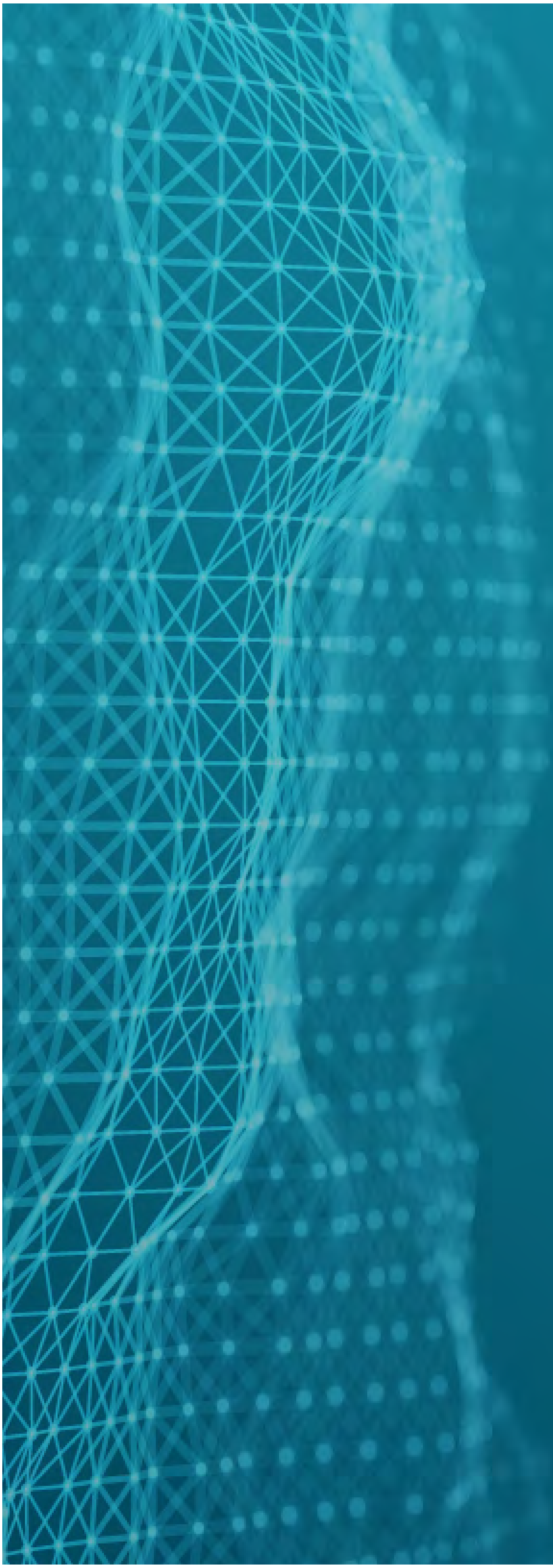
Offshore Energy Digital Architecture
NZTTP Programme

Data Sharing Landscape

OEDA

Net Zero Technology Transition Programme

Net Zero Technology Centre
Technology Driving Transition



“

Overall, the recommendations from the OEDS report have been combined with Data Practitioner expectations and translated into a series of requirements to support the OEDA project.”

Contents

1.0

Summary

Data has a pivotal role to play in delivering an integrated energy sector; without greater active collaboration on shared data sources, toolkits, and best practices, the full potential of the energy transition will not be realised. Many organisations in the offshore energy sector are co-operating on digital transformation implementations, but there remains a lack of clarity on common digital best practices to be adopted. There needs to be a paradigm shift in how organisations delivering the transition can move beyond a model of co-operation to a model of active collaboration around trusted data sources.

The Offshore Energy Digital Architecture (OEDA) project is fundamentally a pilot data sharing platform that enables awareness and access to relevant datasets, to enable shared analytics and increased use of data across the sector to support decision-making, increased use of automation, remote-control technologies, and improved operational efficiency. OEDA "will support the integration of the data and digital infrastructure that is required to deliver the future offshore energy system and demonstrate that we can secure, capture, and make available critical industry data in a manner that is as open as possible", and create a digital energy technology ecosystem which will maximise UKCS-related digital activity[1].

The scope of the OEDA project is to collate the relevant OEDA Technical Requirements from the wider energy sector, determine if it is technically feasible, and then utilise a pilot programme to further inform requirements and next steps for subsequent phases. The pilot platform is Foundry from consortium member Palantir Technologies, which has a proven record for increasing collaboration in the aviation sector through Skywise, a digital platform from Airbus. In practice, Skywise extended beyond the provision of a platform but was a core element of Airbus' digital transformation.

The Net Zero Technology Centre has partnered with InDHu, a start-up that has the principal members responsible for driving the digital transformation at Airbus, to provide a literature review and configure Foundry for the pilot in phase 2 of the OEDA project. The scope and purpose of this report series is to gather existing implementations, recommendations, and best practices from the wider energy sector into a preliminary set of requirements. Experience from the pilot will also help determine and refine the proposed OEDA Requirements to support subsequent phases and other ongoing collaborative data initiatives.

The Offshore Energy Data Strategy (OEDS) Taskforce brought together key stakeholders from across the offshore energy sector and identified two relevant recommendations: Action 2.1: Offshore Energy Data Catalogue (OEDC) and Action 2.2: Data Sharing Fabric (DSF). The recommendations are strategic in nature and are used to define the core deliverables of the OEDA project.

To better understand the context of the OEDS recommendations for a Data Catalogue and Data Sharing Fabric, an extensive literature review was conducted to derive a set of requirements based on a consolidated set of recommendations, best practices, and lessons learned from existing implementations across a chain of eight reports in the wider energy sector, both onshore and offshore, from June 2019 to June 2022. Additional requirements based on the expectations of Data Practitioners were also proposed, a term loosely defined to cover a wider range of job titles that utilise data to meet the same aims as OEDA in terms of decision-making, automation, and operational improvements.

Overall, the recommendations from the OEDS report have been combined with Data Practitioner expectations and translated into a series of requirements to support the OEDA project.

[1] The Oil & Gas Technology Centre (2020) - Net Zero Technology Transition Programme - Appendix VII Offshore Energy Digital Architecture (OEDA) Business Case.

2.0

Offshore Energy Digital Architecture (OEDA)

There are five reports for establishing a sector wide Offshore Energy Digital Architecture (OEDA):

1

OEDA
Data Sharing Landscape

2

OEDA
Technical Feasibility

3

OEDA
Pilot Architecture and Ontology Design

4

OEDA
Potential Business & Cost Model based on Pilot

5

OEDA
Review

This is the first report of the series, to which its purpose is to define and derive technical requirements for a OEDA Data Sharing Platform by evaluating existing implementations, best practices, and recommendations from the wider energy sector and translate them into terms understood within the data industry. The second report determines whether a OEDA Data Sharing Platform is technically feasible using an example Open Source based architecture to perform the evaluation. The third and fourth reports are based on the evaluation of the pilot data sharing platform and associated business and cost models. The final report documents the OEDA project and provides recommendations to establish next steps.

To help determine requirements for a sector-wide data sharing capability, the OEDA project will use Palantir Technologies' Foundry platform along with InDHu as partners for a Pilot. This was primarily due to the success of Foundry in the aviation sector with the implementation of Skywise[2]. Airbus was able to create an ecosystem aimed at accelerating and expanding the exploitation of aviation data across multiple parties, including customers, suppliers, and even competitors in the field of aircraft maintenance. The foundation for their digital platform was Foundry, and many of the key personnel who supported the Airbus Digital Transformation are now part of the InDHu start-up. In the best traditions of the NZTC in trialling new technologies for the offshore energy sector, the OEDA project will evaluate Foundry as a Pilot for the OEDA Data Sharing Platform with the expertise of InDHu in its deployment and configuration.

[2] Airbus (2023) - Skywise | Enhance | Services

3.0

Scope

The purpose of the report is to define and derive technical requirements for a OEDA Data Sharing Platform by evaluating existing implementations, best practices, and recommendations from the wider energy sector, and translate them into terms understood within the Data Industry. Furthermore, this will inform subsequent platform evaluations from other providers.

The consolidated output will be referred to as the OEDA Technical Requirements and used as the single technical basis for subsequent phases of the project, that will eventually lead to a tender for a data sharing platform. The requirements will be based on a chain of eight reports for the wider energy sector, from June 2019 to June 2022.

4.0

Background

In 2020, the Business Case for OEDA (included in Appendix VII of the Net Zero Technology Transition Programme report) identified "the complexity and scale of the challenge to integrate the data from multiple organisations, sectors, technologies, and solutions is substantial. There is a significant risk that meeting the 2045 net zero target will be impossible without investment in deploying key digital technologies in support of this target. Transformation will be excessively costly if these technologies are not deployed in a co-ordinated, collaborative way to avoid a slower, more expensive transformation".

