LTIMindtree

Data Governance using OSDU Data Policy Manager

Authors : **Reetu Ragini, Shankar Velappan, Nitesh Selkari**

Table of Contents

Introduction

In Oil and Gas industry, terabytes of subsurface data viz seismic, well-log, geological and well data are generated during the different phases of data acquisition, processing, and interpretation. The main challenges faced by the E&P companies include storage of these vast volumes of data, data access controls, proper versioning of data, creation of golden datasets and seamless integration of data across multidisciplinary groups. Failure to manage these factors can lead to delays in project completion timelines and even incorrect interpretation leading to project failures.

Over a period of time, Oil and Gas Industry came up with the concept of data governance, which takes into account the processes and methodologies that can be implemented into the business system with standard rules and policies. People accountable for any business function are assigned roles and responsibilities to ensure the correct data is available to the right user group at the right time. Most companies have implemented the data governance framework across their enterprise. Over the last few years, there has been an industry-wide focus on storing data on cloud for efficient storage, retrieval, and data management. However, adopting a cloud native data platform like the OSDU data platform brings major challenges regarding data security and access controls. To address these challenges, OSDU has designed the Policy Service for Entitlement and Obligation of the standard datasets stored. The Policy service ensures that correct data is accessed from the data store by an authorised person only. However, this Policy service may need to be enhanced and customized as per the business requirement.

Business Scenarios

Users from different companies like E&P, data vendors, and oil field service companies access the data stored on the OSDU data platform. A dynamic and robust data governance methodology needs to be implemented. A few specific business cases are mentioned below which show the challenges while dealing with data.

1

Creation of Data Room- When any company initiates a bidding round/farm-out option for its assets, the first step is data room preparation. It includes preparing a list of all important seismic files (along with versions, multiple processing volumes and lineage), petrophysical and well data, and their metadata information. In the old days, the general practice was to store data in some common shared drive, which could be shared with the bidders. Some companies save their data on FTP sites for a certain period when the data can be viewed or downloaded by the clients. In OSDU data platform, temporary access must be provided to the bidders for the specific data elements that are a part of the data room. All these authorizations, access and downloading must be precise and confidential. Any grant of incorrect rights may result into huge data loss and a breach of the security of data.

2

Multi-Client data sets and data sharing as joint venture- E&P companies purchase data from multivendor companies for a specific duration. Data is also shared between companies as a part of a joint-venture opportunity. In these cases, the data is shared between companies based on complex parameters defined in contractual documents (e.g., petroleum sharing contracts, and data purchase contracts). Extraction of these obligation parameters from paper-based documents and dynamically updating the E&O policies can be a challenging task.

3

Economic Sanctions- Different countries impose restrictive measures against a single country or multiple countries, which may include traveling embargo, data access, data exchange, or financial transactions. The E&O policy has to be updated such that any user trying to access the data from the sanctioned country will not be allowed to access the data. As an example, the US has imposed sanctions against Iran. So if anyone from Iran wants to access data in US, it will not be allowed.

In the above examples we observed that the companies need to dynamically control the following:

- Data Sharing
- Data Deletion & Removal of Rights
- Data Storage
- Data Contracts
- Data Access Control
- Data Curation for User's ease
- Data Download Rights

Solution: OSDU Data Policy Manager

The business challenges faced by oil and gas companies were managed by implementing OSDU Policies over data governance Framework, which focus on:

1. Data policies, standards and overall governance strategies
2. Implementation of policies and procedures
3. Enablement of good implementation framework
4. Define strategic roadmap and make continuous improvements
5. Defines roles and responsibilities

OSDU has designed the Entitlement and Obligation policies to address the challenges faced by users from different domains. With these E&O policies, access to the data can be secured by applying secured principles. The policy services will ensure that only correct data is being accessed from the data stores by authorized persons.

