MEDIN DAC WG Meeting 9

13:30-17:00 Tuesday 20th November 2018, Historic Environment Scotland, Edinburgh

|  |  |  |  |
| --- | --- | --- | --- |
| **Attending** | | | |
| Jon Parr – Chair (JP) | MBA | Peter McKeague (PMcK) | HES |
| James Ayliffe (JA) | BODC | Charlotte Miskin-Hymas (CMH) – minutes | MEDIN |
| Matt Arnold (MA) | DASSH | Robin McCandliss – (RM) | MEDIN |
| Chelsea Bradbury (CB) | The Crown Estate | Mary Mowat (MM) | BGS |
| Alasdair Curran (AC) | UKHO | Clare Postlethwaite (CP) | MEDIN |
| Gareth Edwards (GE) | RCAHMW | Jens Rasmussen (JR) | MSS |
| Katie Green (KG) | ADS | Oliver Williams (OW) | Cefas |
| Laura Hanley (LH) | Cefas | Helen Wojcik (HW) | UKHO |
| Anna Luff (AL) | DASSH | Roseanna Wright (RW) | MEDIN |
| Stuart MacDonald (SMacD) | HES | Peter Thijsse (PT) by Skype | MARIS |
| **Apologies** | | | |
| Callum Stone | Met Office |  |  |

1. **Round table introductions**

Welcome to new members Katie Green (ADS), Helen Wojcik and Alasdair Curran (UKHO), Laura Hanley (Cefas). Stuart MacDonald (HES) also attended the meeting as it was in HES and it helped to introduce him to MEDIN.

1. **Review actions last meeting**

**Table 1:** Actions from DAC WG meeting on 25th April 2018

|  |  |  |
| --- | --- | --- |
| **Action** | | **Progress** |
| 8.1 | **Louisa Matthews** to circulate a copy of the photogrammetry interface report to the DACs, as they may be interested in the solution for high volume data types. | Done |
| 8.2 | **Louisa Matthews** to send Dan Lear the ADS wiki process for streaming video files. | Done |
| 8.3 | **Clare Postlethwaite** to feedback to MEDIN Executive that opening the codefest out to a wider audience is not possible with the available budget, but there would be benefit in holding an internal one to share expertise. | Done |
| 8.4 | **Clare Postlethwaite** to clarify what needs to be done to prepare UKDMOS data for hackathon use and to find out what would be useful for users. | Done |
| 8.5 | **Charlotte Miskin-Hymas** to work with DL, JR, GE and MARIS and coordinate the MEDIN hackathon, to happen in Q3 alongside the mega meeting. | Hackathon postponed to Q4 |
| 8.6 | **Charlotte Miskin-Hymas** to contact Peter McKeague for interesting shipwreck images for the website. | Done |
| 8.7 | **Charlotte Miskin-Hymas** to send link to website to DACs and DACs to feedback comments on the text for their DAC pages. | Closed |
| 8.8 | **Robin McCandliss** to include checking DAC webpages as a standing item on the DAC agenda to ensure the text is up to date and remind DACs to check before the DAC meetings. | Done |
| 8.9 | **Clare Postlethwaite** to organise a meeting to discuss archiving MCZ data at the MEDIN DACs for 2018-19. Including Cefas, BGS, DASSH, and possibly UKHO and JNCC. | Done – meeting held in July 2018 |
| 8.10 | **Gaynor Evans** to reiterate with MARIS the need to fix the bugs and to provide a prioritised list. | Closed – on agenda |
| 8.11 | **Gaynor Evans** to circulate the metadata validation errors for each provider to the DACs | Closed – on agenda |
| 8.12 | **Gaynor Evans** to check whether UUIDs have changed and report back at next DAC WG meeting. | Ongoing – CP to follow up with Gaynor Evans |
| **Actions outstanding from earlier DAC WG meetings** | | |
| 7.9 | **Louisa Matthews** to discuss OASIS metadata validation issue with Gaynor Evans and Peter Thijsse | Ongoing: Remapping is needed to MEDIN metadata schema. Issue with getting developer resource and is not likely to happen soon. |
| 7.11 | **Jens Rasmussen** to validate metadata records using both Metadata Maestro and MARIS tool and report back to DAC WG at teleconference in early 2018. | Ongoing: Jens re-exported and validated 180 xml records through the Maestro, however all records failed validation in the MARIS tool. Bugs in the MARIS tool were the likely cause. Not yet tried to run it through again. |
| 6.1 | **Clare Postlethwaite** to ensure UKDMOS/MEDIN document is made available on MEDIN website once finalised. | Not yet done |
| 6.6 | **Mary Mowat** to consult with UKHO about where resource locators in the MEDIN catalogue should point to for data that are held both at BGS and UKHO. | Done. Mary reported discussion with UKHO, adopting a generic option to make it clear that if a user wants backscatter data, they should contact BGS. |
| 6.13 | **Mary Mowat** and **Chelsea Bradbury** to liaise to identify data sets that are present via TCE and BGS and ensure that it does not appear in MEDIN catalogue twice. | Ongoing. The gaps are now identified. Agreed TCE should hold the data from the aggregates industry. Need now to ascertain where existing data are held and where they should be. |
| 5.1 | **Amended action: Clare Postlethwaite** to a) liaise with Data Task Group to finalise MSFD indicator assessment metadata template, which will require a link to a MEDIN metadata record and a DOI for each dataset used in the assessment; b) set realistic expectations for data archive at the DACs to minimise the risk that following the monitoring programmes update, the DACs will be expected to archive the data with little notice. | a) Done. Template completed.  b) Closed – CP has been working on this with the Data Task Group. |
| 2.18 | **Clare Postlethwaite** to speak to JNCC about possibility of them becoming a Bird DAC. | Ongoing |

