

Marine Data Exchange: FAIR data and Machine Learning

Chelsea Bradbury, Senior Marine Data and Insights Manager

The Crown Estate

> Expand list

Topic categories: Birds Environmen...

2017-2022, Nova Innovation, Ynys Enlli, Bird and Marine Mammal Survey

Last series update: August 2023 [MEDIN file](#) |

Collection date: Sep 2017 - Sep 2022

Publication date: Aug 2023

Published by: The Crown Estate

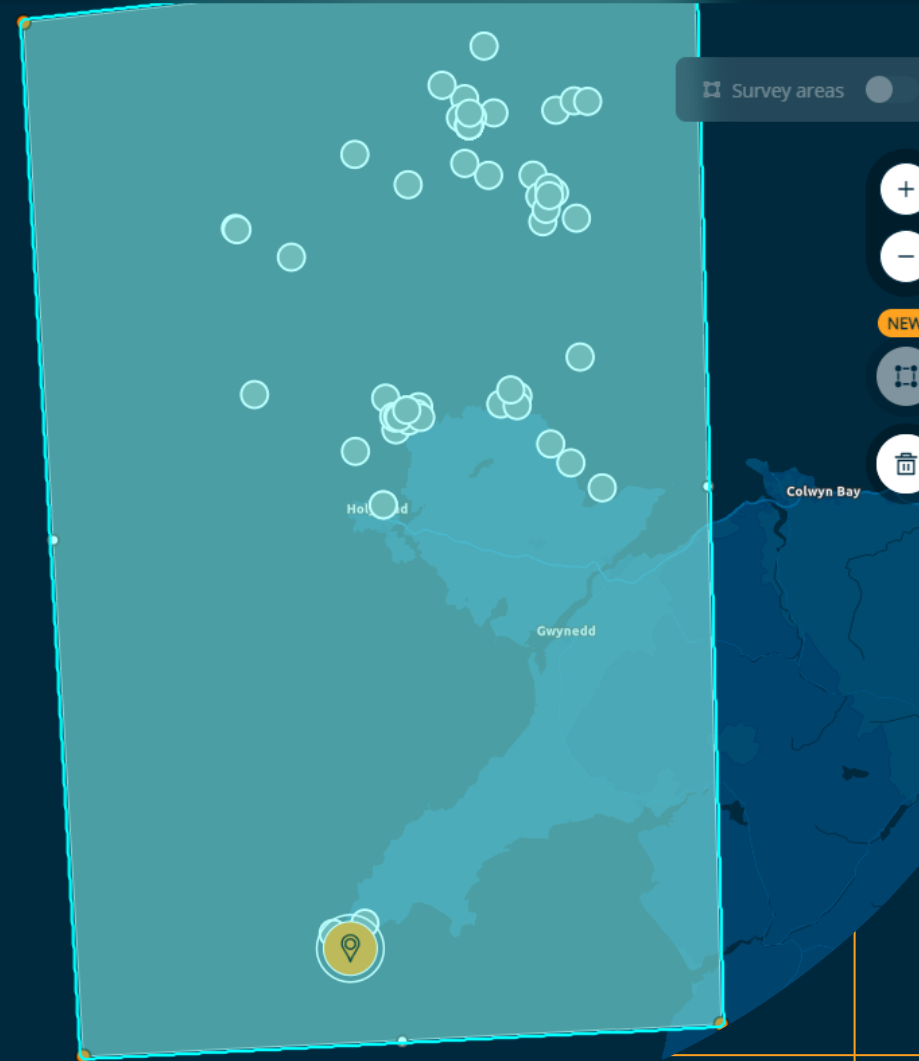
Industry: Tidal

Development: Ynys Enlli

Phase: Pre-construction

Organisation: Nova Innovation

baseline marine
as copied out



What am I going to talk about?

