rate possible + Merge, splice and flip possible + Can manage industry data types such as: mud log data	
Limitations	- Cannot easily store array data - Does not support: variable sampling rates, time logs multi - dimensional curves - LAS 2.0 and 3.0 are incompatible	- Custom reader (dlis) required	- Custom reader required - Outdated

Table 3: *An overview of different data formats used in well logging, including storage medium, advantages and limitations of each. (Seera 2008; Andy McDonald; Dreyer 2019, September 17)*

Appendix 2 – E2E Data Flow Diagram



Appendix 2 - E2E Data Flow Diagram*

List of Figures

List of Tables

Abbreviations

API	American Petroleum Institute
ASCII	American Standard Code for Information Interchange
API	Application Programming Interface
AI	Artificial Intelligence
BO	Branch Office
BI	Business Intelligence
CAPEX	Capital Expenditures
CH	Cased Hole
CWBD	Central Well Data Base
CDC	Change Data Capture
CSP	Cloud Specific Code
CPI	Computed Petrophysical Interpretation
CRS	Coordinate Reference System
DE	Data Ecosystem
DAMA	Data Management Association
DOF	Digital Oil Field
DDMS	Domain Data Management Service
DST	Drill Stem Test
E2E	End-to-End
ERP	Enterprise Resource Planning
E&P	Exploration and Production
EDS	External Data Source
ELT	Extract Load Transform
ETL	Extract Transform Load
GIS	Geographical Information System
HO	Head Office
HOL	Heinemann OiL
HCVD	High Company Value Data
ISV	Independent Software Vendor
IIOT	Industrial Internet of Things
KPI	Key Performance Indicator
LogDB	Log Data Base
LWD	Logging While Drilling
LCB	Low Carbon Business
ML	Machine Learning
MBO	Management by Objectives
MD	Measured Depth
MWD	Measurement While Drilling
MVP	Minimum Viable Product
NPD	Norwegian Petroleum Directorate
OH	Open Hole
OSDU™	Open Subsurface Data Universe
OPEX	Operational Expenditures
OCR	Optical Character Recognition
PTS	Petro Technical Suite

PaaS	Platform as a Service
PSFO	Pro Source Front Office
PoC	Proof of Concept
QC	Quality Control
ROI	Return On Investment
RAFS	Rock and Fluid Samples
RCAL	Routine Core Analysis
SOC2	Service Organization Controls 2 type 2
SaaS	Software as a Service
SCAL	Special Core Analysis
SOW	Statement of Work
SQL	Structured Query Language
SLA	Surface Level Agreement
SOR	System Of Record
TL	Techlog
TVD	True Vertical Depth
UBHI	Unique Borehole Identifier
WKE	Well Known Entity
WKS	Well Known Schema
WLMS	Well Log Management System