**3. DAC network status**

3(a) Brief report from DACs

The DACs gave a verbal report on progress. Items of note:

* ADS – ForSeaDiscovery archive released, increase in data deposits
* BGS – New version of Offshore GeoIndex available
* BODC – 3 years of MSFD data have been pushed to EMODnet, 3000 new parameter codes for contaminants data linking ICES and BODC vocabularies
* Cefas – Building work continues, focusing on scanning paperwork and streamlining physical records
* DASSH – SAHFOS has joined MBA and DASSH now manages CPR data
* HES – Working on reconciliation of Canmore records with wreck locations, metadata cleaning
* Marine Scotland – Scanning old vessel logbooks, archiving routine surveys
* RCAHMW – All wreck data made available on Lle (Welsh government Geo portal)
* TCE – Resolving issue around metadata harvest to MEDIN (resource field not showing correctly)
* UKHO – Developing a new portal and DAC interface, expect it to go live in December. CP and MMOW volunteered to act as DAC WG testers

**4. DAC Work Plan 2018/2019**

4(a) Progress with DAC Work Plan

RM went through the DAC work plan for 2018/19 (Paper P3). Everything planned for Q1 and Q2 has been achieved, apart from progressing the NI FishDAC accreditation – this has been delayed as it is with AFBI’s lawyers. RM will follow up to see if there has been any further progress.

For Q3 tasks:

* The codefest has been postponed to Q4.
* The task to identify MSFD datasets in the DAC network is with the Data Task Group.
* Portal metadata validation is on track.

4(b) Development of the next MEDIN 5 year Business Plan

A first draft of the new MEDIN Business Plan went to the Executive and the Sponsors Board recently and they were generally happy with the content.

The business plan contains 12 High Level Objectives (HLOs). There was a brief discussion to identify the HLOs that were directly relevant to the DACs.