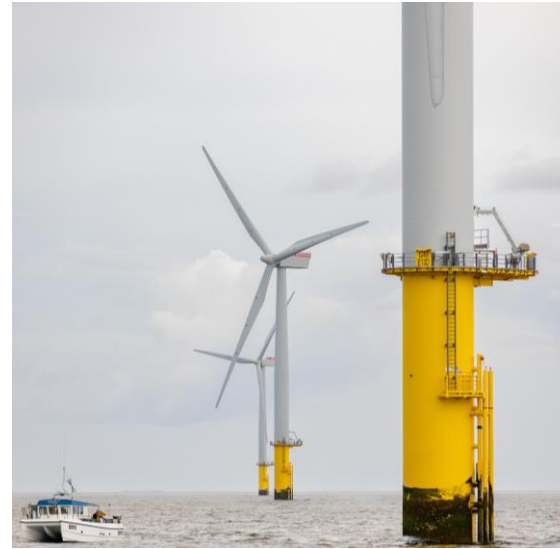
- Introduction to The Crown Estate and the Marine Data Exchange (MDE)
- What Marine Energy Data does the MDE hold?
- FAIR MDE
 - MEDIN guidelines and metadata standards
 - Future
- MDE and Machine Learning
 - Metadata...
 - Improved data transfer
- Future Development

The Crown Estate, the Marine Data Exchange, and marine energy data

Who are The Crown Estate?

The Crown Estate is a unique business with a diverse portfolio that stretches across the country, ranging from central London properties to the **management of the seabed around England, Wales and Northern Ireland**. The Crown Estate was established by an Act of Parliament and **returns 100% of their net revenue profits to the Treasury** for the benefit to the nation.

The Crown Estate is driving **the sustainable development of the seabed**. Our role is central to supporting the delivery of the UK's net zero agenda, including growing offshore wind, supporting the development of carbon capture and hydrogen markets, alongside facilitating the development of other sectors involving the seabed.



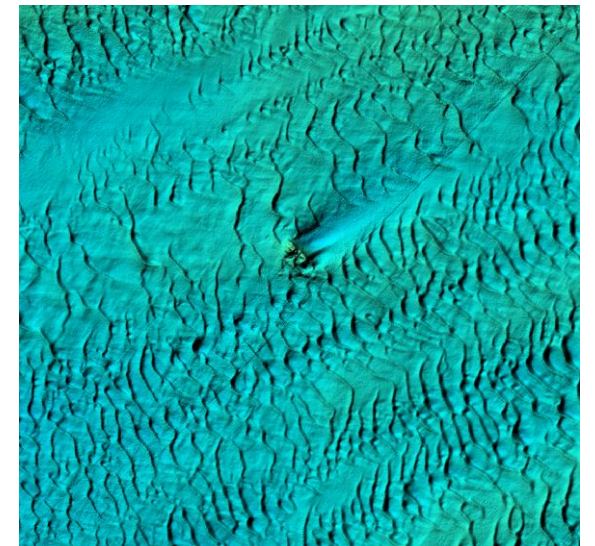
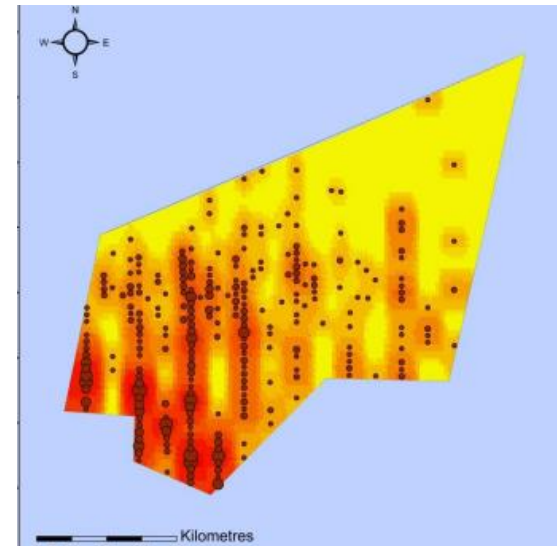
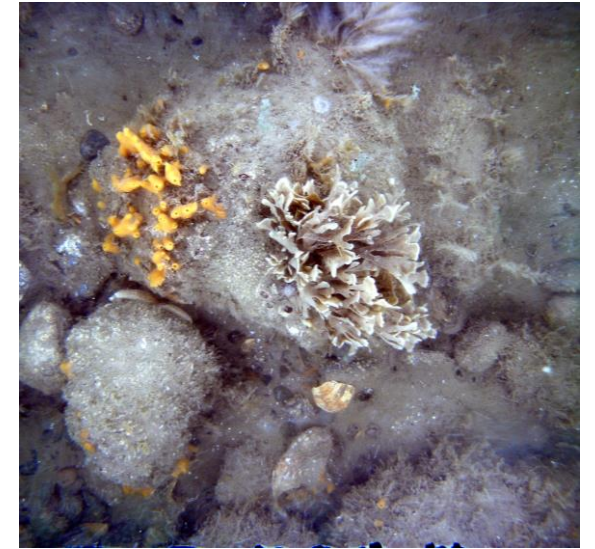
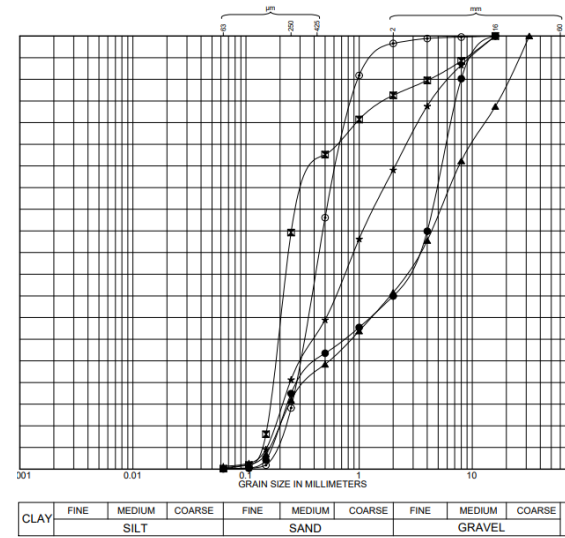
What is the Marine Data Exchange?