LTIMindtree's techno-domain specialist team has designed a solution, "OSDU Policy Manager" which can run on top of this OSDU DPs E&O service, which makes it easy and more effective. The team has built a user interface using which the users can define their own business rules. These enhancements help to add, remove, and update policies based on E&O business rules. Access Control List (ACL) can be updated and controlled using the OSDU policy manager, and the data accessibility can be customized as per project requirements. On top of this, business rules can be used to validate user authenticity and data accessibility (Figure. 1).

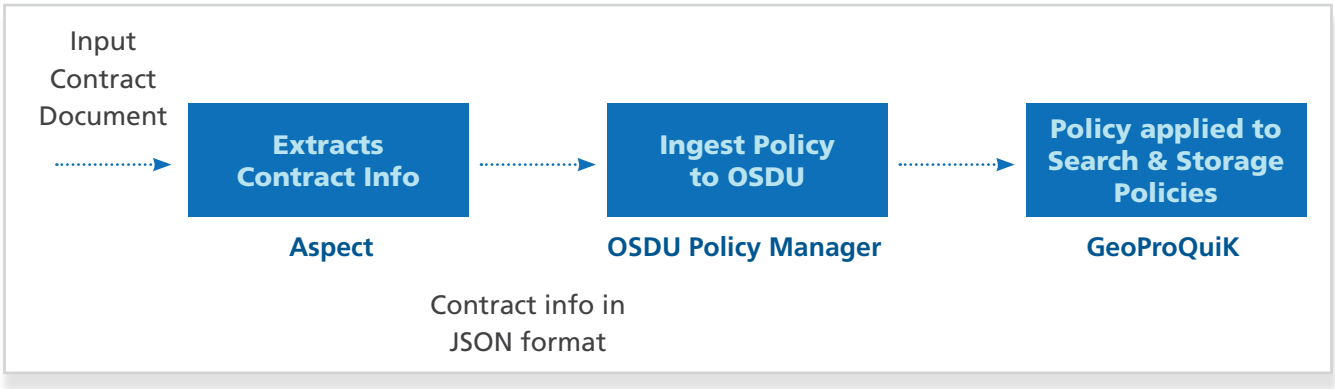

Figure 1: Workflow to extract data using Aspect and integrated with GeoProQuik through the entitlement and obligation business rules and policies.

The creation of this solution for customizing the E&O policy parameters will enable oil and gas companies to address the challenges of data sharing and accessibility.

OSDU Data Platform & Entitlement and Obligation (E&O) Services

Entitlement APIs are mandatory properties to authenticate and authorize access to data by users. These entitlement APIs are used to create different groups. Usually, APIs takes into account the Keycloak JSON (Java Script Object Notation) Web Tokens (JWTs) to validate the user's authenticity. OAuth tokens generated from the Entitlements APIs grants authorization to the requested data partition. This E&O service is implemented using the concept of Data Group, Service Group and User Group.

Data Groups: Validate access of the data. Example- data.wellboredb.viewer and data. wellboredb.owner. (refer from https://www.ibm.com/docs/en/cloud-paks/cp-data/3.5.0?topic=SSQNUZ_3.5.0/svc-osdu/en_api.htm)

Service Groups: Access to the services. Example- service.storage.use and service.storage.admin

User Groups: are created that help in ranking groups of users and the services. Example- users.datalake.viewers , users.datalake.editor and users.datalake.admins

When any user wants to access data, the Client ID, client secret generates Token ID that passes into OSDU. OSDU accepts this refresh token to generate a username and email ID. The Entitlement API gives the list of groups that the user can access if the user account is a member of the specified data partition, if the service account is a member of the specified data partition and service account is a member of a user group in the specified data partition. If these conditions are not fulfilled, the result appears as an unauthorized error. The E&O policies are implemented during data search, data delivery, and ingestion services.

There are two types of E&O security models:

1. **Role-based** – where entitlements are granted to people or groups as viewers, editors ,or admins of services, metadata, or datasets.
2. **Policy-based** – where entitlements are granted through policies using attributes of the users and data.