|  |  |  |
| --- | --- | --- |
| **No.** | **HLO** | **Workstream involvement** |
| 1 | Provide efficient access to arguably the most comprehensive coverage of high quality marine data held by UK organisations via the MEDIN portal and DAC network, thereby enabling wide-ranging and dynamic access to UK marine data, UK marine Reference data, view and download services and other data products in line with user requirements. | WS1 lead |
| 2 | Facilitate frictionless, secure, cost effective, long-term archival of marine data within the MEDIN DAC network. | WS1 lead |
| 3 | Develop and provide an open access suite of easy to use, forward looking data management tools and standards (e.g. data guidelines, Application Programming Interface (API) service guidelines), fostering community sharing of expertise and code, to ensure more efficient handling of all marine data being generated now and in the future. | WS3 lead, WS1 involved |
| 4 | Lead and build a community of engaged users of marine data, fostering good data management practice. | WS7 lead |
| 5 | Facilitate domestic and international assessments of the marine environment through the storage, management and dissemination of UK monitoring data and provision of advice on retrieval and amalgamation of large, multidisciplinary datasets. | WS3 and WS1 |
| 6 | Support the UK marine sector to provide globally interoperable marine data services e.g. machine readable Application Programming Interface (API) standard for DACs and others. | WS2 and WS1 |
| 7 | Achieve full adoption of the MEDIN framework (DAC network, standards, metadata portal, data guidelines, data clause) across UK organisations’ (public and private sector), promoting open data principles to ensure widespread data accessibility. | WS7 lead |
| 8 | Lead the UK marine community in data management for new and emerging technologies, including autonomous and robotic systems, satellite observations, and operational modelling and predictions. | Advice from WS1 |
| 9 | Record and provide access to information on monitoring activities in the UK to support the UK Marine Monitoring and Assessment Strategy (UKMMAS) Community. | WS3 lead |
| 10 | Enhance skills within the marine data community, providing training and education. | WS2 lead |
| 11 | Develop a coordinated approach for users to retrieve marine data distributed throughout the MEDIN DAC network. | WS1 lead |
| 12 | Integrate, coordinate and influence international and cross-domain data initiatives, promoting good data management practice. | WS4 lead, WS1 input |

5**. Promoting the DAC Network**

5(a) DAC web pages

DACs were asked to check their webpages for currency of content before the meeting. Some DACs have edits that need to be made, which have been fed back to CMH.

5(b) Using the Portal for promotion

CP asked the DACs whether they use the portal for promotion purposes. Cefas and BGS have done. CP asked the DACs to let her know if there are things that would make the promotion of the portal easier.

**Action 9.1: DACs to send to Clare Postlethwaite (by end Feb 2019) items that DACs would like to see on the portal that would make it easier to use for promotion of DACs.**

5(c) Linking with other WS7 communications

CMH gave a presentation on the communications methods used to promote and raise awareness of MEDIN. The DACs fed back that the new website is a great improvement. If DACs have things that they would like CMH to disseminate, then please contact her.

**6. Coordinated data archiving**

6(a) Archiving data already published elsewhere

James Ayliffe raised this item for discussion. For example, Marine Scotland publishes Scottish monitoring data on its portal with a DOI for the whole dataset. Part of this data set comprises temperature and salinity time series, which are made available to BODC for ingestion. The provenance of the data in BODC does not relate back to the Marine Scotland DOI.

There are three existing scenarios:

* Data in a data centre is pushed to an international data centre.
* Data held in data centre and is referred to by another data centre (e.g. SeaDataNet).
* A package of data is held in a data centre and elements of the data package are archived at other DACs. For example, MCZ data – Cefas publishes the whole package. Elements of the data set go to other DACs and they have a discovery metadata record for their specific holding.

The issue is around provenance of the data and machine to machine use. Potentially there could be a scenario where duplicate versions of the data are used for the same purpose with the user thinking they are separate datasets as they come from different archives. It is valid to have duplicated data as long as the UUID identifier is present and preserved across the different instances of the data.

**Action 9.2: James Ayliff to collate examples from across the DACs where duplication of data is an issue, to bring to next DAC WG meeting for further discussion.**

6(b) Check if there is an update on pilot of coordinated approach to archiving multidisciplinary datasets (DASSH)

DASSH have had no enquiries to deposit multi-disciplinary data as yet.

**7. DAC Data Access Services**

7(a) QC of metadata for new validation requirements

Peter Thijsse joined the meeting by Skype.

When the new portal was released, the restrictions on metadata validation were relaxed in order to give the DACs time to address validation errors. Many of the errors have now been addressed and we are looking to progressively reinstate restrictions. If there are still false positive errors, the DACs need to let Peter Thijsse know.

There are some records which are not spatial and therefore cannot populate the mandatory spatial reference field. This affects a small number of Cefas records. DASSH and TCE have a lot of third party records, which will be harder to correct. There are some instances of metadata being validated on Maestro, but not appearing on the portal.

**Action 9.3: Chelsea Bradbury to discuss with Peter Thijsse issue around MDE records being validated on maestro but not fully appearing on the portal.**

**Action 9.4: Katie Green to follow up ADS metadata harvesting issue with Peter Thijsse – some ADS records are not being harvested because a mandatory field is null, when they should still harvest.**

It was agreed to:

* Reinstate the filter for metadata owner (all records should have an owner).
* Leave the other filters off for the time being and review again at the next DAC WG meeting.