The MDE is a data management system used to store and publish industry survey data.

Set up in 2013, it is one of the largest databases of offshore survey industry data in the world (over 250 TB).

It is designed too:

- Ensure **reuse** of industry survey data.
- Allow industry data to support **evidence projects** which can then benefit the sector.
- Aid The Crown Estate's **decision making**.
- Avoid **loss of survey data** in the event of project cancellation.





Marine Energy Data on the MDE

The Marine Data Exchange holds data from the marine energy projects from across the UK, including:

- Seagen Strangford Lough tidal project (pictured)
- Ynys Enlli tidal project
- Morlais tidal demonstration project (including the outputs of the Marine Characterisation Research Project, potentially totalling over 80TB of data)

It also holds some early data from Scottish wave and tidal projects, and the outputs of the Wave and Tidal Enabling actions work.

Majority of marine energy projects in the UK have been tidal, with some wave projects.

FAIR MDE

Where are we now?

- As a sponsor of the UK's Marine Environmental Data and Information Network (MEDIN), The Crown Estate utilises a number of their products.
- The Crown Estate requires all data to have associated MEDIN discovery metadata (schema based on the ISO 19115:2003 standard), making MDE series **discoverable** via the MEDIN portal - a UK marine metadata portal.
- As part of our Quality Assurance (QA) process, The Crown Estate also implements MEDIN data guidelines (a set of MEDIN endorsed archiving standards for different data collection types), ensuring data that has been provided is as **interoperable** and **reusable** as possible.



MEDIN discovery metadata standard



MEDIN data guidelines

How is MDE improving the 'FAIR-ness' of our data holding?

Currently:

- Pushing our data into subject specific data portals which utilise standardised formats (OneBenthic, JCDP) – **Findability/Interoperability**
- Metadata cleaning campaign – **Findability/Accessibility**

Over next year:

- Clarifying and streamlining our terms of use – **Accessibility**
- Streamlining of the Metadata creation process... including using machine learning techniques - **Findability**

2022, ORE Catapult, Ynys Enlli , Analysis of Tidal Energy Resource

Powered by:
**THE CROWN
ESTATE**

Don't show this x

What is a series page?

OVERVIEW

Download MEDIN file | ↓

Last series update: August 2023

Collection date: November 2022 - November 2022 Categories: Engineering Design Metocean