Data Governance Framework

The data governance framework and OSDU policy manager , have enabled the companies to effectively manage and implement data security and its restriction on accessibility by authorized persons. Since subsurface data acquired for petroleum exploration are a national asset, their governance has been controlled by data residency rules. Companies that own data were initially unaware of the exact location of the Cloud service provider and their details. With E&O policies, all the details of the location of data and its level of access on the server have been controlled by user-based access control (ACL) defined in user groups (Figure. 2).

OSDU data platform provides a certain set of default policies that apply to all service APIs. However, there are project-specific requirements that are not a part of the default policies. To address these business challenges, the LTIMindtree team has worked on adding custom policies, based on the client's specific requirements. To implement the custom policies, the team has leveraged our in-house tool called GeoProQuik. It is a lightweight web-based OSDU aggregator tool with functionalities such as data search through <Full form> Geographic Information System (GIS) based spatial search, data visualization and data ingestion.

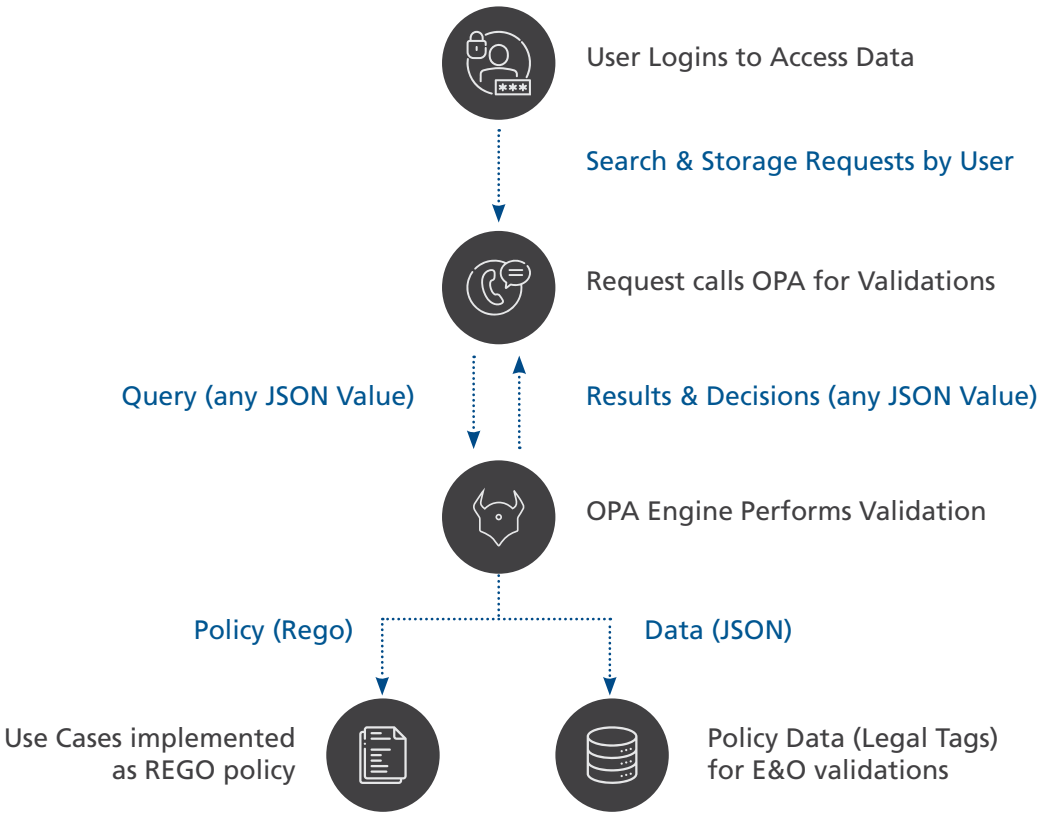

Figure 2: Detailed workflow developed to understand OPA Engine and Policy Services

The E&O service performs two types of validations: user identity and data accessibility. These validations are defined in different sets of rules(policies) and evaluated against the user group information of the user and legal tags of the data sets. These custom rules or policies are written in Rego and implemented using Open Policy Agent (OPA) Engine.

