D5.2 Data Management Plan

Deliverable No	D5.2
Work package No. and Title	WP5 OPTIDRILL Well, Drilling, stimulation and completion database
Version - Status	V4.0 – draft
Date of Issue	30/06/2021
Dissemination Level	PUBLIC
Filename	OptiDrill-D5.2-Data Management Plan v4.0 final

DOCUMENT INFO

Authors

Author	Organization	e-mail
Mohammad Ashadul Hoque	Technovative Solutions Ltd	ashadul@technovativesolutiuons.co.uk
Shahin Jamali	IEG	shahin.jamali@ieg.fraunhofer.de

Document History

Date	Version	Editor	Change	Status
14/06/2021	v1.0	Mohammad Ashadul Hoque	Initial draft	Draft
28/062021	v2.0	Mohammad Ashadul Hoque	Incorporated DMP for open data scenarios	Draft
28/062021	v2.1	Shahin Jamali	Updated file format, standards, and metadata	Draft
29/062021	v3.0	Mohammad Ashadul Hoque	Draft ready for review	Draft
30/06/2021	v4.0	Shahin Jamali	Final draft	Draft

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The OPTIDRILL project aim to develop a drilling advisory system utilizing novel sensor and machine learning methods to predict ROP, lithology, drilling problems, well completion and enhancement, and unite those methods under one system to enable drilling process optimization and intelligent decision making. The project's main objectives therefore include:

- Develop enhanced drill monitoring systems based on measurement while drilling (MWD) systems and acoustic- and vibration-based sensors.
- Develop automated machine learning-based analysis methods to predict drilling parameters using sensor-based data-driven models.
- Develop a real-time drilling monitoring and optimization tool as a unified system to combine the existing data with the newly developed methods.
- Develop coupled drilling optimization models to reduce geothermal drilling costs
- Develop sustainability model of OPTIDRILL.

The consortium's activities, with respect to data can, therefore, be categorised into the following complementary areas:

- Requirement analysis
- Sensor and data acquisition system specification
- Drilling, formation, and well completion historical data from real world wells
- Data from drilling simulation at Fraunhofer IEG
- What-if scenario data
- Drilling Advisory System
- Administrative data.

The current deliverable describes the consortium's strategy for management of the data collected, processed and/or generated because of the above-mentioned activities. The report includes information on 1) handling of data during and after the end of project 2) types of data collected, processed and/or generated 3) methodology and standards applied 4) data sharing, archiving, and security. The OPTIDRILL data management plan is a *living* document ensuring a dynamic data management lifecycle and integration of updates over the course of the project for any significant changes to the consortium's policies and/or the addition of new datasets.

1. INTRODUCTION

The Data Management Plan (DMP) covers several topics established by the EC and it describes the Data Management life cycle for all data sets that will be collected, processed or generated by the research project. The DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Therefore, DMPs will have a clear version number and include a timetable for updates. As a minimum, the DMP will be updated in the context of the periodic evaluation/assessment of the project. If there are no other periodic reviews envisaged within the grant agreement, an update needs to be made in time for the final review at the latest.

The DMP identifies different items that are relevant to the correct identification and management of the data collected and produced during the project. This means that all the data produced in the OPTIDRILL project will have to be identified according to the following items:
  - Data set reference and name
  - Data set description, purpose, and utility
  - Origin, File format, data size
  - Standards and metadata
  - Data sharing
  - Archiving, preservation, security

2. DATA SUMMARY

2.1 Requirement analysis

The OPTIDRILL project aims to develop a real-time drilling advisory system utilising predictive and prescriptive process optimisation to improved drilling and well completion efficiency; that will significantly reduce the cost of deep geothermal drilling on all types of geological formations and to reach greater depths and higher temperatures, cost-effectively. The requirement analysis identifies user's expectations from the drilling advisory system and its scopes, capabilities, various health and safety requirements, regulation, standards, skills, etc.

Data set reference and name	Stakeholder and end-user requirement analysis
Data set description, purpose, and utility	- User expectation from the drilling advisory system - Drilling advisory system scope and capabilities - Health and safety requirements, regulation, standards, skills etc. - OPTIDRILL validation conditions
Origin, File format, data size	Origin: Consortium member File format: MS office file formats, PDFs, images Data size: Variable
Standards and metadata	Not applicable; this is natural language text (descriptive metadata)
Data sharing	The reports generated will be available in the public domain/confidential depending on the status mentioned in the grant agreement. The reference materials in the report will be subject to copyright as per the journal publishing the referenced article.
Archiving, preservation, security	Stored at Fraunhofer ownCloud in accordance with corporate back-up and recovery policies. Data are protected by copyright.

2.2 Sensor and data acquisition system specification

Data set reference and name	Sensor and data acquisition system specification
Data set description, purpose, and utility	Data set: - Sensor and data acquisition system design specification - OPTIDRILL sensor selection and specification Purpose and utility: - Development of enhanced drill monitoring systems based on acoustic- and vibration-based sensors. - Development of automated machine learning-based methods to predict drilling parameters using sensor-based data-driven models. - Real-time lithology prediction of the formation by employing coupled deep learning methods in combination of real-time strings data. - Measurement of drill induced parameters and system identification
Origin, File format, data size	Origin: Consortium member File format: MS office file formats, PDFs, images Data size: Variable

Standards and metadata	Not applicable; this is natural language text (descriptive metadata)
Data sharing	The reports generated will either be available to the public or confidential depending on the status mentioned in the grant agreement. The reference materials in the report will be subject to copyright as per the journal publishing the referenced article.
Archiving, preservation, security	Stored at Fraunhofer ownCloud in accordance with corporate back-up and recovery policies. Data are protected by copyright.