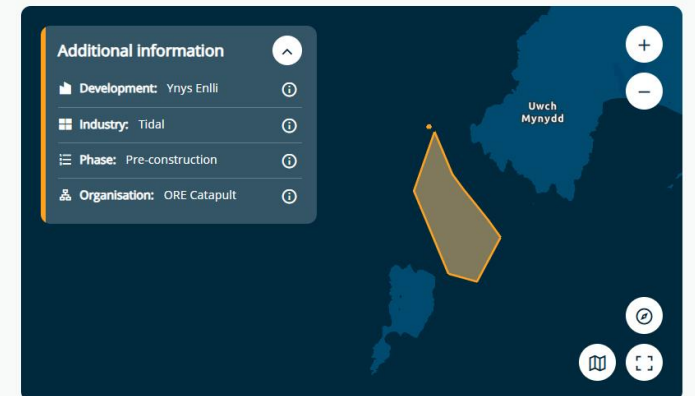
Publication date: August 2023

Published by: [The Crown Estate](#)

Summary

Report providing detailed analyses of data gathered from an Acoustic Current Doppler Profiler (ADCP) deployment at the north end of Bardsey Sound. The ADCP was deployed by Bangor University as part of their SEACAMS2 research programme. Data analysis was carried out by the Offshore Renewable Energy Catapult for Nova. Series includes the link to the ...

[Read more](#) v



MDE: Machine Learning Applications

```

<gmd:identificationInfo>
  <gmd:MD_DataIdentification>
    <gmd:citation>
      <gmd:CI_Citation>
        <gmd:title>
          <gco:CharacterString>2012, EMEC, Shapinsay Sound, TRIAXYS wave data </gco:Char
          </gmd:title>
        <gmd:date>
          <gmd:CI_Date>
            <gmd:date>
              <gco:Date>2012-12-03</gco:Date>
            </gmd:date>
            <gmd:dateType>
              <gmd:CI_DateTypeCode codeList="http://standards.iso.org/ittf/PubliclyAvail
              codeListValue="publication">publication</gmd:CI_DateTypeCode>
            </gmd:dateType>
          </gmd:CI_Date>
        </gmd:date>
        <gmd:identifier>
          <gmd:RS_Identifier>
            <gmd:code>
              <gco:CharacterString>SHAP-AXYS-2012</gco:CharacterString>
            </gmd:code>
          </gmd:RS_Identifier>
        </gmd:identifier>
      </gmd:CI_Citation>
    </gmd:citation>
    <gmd:abstract>
      <gco:CharacterString>Data collected by TRIAXYS buoy at EMEC's Shapinsay tidal test
    </gmd:abstract>
    <gmd:pointOfContact>
      <gmd:CI_ResponsibleParty>
        <gmd:organisationName>
          <gco:CharacterString>European Marine Energy Centre (EMEC) Ltd</gco:CharacterSt
          </gmd:organisationName>
        <gmd:positionName>
          <gco:CharacterString>Hydrodynamicist</gco:CharacterString>
        </gmd:positionName>
        <gmd:contactInfo>
          <gmd:CI_Contact>
            <gmd:address>
              <gmd:CI_Address>
                <gmd:electronicMailAddress>
                  <gco:CharacterString>info@emec.org.uk</gco:CharacterString>
                </gmd:electronicMailAddress>
              </gmd:CI_Address>
            </gmd:address>
          </gmd:CI_Contact>
        </gmd:contactInfo>
        <gmd:role>
          <gmd:CI_RoleCode codeList="http://standards.iso.org/ittf/PubliclyAvailableStan
          codeListValue="originator">originator</gmd:CI_RoleCode>
        </gmd:role>
      </gmd:CI_ResponsibleParty>
    </gmd:pointOfContact>
  </gmd:MD_DataIdentification>
</gmd:identificationInfo>

```

Metadata Autofill

Problem

We have internal 'MDE metadata' and MEDIN Discovery metadata XMLs (which are shared to MEDIN portal to allow discovery) associated with our series. This frequently leads to having 'two sources of the truth regarding metadata.

Solution

Generate MEDIN XMLs based off our internal metadata, augmented by adding INSPIRE fields, and do this at the last possible point before publication.

This solution will rely on document cracking and machine learning to minimise the effort required from the creator, aiming to be able autofill as many of these fields as possible.