The implementation workflow of Entitlements and Obligations using the Open Policy Agent (OPA) Engine is shown in Figure 1. The user requests data by calling any services viz., search or storage. Based on the policy rules, the request gets evaluated in the OPA engine and the response is given as true or false. In the evaluation process, legal tags and user group information is validated against the Rego policies. The results decide whether the user has access to the requested data or not. The data search and retrieval operations will be executed only if the response from the OPA engine is positive.

Use Cases

Multiple use cases for E&O for which policies are defined and evaluated through OPA engine include Economic Sanctions, State Secrets, Export Restrictions, Production Sharing Contract (PSC), joint operating agreements, multi-clients, data subscriptions, data purchase for multi-client data, data exchange, data rooms, organizational policies like data processing policies, data governance policies and data security and legal policies.

During data purchase or exchange or export or entering into PSC, companies have to undergo various milestones like data security from unauthorized users, location, limitation of access, providing metadata, storing lineage and assigning roles to people who are responsible for playing their job roles in smooth and successful interaction between companies.

In a daily life cycle of a user, E&O policies help to save time, give correct and trusted dataset with lineage and metadata captured, and will avoid delay in flagging the error message if there are any data-related issues. In the below example, we can see how the implementation of E&O as a part of the DG framework helps users at different stages to derive a golden dataset and produce results for business decisions (Figure. 3).