OEDA is fundamentally a data sharing platform that enables awareness and access to relevant datasets, demonstrates "shared analytics platforms that are as open as possible", and promotes "increased use of data across the sector to support decision making, increased use of automation, remote control technologies, and improved operational efficiency".

In August 2021, the Scottish Government awarded the NZTC a £16.5million investment programme to accelerate a range of energy transition projects to help deliver Scotland's net-zero economy. The Net Zero Technology Transition Programme is expected to enable £403billion for the economy and 21,000 jobs by 2050; it covers seven projects that have matched funding from industry.

Many of the stakeholders for OEDA include participants in the Offshore Energy Data Strategy (OEDS) Taskforce, which made two key strategic recommendations with regards to a data sharing platform. OEDA is not an isolated initiative but forms part of a significant movement within the wider energy sector that has produced multiple projects and at least eight related reports, for both onshore and offshore, over a three-year period between June 2019 and June 2022.

Energy Hub
NZTTP Programme

Data for Net Zero
NZTTP Programme

OLTER
NZTTP Programme

Advancing Remote Operations
NZTTP Programme

Offshore Energy Digital Architecture
NZTTP Programme

[3] The Oil & Gas Technology Centre (2020) - Net Zero Technology Transition Programme - Appendix VII Offshore Energy Digital Architecture Business Case.
[4] Scottish Government. (2021) - Investing in net-zero technology - gov.scot

5.0

OEDA Technical Requirements

Table 1 is a summary of the key technical requirements derived from an extensive literature review to evaluate existing implementations, best practices, and recommendations. To avoid repetition, only the requirements that have a technical influence are reproduced here. The absence of a particular key recommendation is therefore not due to a lack of importance; but the associated technical requirements have been captured previously.

For example, many of the reports discuss the principle of "Presumed Open" data, which is an important topic but is not stated explicitly below. The associated technical contribution has, however, been captured in Requirement E2 with regards to a Data Sharing Fabric, which itself defines different access policies. The principle of Presumed Open would therefore be accommodated by setting the default policy to one that has very permissive access controls or authorisations.

Similarly, there are varying "standards" and minimum expectations regarding metadata. Rather than list all the key-value pairs contained within the eight reports, the key technical requirement is the flexibility to choose what they are, enforce some as mandatory, some as optional, and instigate a hierarchy. The technical implementation and hence requirements are not dependent on knowing what the metadata represents.

The table below is a consolidated list of requirements derived from the recommendations within the wider energy sector reports. It should be noted that there are additional expectations of a data sharing platform from typical data industry users that are not captured within the energy sector reports. In Appendix A, a series of additional requirements are proposed that are based on industry experience that the offshore energy sector would benefit from and have been included for consideration.

Req. ID	Requirement	Source(s)
E1	OEDA shall support the OEDS defined Data Catalogue.	From Action 2.1: Offshore Energy Data Catalogue (OEDC).
E2	OEDA shall support the OEDS defined Data Sharing Fabric.	From Action 2.2: Data Sharing Fabric (DSF).
E3	OEDA shall be based on Open-Source software and open standards. It should facilitate the Presumed Open principle.	The principle of being as "Open as possible" as expanded in the EDiT[5] report as: "Wherever possible, it is proposed that these should be based on Open-Source software, open data licences and open standards"
E4	OEDA shall support a customisable set of attributes to act as metadata and have the means to define differing levels of priorities and controls.	Several metadata attributes have been defined, in effect the superset from Ice Breaker One on Open Net Zero[6], EDVP[7] and Dublin Core[8] but recognising the need to set and control differing priorities.
E5	OEDA shall support external URL redirects, HTTP based APIs, the means to redirect to static files and other protocols to support streaming applications.	The ONS Energy Data Visibility project stated the protocols initially should be HTTP based, but recognised with maturity it should support streaming applications.
E6	OEDA shall support metrics regarding the data.	The EDVP identified the need to surface and measure data quality - the subjective component in assessing data quality will be influenced by existing Industry standards-based initiatives. The implication is users manually submitting feedback.
E7	OEDA shall support means for prioritising data sets, either for release, update or additional context.	Multiple reports including EDVP and EDTF cited a two-phase approach to data sharing, where users can see a list of potential sources and request them. These are then prioritised for release based on requests received.
E8	OEDA shall support a mechanism to enable users to provide direct feedback to Data Providers.	Multiple reports have cited providing feedback between users and Data Providers, the former to help improve the data sources and the latter to support internal business cases.
E9	OEDA shall display lineage or provide the means to define a lineage between datasets. OEDA shall support datasets to be related using attributes.	EDVP also identified the need to establish both Data Provider led and user driven relationship mapping between datasets.
E10	OEDA shall support and maintain support for the highest security standards in the field of Authentication, Authorisation and Zero Trust (including defence in depth).	OEDS report states in Action 3.2 Cyber Security: "The offshore energy sector should continue to prioritise cyber security, adhering to cyber security best practice and disseminate progress to the wider sector to help developing industries."

Table 1: Technical Requirements derived from the Energy Sector

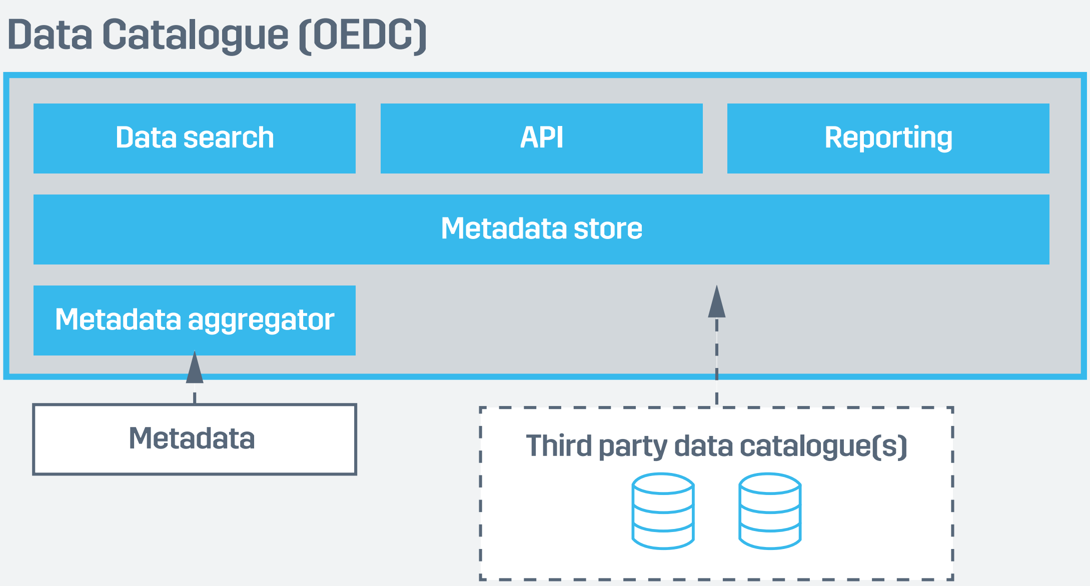
[5] Energy Systems Catapult (2022) - Delivering a Digitalised Energy System
[6] Icebreaker One & Open Net Zero (2023) - Open Net Zero by Icebreaker One
[7] Hippo Digital (2020) - Energy Data Visibility [Discovery report]
[8] Dublin Core Metadata Initiative (2023) - DublinCore

6.0

OEDA Data Sharing Platform

The Energy Systems Catapult (ECS), under the umbrella of the Offshore Energy Data Strategy (OEDS), recommended the delivery of a common data toolkit with two explicit actions for an Offshore Energy Data Catalogue (Req. ID. E1) and a Data Sharing Fabric (Req. ID. E2). It is proposed for the offshore energy industry that an OEDA Data Sharing Platform shall be defined as the combination of these two elements. The purpose of this section is to describe the two recommended actions from a technical point of view to enable evaluation of the existing implementations and also consolidate the key technical requirements.

The OEDC is different from the Data Catalogues typically used within the data industry or in academia, as per this extract:



High-level schematic of a data catalogue system. Metadata can be ingested through a centrally operated metadata aggregator that retrieves and updates metadata from relevant organisations. Organisations with more mature data management processes may wish to integrate more directly with the system and push their own updates to the data catalogue or synchronise compatible systems.