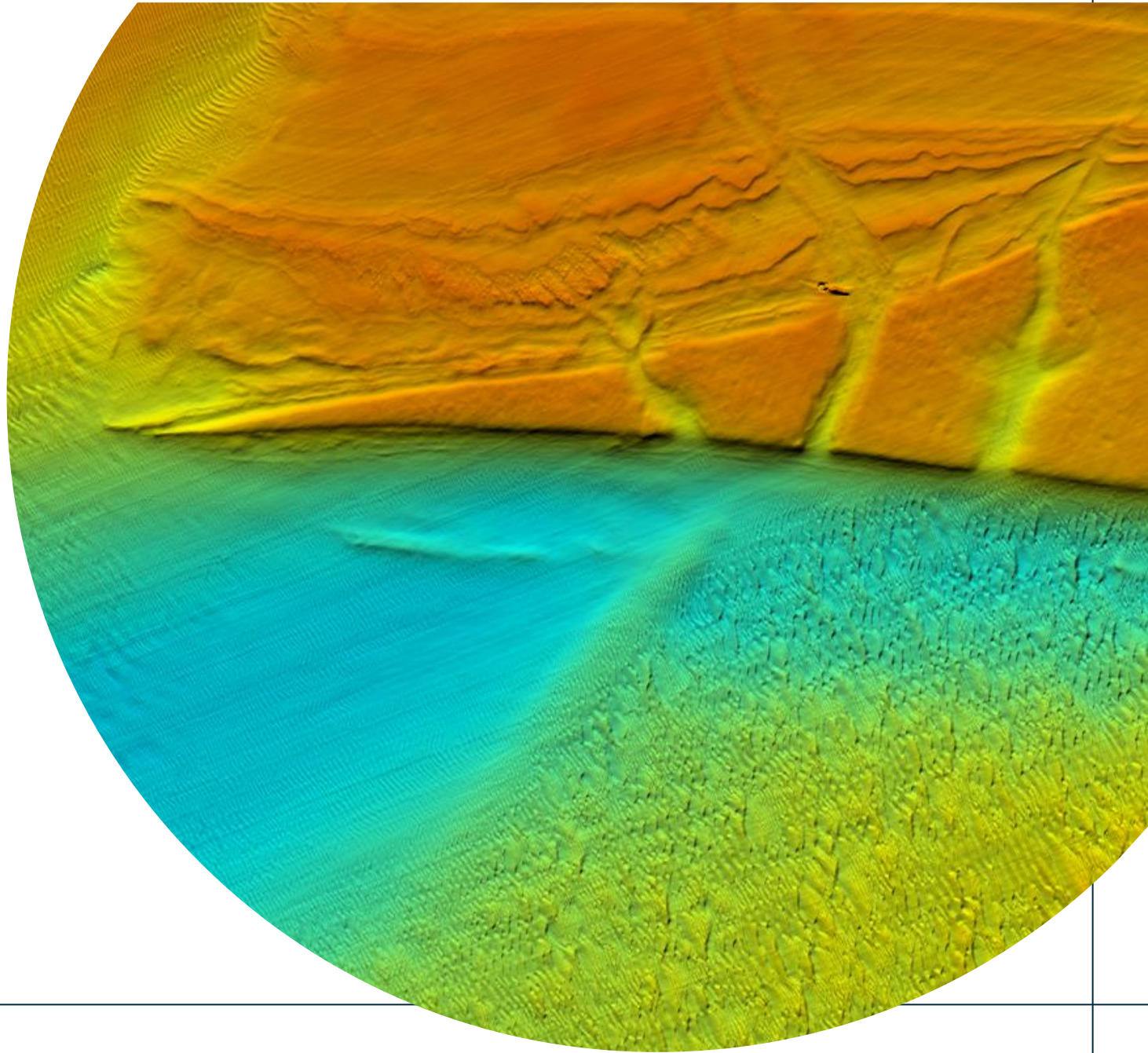
This approach will reduce errors in metadata creation, improve the discoverability of historical records by improving their accuracy, and mean we have one source of the truth for MDE metadata.

Automated data delivery sorting

MDE is built off Microsoft Azure, and we use blob containers for large deliveries of data.

We often receive significant quantities geophysical and other data types, in complicated file structures.

We want to utilise machine learning techniques to sort and identify where this data should be delivered to, to ensure it remains well-structured and reduce human error.



The future..?

- Utilising machine learning techniques to optimise processes and improve the FAIR-ness of the MDE data will continue.
- Using AI techniques such as LLM to interrogate the reports and data we have is increasingly a focus for the MDEs development, as is building 'chatMDE'... Stay tuned for updates!
- If you'd like to sign up to the MDE newsletter, and hear about how the MDE is developing and other MDE news, scan the QR code on screen.