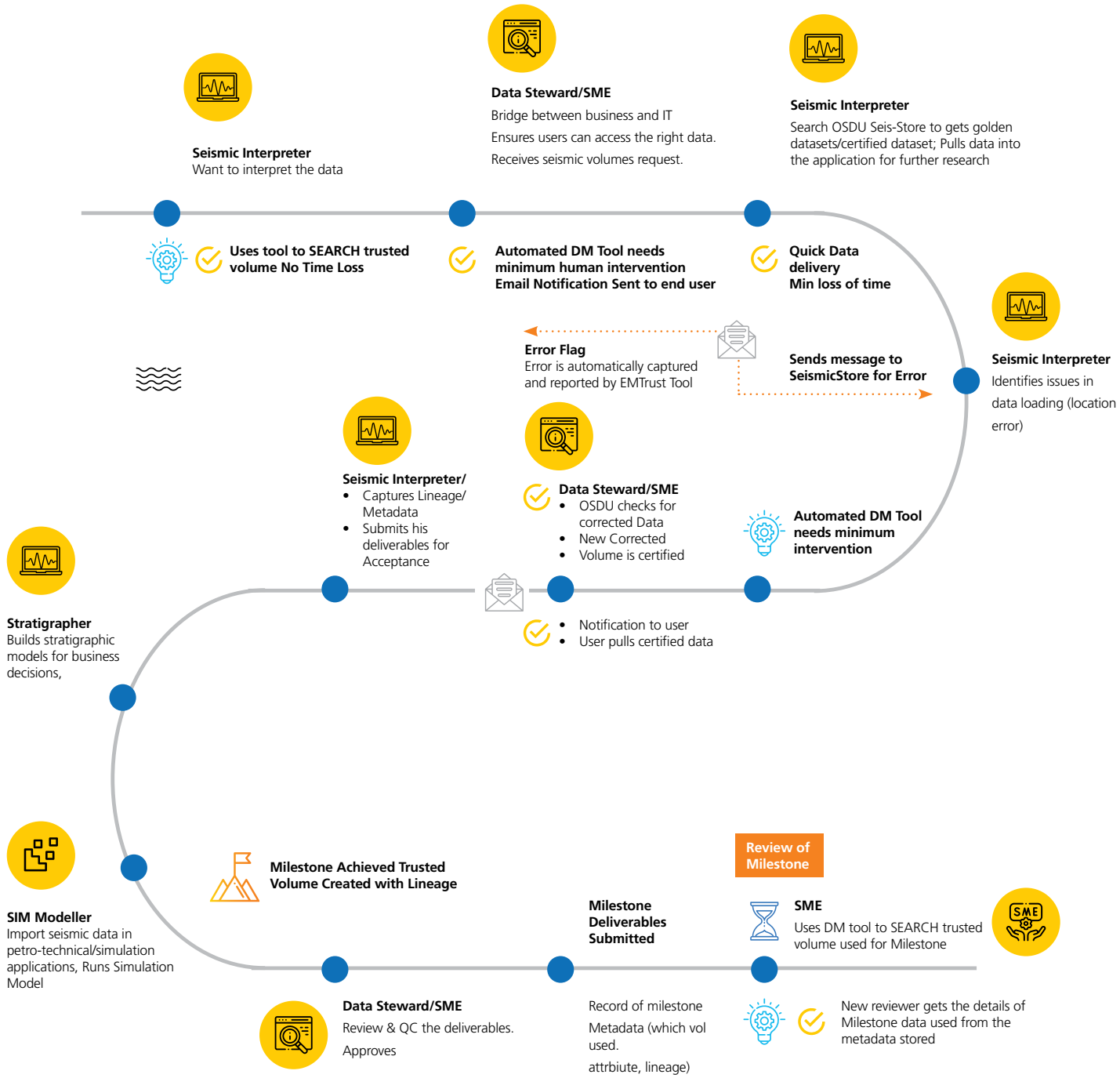

Figure 3: Data management tool helps the user to access a golden set of data and flag errors to the concerned person by sending notifications.

OPA Policy Services

The Entitlement & Obligation services, as a part of OSDU Policy Manager, brings a set of guidelines for mandatory identification information at user level. This information is documented based on the combined decision of data owners and data users. Different processes and functions in the obligatory rule-based OPA Engine is shown in Figure 4. The two major types of information used are user information & legal tag.