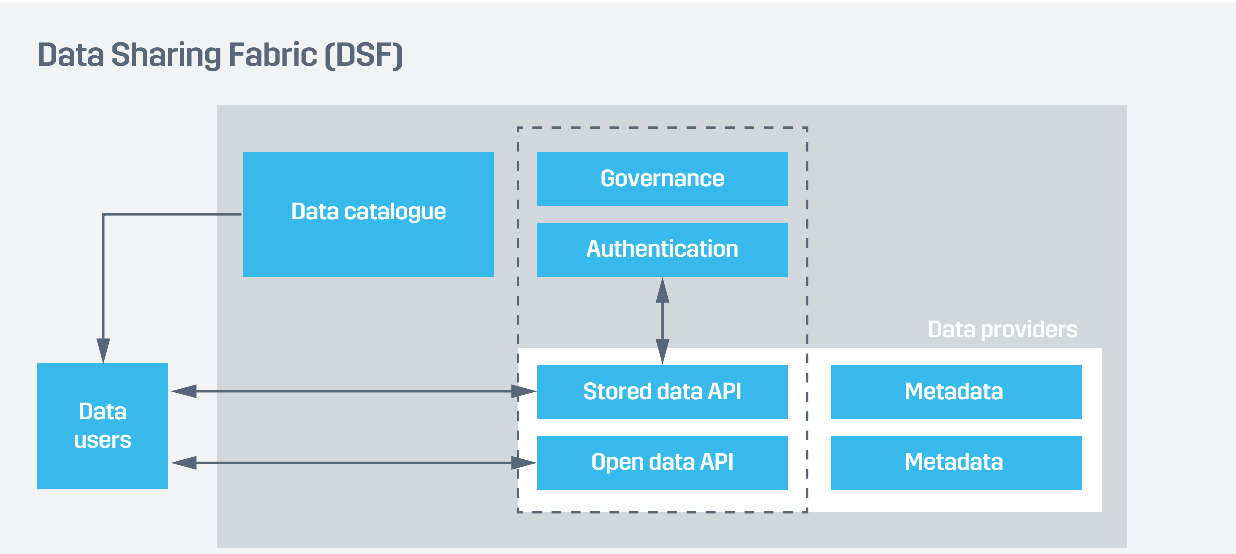
Figure 1: Extract of Data Catalogue from Figure 9, p30 of ECS OEDS Digitalising Offshore Energy Systems[9]

9 Energy Systems Catapult (2022) - Digitalising Offshore Energy Systems

The Data Catalogue consists of three high level functions: a mechanism to search the catalogue using a direct visual interface; an Application Programming Interface (API) to support a range of manual, automated, and machine-to-machine interactions; and finally, a reporting capability around the catalogue utilisation, such as the number of sources, their relative usage, etc.

These functions interact with a metadata store, which is populated through two mechanisms. The first, and to which is the distinguishing feature relative to Data Industry definitions of a catalogue, is the metadata aggregator, an automated means to retrieve and update data from relevant data providers. The second is direct access for more mature data providers to provide metadata directly or synchronise with compatible systems; this behaviour is often referred to as federation of services i.e., peers mutually exchanging data in support of a collective goal.

The catalogue does not stipulate any access requirements, as they are captured in the Data Sharing Fabric:



High-level schematic of a Data Sharing Fabric, detailing the use of an industry governed authentication service to facilitate the exchange of shared data on a limited basis to Data Users. It also shows how the Data Catalogue can be used in conjunction with the Data Sharing Fabric to facilitate authenticated data requests.

Figure 2: Extract of Data Sharing Fabric from Figure 10, p31 of ECS OEDS Digitalising Offshore Energy Systems

The DSF is an industry-governed authentication (AuthN) and authorization (AuthZ) service to facilitate the exchange of shared data on a controlled basis that can control access to Application Programming Interfaces (APIs). Within the data industry, it could be considered as a combination of an Identity Provider (idP) and an API Gateway.

The system should permit an authenticated user for the catalogue, depending on the level of authorization, to access the upstream data providers. The schematic above illustrates a mechanism for governance, which defines the users that can connect, their level of authorization, and an authentication service that delivers the required access. In practice, the governance could be seen as a series of individual or shared access policies, where multiple policies may be applied to a user to provide granular control.

The bottom two green boxes are shown as two separate APIs, one for shared data and a second for open data, whereas in practice a single API Gateway would redirect a request (given sufficient authorization) to the right API. In effect, the APIs would be linked to an access policy, which could use the same tags as above (e.g., Shared and Open) to provide the same effect. To meet the Presumed Open stance, a default policy could be applied to the API Gateway to permit permissive access to resources, including (if required) no controls or restrictions at all.

The balance of the technical challenge is the provision of the Data Catalogue; to map the requirements with sufficient fidelity, additional context about the data, the likely metadata, its management, and some of the use cases is required. A key assumption is the technical interfaces (e.g., protocols) supported by the Data Providers, which directly influence the technical burden of the metadata aggregator within the OEDC. For example, if Data Providers use REST based APIs, this would minimise the technical burden on the metadata aggregator, as this is a tried and tested technical design pattern for automation. In contrast, a Data Provider that uses a custom authentication system would lead to bespoke integration efforts for the metadata aggregator in creating suitable bots to automatically log-in and extract the relevant information. In addition, it would be susceptible to breakdown should the upstream provider make any changes to their system.

Both the OEDS recommendations for a Data Catalogue and Data Sharing Fabric reference several other reports and packages of work across the energy sector. Therefore, to establish a fuller set of requirements for a OEDA Data Sharing Platform, an extensive literature review has been conducted to understand the energy sector landscape. The intention was to evaluate existing implementations, best practices, and recommendations from the wider energy sector and consolidate them (with the right context) into a single set for the OEDA Technical Requirements.

7.0

Energy Sector Landscape

To better understand the context of the OEDS recommendations for a Data Catalogue and Data Sharing Fabric, evaluate existing implementations, and collate key recommendations, the following chain of inter-related energy sector reports were assessed from 2019 to 2022:

The reports were grouped into four segments for convenience and addressed in reverse chronological order to capture the most relevant and recent information first. Although not explicitly mentioned within this chain, the Open Subsurface Data Universe (OSDU) will also be briefly addressed:

1.1. Energy Digitalisation Taskforce (EDiT) - Jan 2022
1.2. ONS Energy Data Visibility - June 2020
1.3. Energy Data Visibility Project (EDVP) - July 2021
1.4. Energy Data Taskforce (EDTF) - July 2019
1.5. Open Subsurface Data Universe (OSDU)