2.3 Drilling, formation and well completion data from real world wells

Data set reference and name	Historical drilling and logging data
Data set description, purpose, and utility	Data set: • Daily drilling reports • Drilling and completion reports • Formation data • Real-time and Memory LAS data • Real-time and Memory Logs • Other drilling data • Hard drilling data Purpose and utility: ML models development
Origin, File format, data size	Origin: Consortium member and open data File format: various file formats such as Binary database, PDFs Data size: Variable
Standards and metadata	Standards: various standard defined mainly by the data provider Metadata: various metadata
Data sharing	Data sharing will be on need basis within the consortium for the purpose of ML model development, statistical validation and OPTIDRILL drilling advisory system development. Data will not be shared outside the consortium.
Archiving, preservation, security	Copyrighted data will be encrypted by partners before uploading to their private folder on Fraunhofer cloud server. The encrypted data will be copied to local storage (air gapped) local storage for decryption, processing, annotation of sensitive data, sterilization and extraction of relevant data for ML training and validation. Data will be completely deleted at the end of the project. Copyright owner continues holding the copyright of the data. Open data will be stored at Fraunhofer ownCloud server with corporate backup and recovery policy.

2.4 Data from drilling simulation at Fraunhofer IEG

Data set reference and name	Drilling simulation data
Data set description, purpose, and utility	Data set: • Data generated in Match.BOGS Drilling Simulator • Data generated in i.BOGS-Mini autoclave systems

	• Data generated in Field scale test stand Bo.Rex • Data generated in MoUSE system Purpose and utility: Generate missing drilling and logging data in order to develop robust ML models
Origin, File format, data size	Origin: IEG File format: Binary files Data size: Variable
Standards and metadata	Standards: various standards depending on the measurement system used Metadata: various metadata
Data sharing	Data sharing will be on a need basis within the consortium for the purpose of ML model development, statistical validation, and OPTIDRILL drilling advisory system development.
Archiving, preservation, security	Data is stored in Fraunhofer ownCloud server with corporate backup and recovery policy. The ML development team within the consortium have access to the data directories on a need basis for OPTIDRILL development and validation.

2.5   What-if scenario data

Data set reference and name	What-if scenario data of drilling problems
Data set description, purpose, and utility	Data set: descriptions of drilling scenarios, the problems that may occur, drillers' response to each problem and the desired OPTIDRILL response. Purpose and utility: Development of a robust OPTIDRILL drilling advisory system that can detect and predict drilling problems as soon as (or even before) they occur
Origin, File format, data size	Origin: Consortium File format: MS office file formats, PDFs, images Data size: Variable
Standards and metadata	Not applicable; this is natural language text (descriptive metadata)
Data sharing	Data sharing will be on a need basis within the consortium for the purpose of ML model development, statistical validation and OPTIDRILL drilling advisory system development. Data will be shared outside the consortium for research purposes and give industry confidence in OPTIDRILL drilling advisory system.
Archiving, preservation, security	Data is stored in Fraunhofer ownCloud server with corporate backup and recovery policy. Access is restricted to Consortium members only.

2.6   Drilling Advisory System

Data set reference and name	Drilling advisory system design documents, ML modes, and codebase
Data set description, purpose, and utility	The drilling advisory system is the final output of the OPTIDRILL project, which will recommend optimum drilling parameters and detect and predict drilling problems. Confidentiality of these data are crucial for the commercialisation of the OPTIDRILL project
Origin, File format, data size	Origin: Consortium

	File format: MS office file formats, PDFs, images, and programming projects and codes in different languages Data size: variable
Standards and metadata	Best practices in Industrial Control System (ICS) design, development, and commissioning, best practices in ML model training & deployment, and best practices in IT infrastructure design and deployment
Data sharing	Data sharing will be on need basis within the consortium for the purpose of OPTIDRILL drilling advisory system development. Data will not be shared outside the consortium.
Archiving, preservation, security	No personal data will be stored. Stored at IEG and TVS in accordance with their backup and recovery policies. Data are protected by copyright.

2.7 OPTIDRILL administrative data

2.7.1 E-mail lists

At this stage of the project, we do not operate any project internal e-mail distribution list. The e-mail address info@optidrill.eu is publicly available on the website and can be used by potential users to submit queries on the project[1].

2.7.2 Online document library, working areas, meetings

The project uses ownCloud hosted within Fraunhofer for storing all project related information. It is accessible only to the consortium partners. The meetings are hosted using Microsoft Teams applications.

2.7.3 Dissemination material

The project dissemination materials include flyers, brochures, deliverables, posters, and presentations. The deliverables intended for public dissemination will be shared via the project website. In addition, all publications (peer-reviewed and newsletters) will be open access.

---

[1] All website data and website will be in use until 5 years after the end of project.