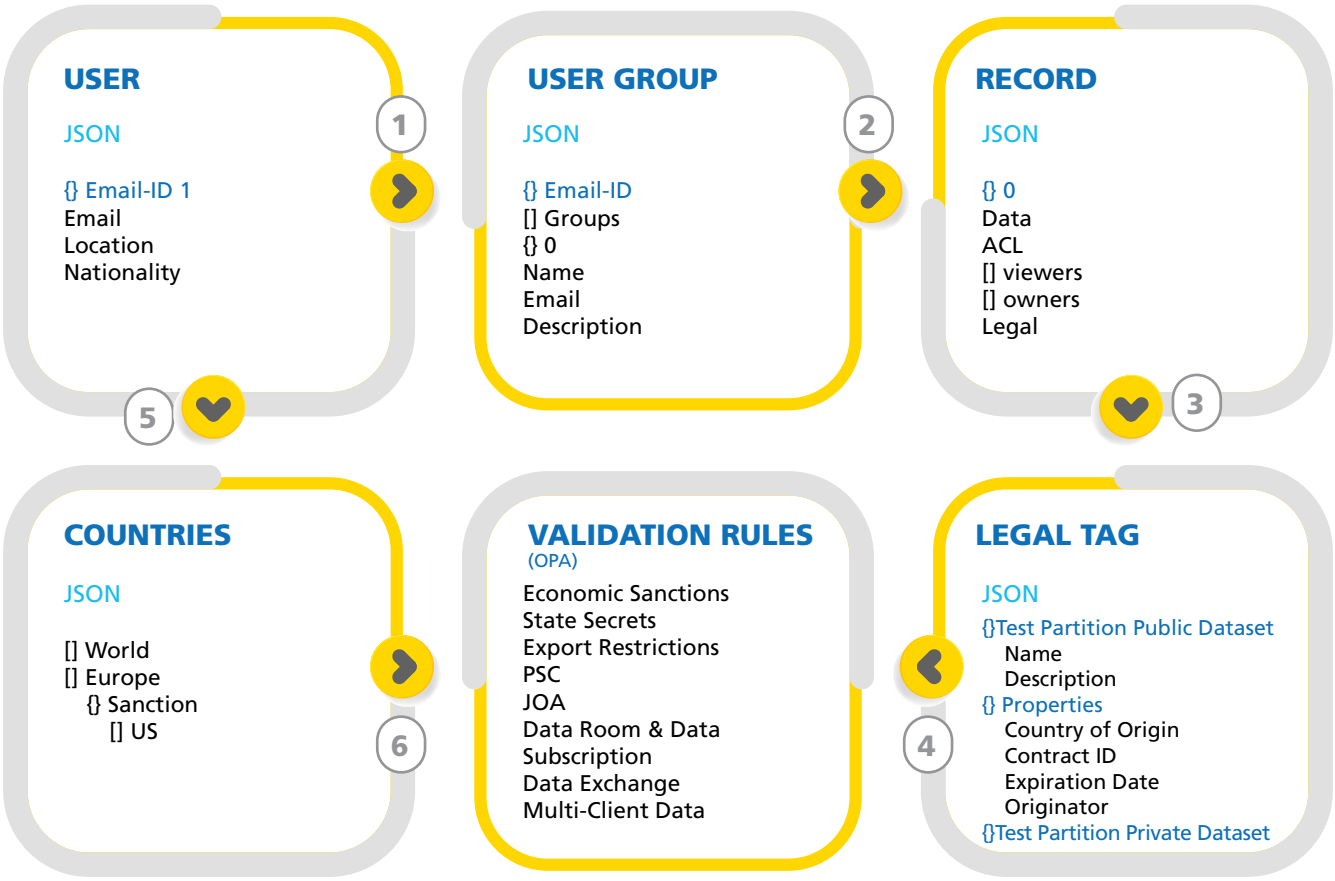

Figure 4: OPA Policy Services and the workflow and business rules applied to secure the data

The different components defined to run the OPA engine are as follows:

USER

When the user completes his login on the OSDU data platform, at the backend the API reads the user email ID, permanent location, and nationality and stores it in the "JSON" format.

User Groups

Based on the user credentials (e.g., email ID) the search result checks the defined user groups to define which type of access the user has been provided in the database. In above Figure it is of "VIEWER" group. The information in user group is matched with the record where data types and its access are defined.

Record

It includes three different attributes:

- Data type - (well, wellbore, marker, seismic etc) and it's kind which is defined the as entity of the data. An ID is created here, which is unique to each data type.
- Access Control List (ACL) is used to define whether the user ID have "data default VIEWER" or "data default OWNER" access for the data type.
- Legal- tag defines the legal rules valid for different countries and nationalities.

Legal Tag

- When a user wants to ingest or store data in OSDU, the legal tag attachment is mandatory. The legal tag API includes different attributes of the data types like Country of Origin, contract ID, expiration data, data type, security class and its export classification. So, the user can insert and also retrieve data from OSDU based on the legal tag.
- In the Policy Manager, the User-User attributes and User-Legal tag attributes are compared and validated to give results as True and False. So, all the use case rules are validated under OPA rules. For example, if a country like US has applied economic sanctions upon other countries like Iran, Syria, Sudan, Russia, etc., then if any person is trying to access data from those locations, he will not have access to US data.

OPA Services Approach

E&O Policies have been defined between OSDU Policy API and Cloud native data-Lakes to restrict user access to the database. Policy services are created to manage the development stages of the policies in OSDU. OPA services apply the validation rules for the user's input query and give results based on the policies applied. The records data are validated based on the Access Control list (ACL) parameters defined for the dataset.