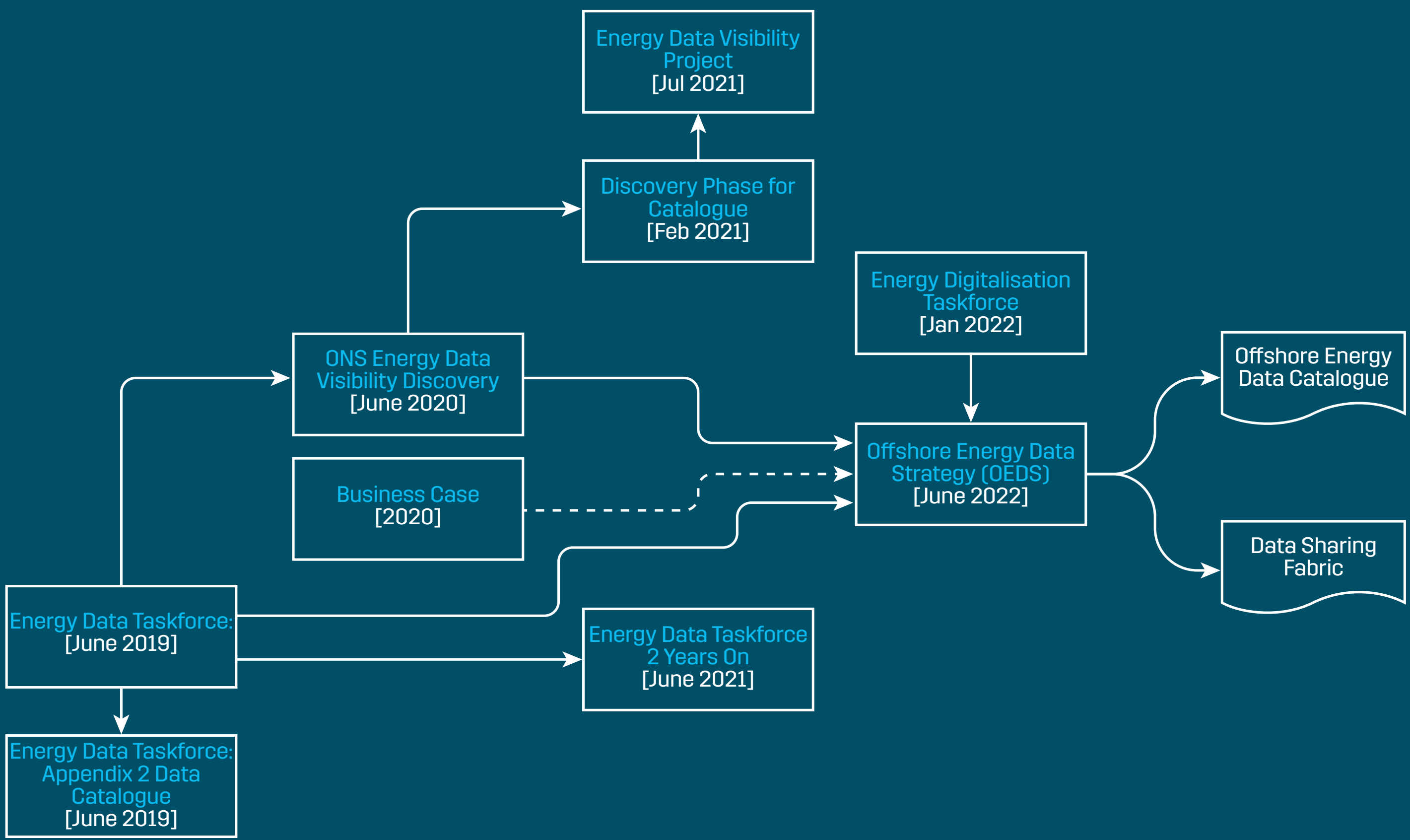


Figure 3: Chain of recommendations from the Energy sector

7.1 Energy Digitalisation Taskforce (EDiT) - Jan 2022

Figure 4: Energy Digitalisation Taskforce reports chain

One requirement (Req. ID. E3) is derived, and one implementation assessed from this segment of the reports chain. The OEDS report cites the Energy Digitalisation Taskforce (EDiT), which recommends an open approach: "Wherever possible, it is proposed that these should be based on Open-Source software, open data licences and open standards to maximise the utility of the sector investment, increase the likelihood of international adoption and minimise technical monopolies" (Req. ID. E3). A working example of a Data Catalogue and Data Sharing Fabric is also provided from Icebreaker One called Open Net Zero . The figure below shows the Open Energy Architecture from the Open Net Zero project annotated to map to OEDS terminology:

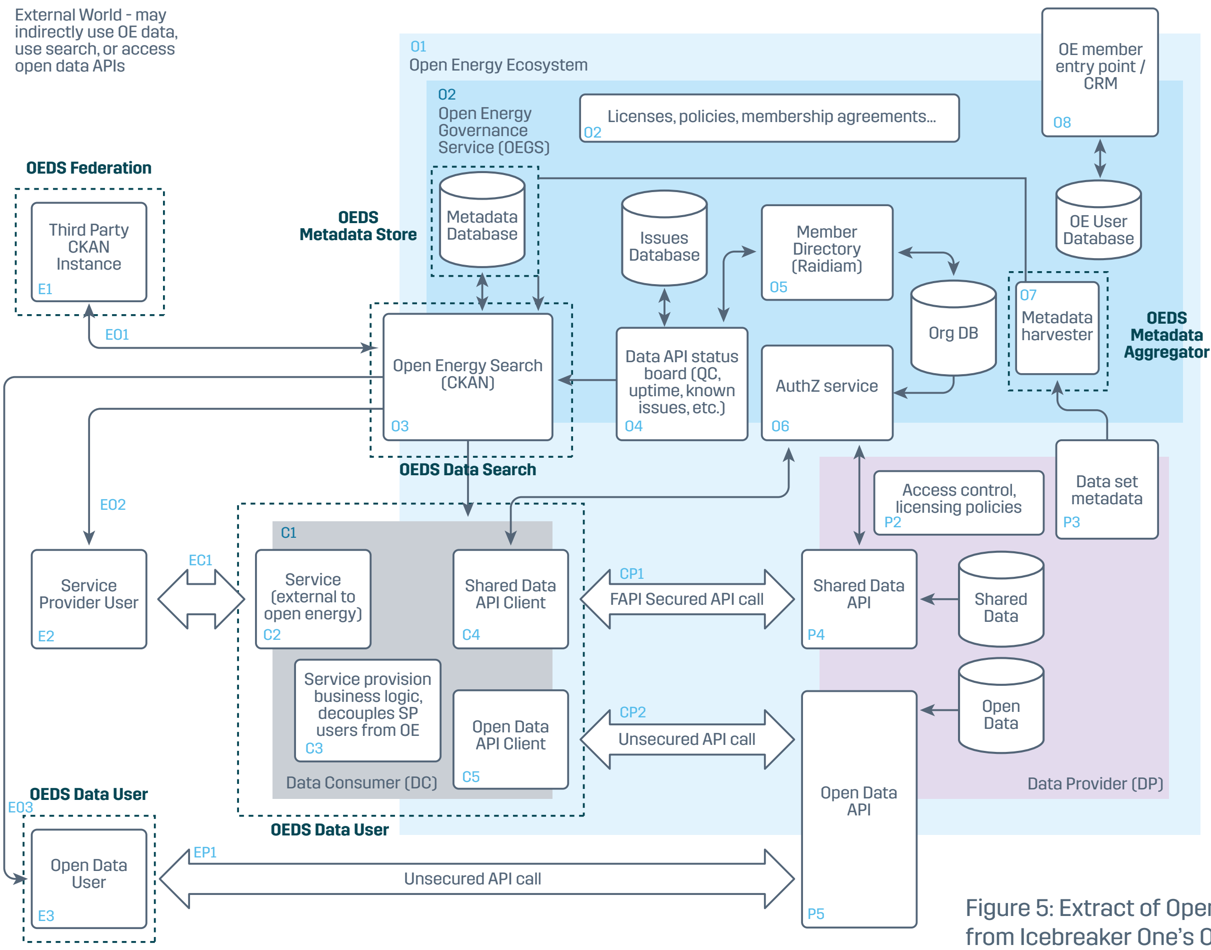


Figure 5: Extract of Open Energy Architecture from Icebreaker One's Open Energy project

[10] Icebreaker One (2023)
[11] Icebreaker One (2023) - Open Net Zero
[12] Icebreaker One (2022) - Open Energy Architecture

Overall, it has similar characteristics to the Offshore Energy Data Catalogue (OEDC) in that the Data Consumer (or Data User in OEDC terminology) can utilise a Shared Data or Open Data API. The Open Energy Search is provided by an open-source data management system called CKAN[13] and is comparable to the OEDC Data Search. the Metadata Database is the OEDC Metadata store, and the Metadata harvester is a type of OEDC metadata aggregator. There are some fundamental differences, in that there is no API to conduct the Data Search, limiting access to manually entered search terms in a web portal rather than automated discovery and monitoring. It should be noted that the CKAN itself provides such functionality, but it is not available in the Icebreaker One implementation.

The Metadata harvester is predicated on the Data Provider providing a suitable API (bottom right of diagram), securing it to the Icebreaker One's Trust framework standard[14]. In practice, it requires Data Providers to provision a suitable API in front of their existing data using an open standard definition (Open API[15]). The Open Energy Architecture also defines the elements of the OEDS Data Sharing Fabric (DSF) in that the Member Directory and Org Database are comparable to the OEDS Governance block, the AuthZ (authorization) service is comparable with the OEDS authentication service (as discussed previously, also includes authorization), and the two defined APIs (Open and Shared). As the Icebreaker One solution utilises the CKAN framework, it is compatible with other CKAN instances comparable to the Federation requirements in the OEDS-defined Catalogue and Data Sharing Fabric.

The EDiT also states, "open data (in class OE-O) MUST NOT require any form of authentication prior to access", which may pose challenges in providing a cost-effective service, ensuring fair access (preventing a minority of users from consuming most resources), and conforming with current cybersecurity best practices.

An example is a misconfigured automated pipeline that could accidentally overload the upstream Data Provider's system with too many requests, degrading the service for other users. Typical identifiers common with human users will be absent for machine-based communication; therefore, isolating the disruptive user or users will be limited to more crude measures such as rate limiting based on IP address. In the Cloud era, this can have unintended consequences by blocking users from the same provider, like blocking all users of a university or a Company where access is behind a common institutional proxy. This is commonly mitigated by the inclusion of an authentication token, even if issued with minimal registration data, as it will disproportionately improve the service's resilience to these types of accidents or attacks.

As further discussion is required to balance the key requirement for Open Data with the additional challenges from a resource consumption and cyber security perspective, the technical requirements will accommodate both sets of needs, allowing for maximum flexibility.

[13] Open Knowledge Foundation (2023) - CKAN
[14] Icebreaker One (2022) - Data Provider Guidelines
[15] The Linux Foundation (2023) - The OpenAPI Initiative

7.2 ONS Energy Data Visibility - June 2020

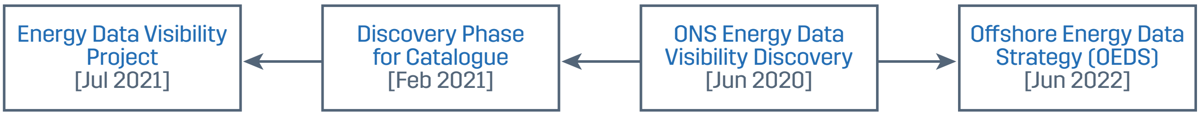


Figure 6: Energy Digitalisation Taskforce reports chain

This segment of the report chain provides five requirements in total (Req. ID 4-9) from two reports: the Office of National Statistics (ONS) Energy Data Visibility Discovery[16] and the follow-on activity called the Energy Data Visibility Project. The Discovery Phase for the Catalogue provides good background material, but the relevant points are better captured by the other two reports.
The ONS EDVD report, compiled by Hippo Digital, largely echoes the findings captured previously in that "We heard from users that today's data is difficult to discover, search, and understand, verifying the findings from the EDTF report". It outlines two (additional) relevant recommendations for OEDA:

- Recommendation 2: Start with a simple set of attributes and evolve to meet a key principle: "We recommend that datasets have a set of characteristics that support consumers in making the decision as to whether they are right for the task" (Req. ID. E4).