**Action 9.5: Peter Thijsse to reinstate the metadata owner field filter.**

**Action 9.6: Clare Postlethwaite to identify critical fields to test for metadata validation and to identify how many records would fail at present.**

7(b) Plan for next Codefest

The Codefest has been postponed from Q3 to Q4. The intention was to use monitoring observations, e.g. UKDMOS. Dan Lear has also suggested aligning it with biodiversity workflows.

There was a suggestion to have a workshop on biodiversity workflows followed by a hackathon using UKDMOS data. There is a need to pin down a date in Q4.

**Action 9.7: Jon Parr and Matt Arnold to discuss pinning down a date for hackathon with Dan Lear.**

7(c) Megameeting

CP asked the DACs whether there is anything that should be raised at the standards working group meeting the following day (21st Nov).

* Jens Rasmussen mentioned that at IMDIS there was discussion about annotation aligned with video and imagery and the lack of a standard associated with it.
* Consider how DACs should feed into update of standards.
* DACs need to highlight where standards don’t fit the data or become out of date.

**9. AOB**

9(a) Date of next meeting

RM will poll for meeting dates in April/May 2019.

**ACTION LIST**

**New actions from DAC WG meeting 20th November 2018**

|  |  |  |
| --- | --- | --- |
| **Action** | | **Progress** |
| 9.1 | DACs to send to Clare Postlethwaite a list of things that DACs would like to see on the portal that would make it easier to use for promotion of DACs. |  |
| 9.2 | James Ayliff to collate examples from across the DACs where duplication of data is an issue, to bring to next DAC WG meeting for further discussion. |  |
| 9.3 | Chelsea Bradbury to discuss with Peter Thijsse issue around MDE records being validated on maestro but not fully appearing on the portal. |  |
| 9.4 | Katie Green to follow up ADS metadata harvesting issue with Peter Thijsse – some ADS records are not being harvested because a mandatory field is null, when they should still harvest. |  |
| 9.5 | Peter Thijsse to reinstate the metadata owner field filter. |  |
| 9.6 | Clare Postlethwaite to identify critical fields to test for metadata validation and to identify how many records would fail at present. |  |
| 9.7 | Jon Parr and Matt Arnold to discuss pinning down a date for hackathon with Dan Lear. | Complete: The Hackathon is scheduled for the 7th March.  To run alongside the HBDSEG Biodiversity Data Flow meeting. |
| **Actions outstanding from earlier DAC WG meetings** | |  |
| 8.5 | **Charlotte Miskin-Hymas** to work with DL, JR, GE and MARIS and coordinate the MEDIN hackathon, to happen in Q3 alongside the mega meeting. | Closed. |
| 8.12 | **Gaynor Evans** to check whether UUIDs have changed and report back at next DAC WG meeting. | Ongoing – CP to follow up with Gaynor Evans |
| 7.9 | **Louisa Matthews** to discuss OASIS metadata validation issue with Gaynor Evans and Peter Thijsse | Ongoing: Remapping is needed to MEDIN metadata schema. Issue with getting developer resource and is not likely to happen soon. |
| 7.11 | **Jens Rasmussen** to validate metadata records using both Metadata Maestro and MARIS tool and report back to DAC WG at teleconference in early 2018. | Ongoing: Jens re-exported and validated 180 xml records through the Maestro, however all records failed validation in the MARIS tool. Bugs in the MARIS tool were the likely cause. Not yet tried to run it through again. |
| 6.1 | **Clare Postlethwaite** to ensure UKDMOS/MEDIN document is made available on MEDIN website once finalised. | Not yet done |
| 6.13 | **Mary Mowat** and **Chelsea Bradbury** to liaise to identify data sets that are present via TCE and BGS and ensure that it does not appear in MEDIN catalogue twice. | Ongoing. The gaps are now identified. Agreed TCE should hold the data from the aggregates industry. Need now to ascertain where existing data are held and where they should be. |
| 2.18 | **Clare Postlethwaite** to speak to JNCC about possibility of them becoming a Bird DAC. | Ongoing. |