The workflow for this process can be summarised as follows: When any user enters his credentials and login details, a token ID will be generated. This token ID contains all the mandatory details like name, nation, location, access details, whether it is owner, viewer, admin and other essential details. The "Search API" will send the token details to the "Search Service" Request to Policy Services, which then goes to the OPA engine where all API codes have been mentioned. Here there will be a thorough evaluation of token details with the mandatory fields mentioned in REGO and if the token details pass all attributes, it will return the evaluated results to the policy service, and it will be passed on to the Search service. If the user has access, the result will pop up in the Search service response else the user will fail to get valid search results. This is reflected as True/ False in the Policy Service results. The same flowchart is designed for "Storage API" where when the user feeds his initial login credentials, the request is sent to Policy User and then to OPA engine, where his data entries are validated with the REGO Policy rules (Figure 5).

When using the OPA policy engine, its scalability and performance are one of critical trade-offs. However, LTIMindtree techno-domain specialist team successfully created rule-based codes using REGO language and implement it in the OSDU Data Platform.

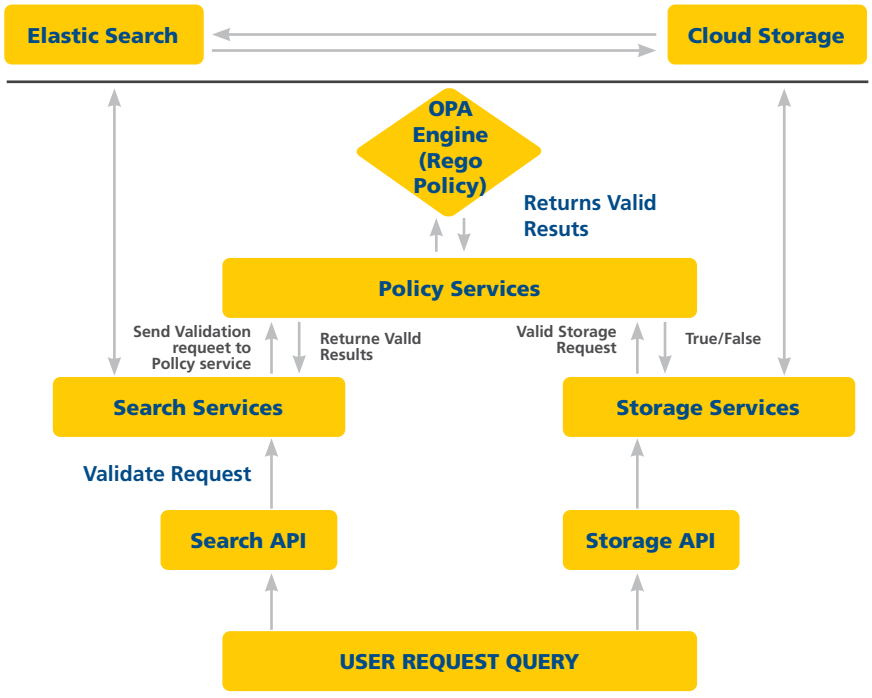

Figure 5: Flowchart depicting the processes involved when a user request goes through the OPA Engine

At an enterprise level, the E&O policies are based on a set of obligations that are defined in contractual documents like Production Sharing Contracts, Subsurface Data Acquisition/Purchase/Sale contracts. Extraction of these parameters manually from paper-based contract documents can be time intensive and prone to human errors.

LTIMindtree has developed an end-to-end solution to extract the metadata from contract documents and update the ACL parameters of the associated datasets on the OSDU data platform. The smart document processing application called "Aspect", can be trained to automatically extract the metadata from contractual documents and converts them into JSON format. Our in-house data QC, Search and OSDU data ingestion application called "GeoProQuik" is used to ingest the JSON data and update the ACL parameters of the concerned datasets on the OSDU data platform. The contractual obligations can then be met using the customized Entitlement & Obligations services of the OSDU. Typically, the major parameters extracted from the contractual documents and used for this workflow include:

1) Legal Tag information
2) Email ID
3) Company Name (s)
4) Block Name
5) Location
6) Contract ID
7) Date of signing of contract
8) Contract Expiration Date

The E&O policies are validated at the user level. All users of a project are assigned to the project group based on the email ID. Whenever the user requests for a data set, the user details are validated against the ACL policies defined for that dataset in the E&O policies of the OSDU. Based on the response (valid/not valid) the user is authorized to access.