- Recommendation 4: Visibility and access: "Data consumers do not view visibility and access as separate concerns" (Req. ID. E2).

There are several quotations from participants that provide an insight into the underlying principles, for example, in support of Recommendation 2: "And if the catalogue in the front had information on a data dictionary or information on what's in the dataset (emphasis added), then that would be really useful." The context is that most existing implementations provide links to other locations with minimal context, and therefore the user must hop to multiple (and contrasting) portals to access the information they need.

The detailed recommendations provide a set of attributes to consider in terms of metadata:

- Description
- Organisation lineage and ownership
- Aggregation method
- Update frequency
- Date last updated
- Confidence/ Quality level
- Format
- Units
- Risks
- Contact
- Keywords
- Licence conditions

[16] Hippo Digital (2020) - Energy Data Visibility Discovery

7.2 ONS Energy Data Visibility - June 2020

In the “Service Evolution” section, the project outlines a roadmap of maturity that starts with a list of links and extends to a “Full Ontology”, where the search is knowledge based and not on attributes. It is the only document in the series examined to outline that basic keywords are fundamentally insufficient but a good starting point:

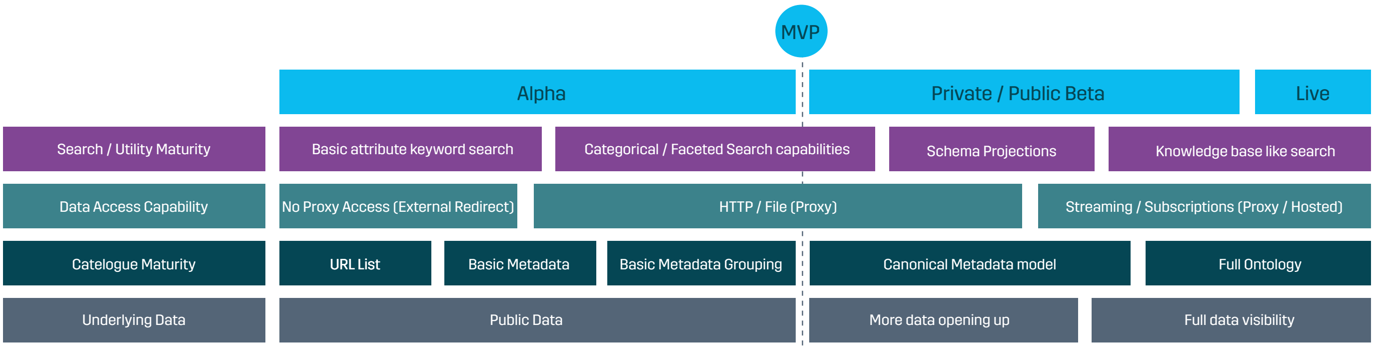


Figure 7: Hippo Digital Roadmap beyond Alpha, p52, ONS Energy Data Visibility Discovery

The “Data Access Capability”, however, is predicated on HyperText Transport Protocol (HTTP) but matures to “Streaming/Subscription (Proxy/Hosted)”:

Catalogue Maturity	Access Capability	Search Capability
URL List	External redirect	Basic search based on provided dataset titles / strings; the utility this provides is that data is centrally visible, but not discoverable, seachable, or understandable in a meaningful way.
Fully attributed datasets and canonical metadata model	HTTP / file proxy	Categorical / faceted search capabilities, which means data is organised into logical categories based on relevant attributes, and users can search based on these relevant attributes.
Fully ontology	Streaming (Proxy / Hosted)	Knowledge-base like search across dataset at-tributes, with well-defined relationships that users can search based on.

Figure 8: Hippo Digital Summary Table, p52, ONS Energy Data Visibility Discovery

7.2 ONS Energy Data Visibility - June 2020

Participants in the study noted that "To this end, we believe that a mere U[niform] R[esource] L[ocator] list without a well-developed set of attributes is unlikely to deliver value for users (emphasis added) since it wouldn't have the appropriate context that users need to truly obtain value from data and would not enable search capabilities of sufficient quality". The key underlying principle is the need for context around the data, which this study has expressed in terms of attributes, but as the figure outlines above, this provides only the most basic support. OEDA should therefore support a range of means to provide context around the data, up to supporting a "Full Ontology", but at its minimum, support a series of attributes.

While HTTP links and subsequently direct access to assets using REpresentational State Transfer (REST) APIs, predominantly based on HTTP, are suitable for smaller payloads, there are limitations in applying this approach to many offshore industrial datasets. The following table outlined by the OEDS shows, for the Renewables sector, the data types broken down by lifecycle stages (see table below).

Much of the data, from forecasts to Metocean assessment data, will be significantly larger than typical reports or some of the data in the onshore sector. For example, Sea Surface Altimetry data over a 7-year period at a one-day resolution can consume over 500 GiB of storage and is unlikely to be accessed using HTTP based APIs (i.e., REST APIs). This may appear to be an extreme outlier, however, for a Machine Learning model for a time series dataset, this only represents 2,555 data points, which is quite low for this type of application given the spatial variability of the data. Larger datasets with greater fidelity will therefore be required, which is not unusual in Climate modelling and measurement.

The key requirements are therefore to support a customisable set of metadata attributes (Req. ID. E4) and the support for external URL redirects, HTTP based APIs, the means to redirect to static files, and other protocols to support streaming applications (Req. ID. E5).

	Application	Construction	Operation	Decommissioning
Renewables	• **Site leasing territory** • **Contract for difference** • **Terms** • **Farm specification** Sizing, distance from shore	• **Onshore electrical installation requirements** Substation, grid connections, voltage options • **Engineering works** Turbine and farm sizing, offshore plant	• **Wind forecasts** • **Load and output forecasts** • **Respose and availability** • **SCUDA** Turbine, balance of plant • **Alarm / Event logs** Outages, curtailment, remote control • **Seabed surveys**	
	• **Metocean assessment** Historical reference datasets, proposed site meteorology, oceanographic behavior, geotechnical • **Site specific modelling outputs** Operational and extreme weather, load / demand studies			

Figure 9: Data landscape map extract from Fig. 5, p15 OEDS report

[17] Copernicus Marine Environment Monitoring Service (2008) via Pangeo

7.3 Energy Data Visibility Project (EDVP) - July 2021

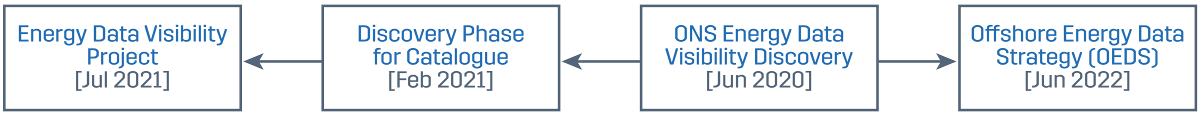


Figure 10: Energy Digitalisation Taskforce reports chain

Following-on from the ONS Energy Data Visibility project, Hippo Digital supported a consortium of companies, including ARUP and most notably Icebreaker One, to extend the initial Visibility activity "to test and validate metadata standards, glossaries and test a prototype solution with the energy community"[18]; this was the basis for the Open Net Zero but itself was based on the Open Energy data catalogue launched in 2019.

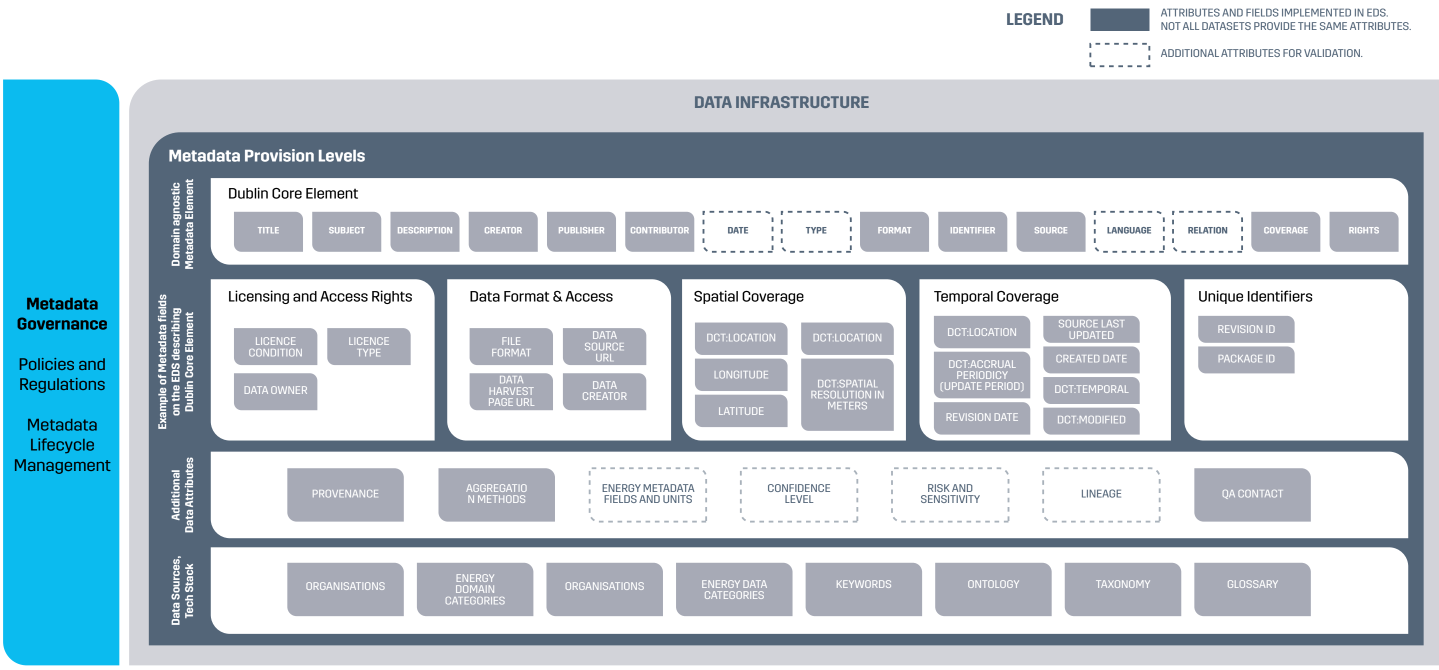


Figure 11: Metadata Governance extract from EDVP Alpha Report

[18] Icebreaker One, ARUP & Hippo Digital (2021) - Energy Data Visibility Project

7.3 Energy Data Visibility Project (EDVP) - July 2021

Most of the report focuses on metadata standards, including worked examples. From a technical requirement perspective, it extends the minimum metadata based on the Dublin Core Element with additional context specific attributes as per the figure above and examples listed below (Req. ID. E4):

- Spatial Coverage
- Spatial Granularity
- Temporal Coverage
- Temporal Granularity
- Provenance •Including a way to surface source datasets and versioning
- Methodology
- Risks / Considerations
- Documentation

The project also identified the following needs:

- Surfacing and measuring quality (Req. ID. E6)
- Data requests (Req. ID. E7)
- Feedback mechanism (Req. ID. E8)
- Surfacing relationships (Req. ID. E9)

The report focuses on subjective measures of quality, which are the purview of the Data Provider and necessary for downstream consumers to consider when utilising the data. The term Data Quality in the Data Industry is often related to other more objective measures, including the presence of invalid, incomplete, inconsistent, and missing data where they are marked within the data itself. An example of Data Quality issue could be where a handful of rows contain text whereas the rest of the dataset consists of integer numbers. This type of quality assessment is conducted using Data Profiling and is addressed as a Data Practitioner expectation in Appendix A.

The data request approach echoes the recommendation from the Energy Data Taskforce (EDTF), which made similar recommendations. Data Providers were seeking early feedback on what datasets to prioritise by following a two-step process. Initial basic descriptions would be shared with the community and based on feedback (votes or comments) prioritise, the dataset would be prioritised for internal assessment prior to release (Req. ID. E7). This is tied in with the recommendation to provide a Feedback Mechanism (Req. ID. E8) to surface Data Quality issues and to also, in effect, vote on what datasets to prioritise for release.

The final key recommendation was to surface relationships between datasets from both a Data Provider perspective in terms of Provenance and Lineage and also from the Community perspective by supporting User-defined relationships (Req. ID. E9). This would aid discovery by downstream users both vertically (lineage) and laterally (linked but not related).

7.4 Energy Data Taskforce (EDTF) - July 2019

Much of the content within this segment of the chain of reports has previously been covered in the ONS Energy Data Visibility Discovery project, which was launched based on a recommendation in the Energy Data Taskforce report itself. There are no new requirements derived from this segment, but additional evidence to support the existing requirements for metadata (Req. ID. E4) and prioritising datasets for release (Req. ID. 7), including a feedback mechanism for Data Providers (Req. ID. E8).

The key additions for OEDA are that the EDTF cites the UK Energy Research Centre (UKERC) Energy Data Centre (EDC) as a good example appearing in the main report and in the appendix related to defining a Data Catalogue. The ERC supports both hosting the data or the links to it, with a flexible approach to the associated metadata as per its guidance[19].

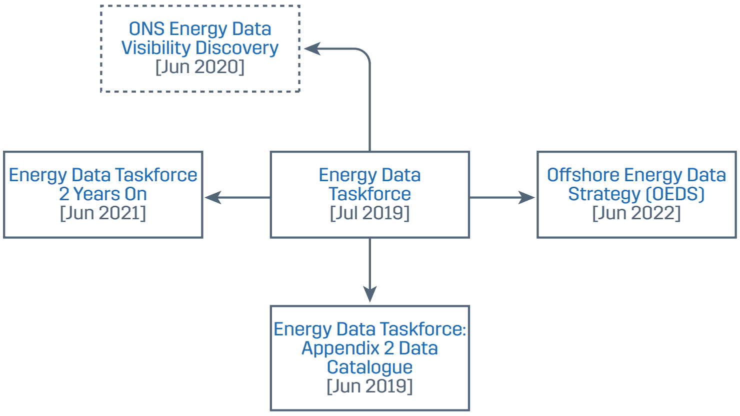


Figure 12: Energy Digitalisation Taskforce reports chain

The example below relates to solar photovoltaic installations[20] in the UK and is referred to as a Linked Dataset, which has its origin in the Nature publication:

The EDC uses the Dublin Core Metadata framework, but the quality and completeness vary between datasets. While its approach provides maximum flexibility in metadata input, this also makes it prone to human error and the usual maintenance responsibility from the Data Provider in manually keeping the entries up to date. The ERC provides the equivalent of Data Search, a Metadata Store, and a mechanism to provide internal Reporting, but there is no evidence of an API or a metadata aggregator. There is also no equivalent to a Data Sharing Fabric or a means to authenticate with the Catalogue and upstream Data Providers simultaneously. Overall, this makes the Energy Data Centre unsuitable for the Offshore Energy Data Catalogue.

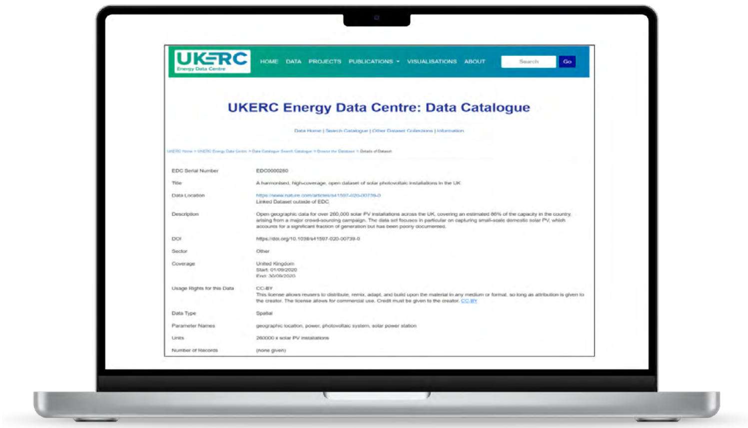


Figure 13: Example Data Catalogue entry in the UK ERC

[20] ERC (2023) - entry EDC0000280
[21] DublinCore (2020) - DCMI Metadata Terms

7.5 Open Subsurface Data Universe (OSDU)

Outside of the onshore energy sector, there is an additional repository referred to in the offshore sector, and that is the Open Subsurface Data Universe(OSDU)[22]. The OSDU is the product of the Open Group[23], which itself consists of several technology companies galvanised around the usage of data. The OSDU mission is to deliver a "standards-based, technology-agnostic data platform for the energy industry that stimulates innovation, industrializes data management, and reduces time to market for new solutions" and identifies three domains:

Platform Provider | External Data Provider | Application Provider

Although its website provides access to a Data Catalogue, it isn't explicitly mentioned within the Technical Specification for OSDU, nor are there any publicly available examples of data sources. As the organisation itself identifies, it defines a standard (for members only) to implement an ecosystem that in effect has strict data definitions[24] (or schemas) with well-defined semantic relationships between entities to create a canonical model or an Ontology. Although the Technical Specification permits the creation of any type of object, the existing templates are focused around Oil and Gas extraction.