Conclusion

OSDU policy services have been created to address the concerns of data security and accessibility in OSDU data platform. It ensures data stored in the OSDU data platform is safe and secure and accessed by authorized person only. A default set of rules have been provided for generic purpose. The LTIMindtree team has successfully built and implemented custom policies in the E&O services based on specific project requirements. These code-based rules were written in Rego and implemented using the Open Policy Agent (OPA). Due to these E&O service enhancements, data access permissions for multiple use cases like economic sanctions, export restriction, Production Sharing Contracts (PSC), Joint Operating Agreements can be successfully implemented.

References

- https://www.ibm.com/docs/en/cloud-paks/cp-data/3.5.0?topic=SSQNUZ_3.5.0/svc-osdu/en_api.htm
- https://community.opengroup.org/osdu/platform/security-and-compliance/home/-/wikis/Design
- https://www.openpolicyagent.org/
- https://community.opengroup.org/osdu/documentation/-/wikis/OSDU-(C)/Architecture-Exploration-Topics/OSDU-R3-Architecture/Entitlements-Use-Cases
- https://osdu.pages.opengroup.org/platform/domain-data-mgmt-services/seismic/open-vds/index.html
- https://www.ibm.com/docs/en/cloud-paks/cp-data/3.5.0?topic=SSQNUZ_3.5.0/svc-osdu/en_api.htm
- https://community.opengroup.org/osdu/platform/data-flow/ingestion/ingestion-workflow/-/issues/56
- https://community.opengroup.org/osdu/platform/domain-data-mgmt-services/seismic/open-vds/-/blob/master/docs/connection.rst

Authors

REETU RAGINI

Reetu is a geoscientist with 15+years of experience in Upstream Oil & Gas industry. In her career across multiple O&G companies she has worked extensively in exploration and development projects. Her core competencies include seismic and well data interpretation, real-time data analysis, and subsurface data management. She has also worked as a part of Managed Services Team and successfully delivered multiple projects in application modernization, data migration and data management. She is highly skilled in G&G Petro-technical software. She is also a Certified Scrum Master.

SHANKAR VELAPPAN

Shankar has 20+ years of experience in the IT Industry in Data Analysis and Application development. He has experience working with major Oil & Gas companies in implementing solutions for Subsurface Data Management and Drilling applications. He is an industry expert in implementing solutions in areas of Data audit, migration, quality analysis, business rules creation & dashboarding. He has rich working experience in various G&G applications, Data Models and cloud-native platforms like OSDU.

NITESH SELKARI

Nitesh is a software developer with 9+ years of experience in application development and testing. He has worked widely across E&P and OFS companies. His expertise lies in automating workflows, application re-engineering, designing microservices-based solutions, and deployment strategies. He is well versed in DevOps process to deploy services in clients' environments for production release. He has contributed significantly to OSDU's Entitlement and Obligation to incubator project, testing, and certification teams.

LTIMindtree is a global technology consulting and digital solutions company that enables enterprises across industries to reimagine business models, accelerate innovation, and maximize growth by harnessing digital technologies. As a digital transformation partner to more than 700 clients, LTIMindtree brings extensive domain and technology expertise to help drive superior competitive differentiation, customer experiences, and business outcomes in a converging world. Powered by more than 84,000 talented and entrepreneurial professionals across 33 countries, LTIMindtree — a Larsen & Toubro Group company — combines the industry-acclaimed strengths of erstwhile Larsen & Toubro Infotech and Mindtree in solving the most complex business challenges and delivering transformation at scale. For more information, please visit www.ltimindtree.com.