The documentation cites a minimum set of attributes for the metadata like that collated for this report based on the wider energy sector reports. From a technical perspective, it defines a framework for how the data should be organised, defined, and utilised. The architecture is significantly more complicated than that proposed for the OEDC or DSF, as the example of the Geospatial Consumption Zone (GCZ) illustrates[25].

Overall, the potential exists for OSDU to satisfy the requirements for OEDA, but the current state of publicly available documentation on the Platform suggests that the overall programme is still in development, and therefore it's not possible to make an effective assessment.

[22] The Open Group (2023) - OSDU
[23] The Open Group (2023)
[24] The Open Group (2023) - OSDU Data Definitions
[25] The Open Group - OSDU Geospatial Consumption Zone Architecture

8.0

Conclusion

A consolidated set of OEDA Technical Requirements in Table 1 of this report has been derived from the assessment of existing data sharing implementations, best practices, and recommendations from the wider energy sector and translated into terms understood within the data industry. In addition to the 10 requirements gathered from a chain of eight energy sector reports from 2019 to 2022, six additional requirements have been proposed based on Data Practitioner expectations.

Any deviation from guidelines and recommendations in previous and current energy sector initiatives is based on the ambition for industry to move beyond a model of exchanging data and information to a model whereby organisations can apply best practice to actively collaborate around trusted data sources using common data toolkits with the required level of functionality to further develop the digital ecosystem in the offshore energy sector.

The technical requirements outlined in this report will be articulated in the subsequent architecture design report and applied in the pilot for the OEDA project.

9.0

Appendix A: Data Practitioner Expectations

The term Data Practitioners is used in this report to cover a wider range of job titles that utilise data to meet the same aims as OEDA in terms of decision-making, automation, and operational improvements, with an emphasis on using the data. There are naturally other stakeholders who would benefit from OEDA, such as auditors keen to ensure compliance or document that such datasets exist. These other roles are ancillary to the primary goal of delivering Net Zero.

The value of the data is derived from its use, and the role of OEDA is to enable effective engagement with such practitioners through data sharing. To enable use cases that support the delivery of net zero through automation, the application of machine learning, and data science leading to the use of artificial intelligence is the key difference between OEDA and the requirements of the other catalogues in the energy sector. The target community, where such skills and expertise exist for such applications, has different expectations from those seeking a catalogue to meet compliance or regulatory requirements.

The following figure illustrates some of the differences in expectations between Data Practitioners and potential users of other Data Catalogues:

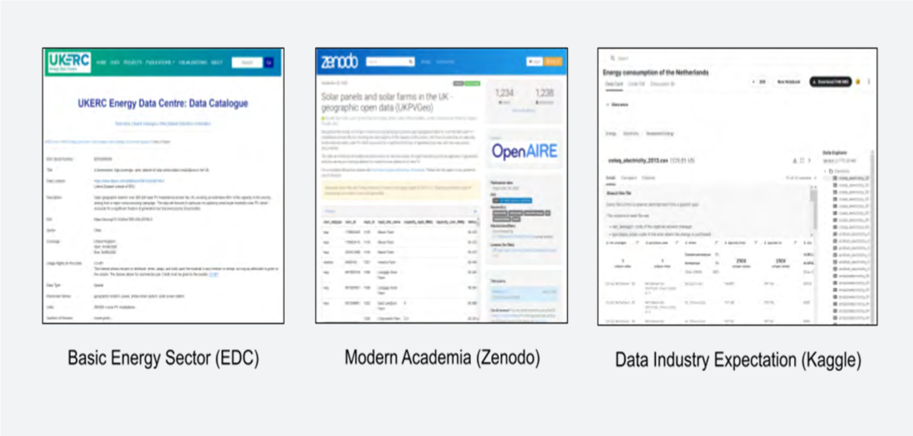


Figure 14: Comparison of Data Catalogue interfaces

The first example is from the Energy Data Centre (cited by EDTF as an example industry-led catalogue) and relates to solar photovoltaic installations[26] in the UK and has been discussed previously. The second example is where the same data is hosted on an academic data platform called Zenodo, and the final example shows another energy sector dataset hosted on Kaggle. Although all three host similar core metadata, the presentation of the latter two is more engaging and highlights additional features that are expected within the data industry.

Accessing the Linked Dataset within the EDC example shows an article from the publication Nature, which in turn has a link to the data on the Zenodo[27] platform:

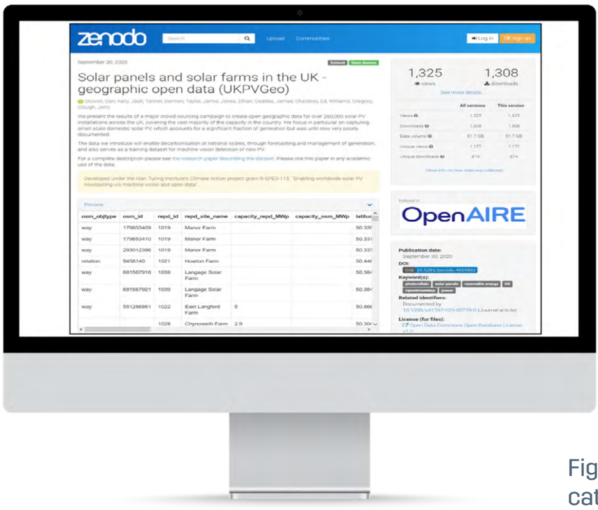


Figure 15: Example data catalogue entry on Zenodo

[26] UKERC Energy Data Centre (2023) - EDC0000280
[27] Stowell et al (2020) - Solar panels and solar farms in the UK - geographic open data (UKPVGeo)

In addition to displaying very similar information in a modern format, the Zenodo platform provides three key features sought by Data Practitioners. Immediately, a preview of the data (not entirely uncommon in the energy sector,- e.g., National Grid ESO[28]) for them to make their own evaluation of the data (Req ID. D3). Outside of the screenshot is a list of assets, their format and critically their size. If the data is in a machine unfriendly format like PDF, or has a custom structure - it is unlikely to be selected for further exploration or engagement. The third feature is the existing engagement of the data (like the OEDC Reporting functionality) in the top right in terms of Views and Downloads (Req. ID E6).

The Kaggle site for a similar dataset[29] has the following format:

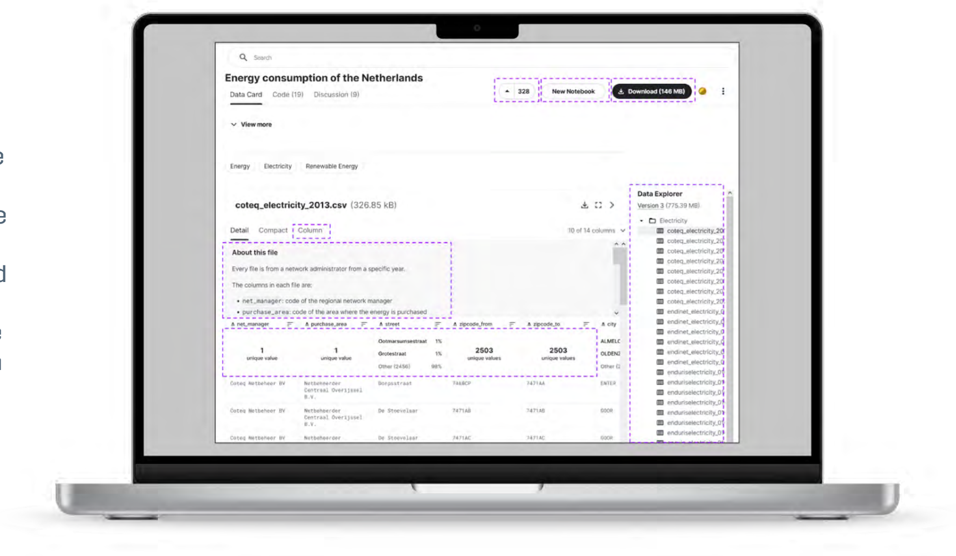


Figure 16: Example Energy related data catalogue entry on Kaggle

The Kaggle layout emphasises access first, and independent of the dataset location (whether it is linked or directly hosted), it provides samples, which are hosted on the platform to accelerate engagement. i.e., users can understand enough about the data and its context (Req. ID. D3 & D6) - due to the rich formatting (Req. ID. D4) support – prior to clicking through to the upstream Data Provider.

At the top, there is an opportunity to provide immediate feedback (Req. ID. E6), much like in Open-Source software development on GitHub, where users "Star" their favourite packages (Req. ID. D1). These are used on GitHub as community engagement metrics for a variety of purposes including for start-ups to receive funding. The feature in Kaggle is intended to provide similar feedback on adopting an Open-Source culture and development environment (Req. ID. D1).

Likewise, there is an opportunity to explore the data (Req. ID. D5) using a computational notebook (New Notebook), followed by an aggregate download link for the impatient. Note that the catalogue is optimised for data consumption and usage, even offering a computational platform for free.

The right-hand side offers a Data Explorer that based on context, allows a user to peruse files hosted on the platform, or if no data is hosted but an API is provided, may contain assets that support the use of the API. Often, even if the data access is via an API, Kaggle hosts samples of the payloads available through the API to reduce the friction for Data Practitioners.

The centre of the image is dominated by rich inline documentation (Req. ID D4), and the "Column" tab shows detailed Data Profile output (Req. ID D3), including but not limited to basic statistics of the data, cardinality, number and popularity of categories, and numerical description (e.g., max, min, median, mode, etc.). Outside of the screenshot, there is an Activity Overview, which shows the page has been viewed 114k times and the data downloaded 108k times (Req. ID E6). The layout, display, and interactivity are key components in making Kaggle an engaging environment to seek and use data.

The contrast between the Energy Data Centre, which may functionally meet most of the requirements of OEDA and the Kaggle environment is significant. These additional features have not been highlighted in the chain of energy sector reports, which are predominantly concerned with releasing data and Data Provider led provision of metadata attributes rather than the stronger focus on the user and functionality of these other platforms.

[28] National Grid ESO (2023) - Balancing Contract Enactment FY 22-23
[29] Luca Basanisi (2020) - Energy consumption of the Netherlands

9.1 Additional Requirements

The observations from Zenodo and Kaggle have been combined to form the Data Practitioner based requirements. The EDiT offered a definition of Open including software, standards, and licences. We propose the inclusion of the development culture and documentation associated with Open-Source software through a system like the git Version Control System (VCS). For example, typically, the Data Provider compiles the documentation and metadata attributes internally before making them available to various catalogues. Adopting an Open-Source culture, would empower them to host such context information in a controlled but public/shared git repository. Much like most Open-Source software, instead of users asking for corrections or amendments, they can edit a copy of the source material and propose to the Data Provider to accept their changes (making a Pull Request).

Users can therefore alleviate some of the burden from Data Providers and enable them to not only directly provide feedback but also propose corrections, engage in making improvements and support a community with the provider. It also enables the Provider to support multiple types of Data Catalogues whilst maintaining their own Single Source of Truth.

- Req. ID. D1: OEDA shall support the use of internal and external repositories for dataset documentation, context, data samples, API definitions and other assets.

Inherent in the structure of Kaggle is the support for data platforms and therefore, Machine to Machine communication using APIs but also long held security tokens or the means to authenticate Machines for direct interaction.

- Req. ID. D2: OEDA shall support the use of long held security tokens including but not limited to client and server-side certificates – mutual Transport Layer Security (mTLS) with Hardware Security Modules (HSM) and/or rotated authentication tokens (i.e., Oauth 2.0 / OIDC).

Kaggle and Zenodo have demonstrated that enabling users to make their own judgements on the data, independent of the provider's use of attributes, is more engaging using Data Profiling (providing relevant insights of the data) and the ability to preview datasets.

- Req. ID. D3: OEDA shall support data profiling for machine readable formats and support the hosting of sample data for user preview.

The Open-Source development culture also provides rich documentation around a project that users can collaborate on, which can also be hosted externally. These are typically in the Markdown format, with extensions that in effect supports a range of features from tables, images, equations, links and more baseline formatting such as bold, italics and underline.

- Req. ID. D4: OEDA shall support rich formatting of content.

Kaggle has demonstrated that users prefer to make their own assessments of the data rather than rely on Data Provider attributes. This includes the principle of the data being Open to Explore, either externally much like the Python Data ecosystem with Binder[30] or internally through hosted Jupyter[31] computational notebooks.

- Req. ID. D5: OEDA shall support the exploration of data with either internal or external platforms.

Most of the current industry data is either relatively small tabular datasets, reports or other content captured in formats such as PDFs. As the OEDS report has demonstrated, it is likely that larger datasets will need to be shared and collaborated with. Current implementations of Data Catalogues have gravitated towards HTTP based links or data access. Whilst they are suited for payloads potentially up to 100 MiB or so, most of the supporting networking ecosystems are not expecting such large payloads and may disrupt access.

Climatology data could be used to inform the design strength of Wind Turbine blades or the placement of tidal generation systems. This type of data is likely to consist of environmental measurements at high granularity across a significant time period. Extracting a subset of the data at a target location is not possible if typical Energy industry storage formats and technologies are used, leading to the whole dataset being loaded into memory prior to extraction of the requested sample.

Given that climate data can consume 300 TiB of memory, it is a significant challenge to load and extract the required data when the largest single compute node from a cloud provider in 2023 Q2 supports only 24 TiB. However, given the right storage technology (e.g., distributed object storage) and file format (e.g., Zarr[32]), a Raspberry Pi with only 4 GiB of memory (or 0.004 TiB, which is less than most modern phones) can sample the relevant data directly due to how the metadata is structured and stored:

[30] The Binder Team (2022) - Binder
[31] Project Jupyter (2023) - JupyterHub
[32] Zarr Developers (2022) - Zarr-Python

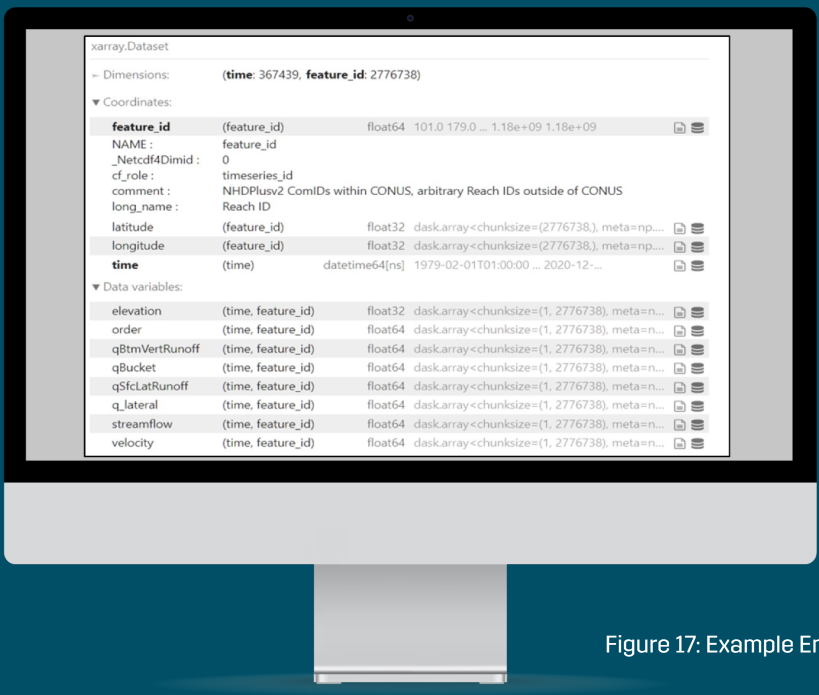


Figure 17: Example Energy Zarr metadata structure

- Req ID. D6: OEDA shall have the means to support a variety of metadata formats (beyond the current attribute-oriented needs).

This list of additional requirements is not exhaustive but provides sufficient coverage to align OEDA with data industry expectations. As these have not been widely discussed previously within the offshore energy sector industry, they are presented as a potential extension to the OEDA Technical Requirements:

Req. ID	Requirement	Source(s)
D1	OEDA shall support the use of internal and external repositories for dataset documentation, context, data samples, API definitions and other assets.	Data Industry expectations around Open Source software development and documentation culture.
D2	OEDA shall support the use of long held security tokens including but not limited to client and server-side certificates - mutual Transport Layer Security (mTLS) with Hardware Security Modules (HSM) and/or rotated authentication tokens (i.e., OAuth 2.0 / OIDC).	Recommendations from wider Energy sector reports are tilted towards Human interaction. The OEDS report explicitly states the use of machine-to-machine interactions. The Data Industry expects the use of standard protocols and approaches.
D3	OEDA shall support data profiling for machine readable formats and support the hosting of sample data for user preview.	Data Industry expectations for data format, structure and size are required prior to previewing the data - particularly important for larger datasets.
D4	OEDA shall support rich formatting of content.	The Open Source development culture also provides rich documentation around a project that users can collaborate on, which can also be hosted externally.
D5	OEDA shall support the exploration of data with either internal or external platforms.	Kaggle has demonstrated that users prefer to make their own assessments of the data rather than rely on Data Provider attributes. This includes the principle of the data being Open to Explore, either externally much like the Python Data ecosystem with Binder or internally through hosted Jupyter computational notebooks.
D6	OEDA shall have the means to support a variety of metadata formats (beyond the current attribute-oriented needs).	Data Industry expectations for data format, structure and size are required prior to previewing the data - particularly important for larger datasets.

Table 2: Proposed Requirements from the Data Industry

Technology Driving Transition

Contact number:
+44 (0)1224 063200

Media enquiries:
pressoffice@netzerotc.com

Net Zero Technology Centre
20 Queens Road, Aberdeen AB15 4ZT

www.netzerotc.com