MEDIN DAC WG Meeting 12

11:00-15:30 Wednesday 11th June 2020, Zoom meeting

|  |
| --- |
| **Attending** |
| Jon Parr – Chair (JP) | MBA | Clare Postlethwaite (CP) | MEDIN |
| Caitlin Allan (CA) | Cefas | Jens Rasmussen (JR) | MSS |
| James Ayliffe (JA) | BODC | Andy Richardson (AR) | UKHO |
| Chelsea Bradbury (CB) | The Crown Estate | Callum Stone (CS) | Met Office |
| Gareth Edwards (GE) | RCAHMW | Kay Thorne (KT) | BODC |
| Tim Evans (TE) | ADS | Oliver Williams (OW) | Cefas |
| Laura Hanley | Cefas | Helen Wojcik (HW) | UKHO |
| Dan Lear | DASSH | Robin McCandliss – (RM) | MEDIN |
| Mary Mowat (MM) | BGS |  |  |
| **Apologies** |
| Peter McKeague (PMcK) | HES |  |  |

1. **Round table introductions**
2. **Review actions last meeting**

Minutes from previous meeting (4th Dec 2019) were accepted without amendment.

**Table 1:** Actions from DAC WG meeting on 11th December 2019

|  |  |
| --- | --- |
| **Action** | **Progress** |
| 11.1 | Robin McCandliss to circulate DAC work plan roadmap, highlighting activities at risk. | Done |
| 11.2 | James Ayliffe to send Jens information on the BODC EDMO SPARQL endpoint. | Done |
| 11.3 | Helen Wojcik to send CP information on GSC provenance work. | Done – follow up contact made with Nick Moore |
| 11.4 | Clare Postlethwaite to feed back to MEDIN Executive Team that the £11K funding is enough to maintain basic involvement with MEDIN but additional funds should be allocated when there is a big change (e.g. to standard etc). | Done |
| 11.5 | Clare Postlethwaite to amend KPIs to take into account comments from DACs before they are shared with Executive Team. | Done |
| 11.6 | Robin McCandliss to include in metrics from portal - number of records from MSCC organisations per DAC in total and in past year, and look at pulling out country information for portal records. | Done |
| 11.7 | Robin McCandliss to adjust annual report template to account for MSCC and country metrics, and to include international meetings section. | Done |
| 11.8 | Robin McCandliss to send DACs the finalised DAC annual report template before end of January. | Done |
| 11.9 | Robin McCandliss to liaise with HES and UKHO (as these organisations are due for MEDIN reaccreditation in 2020) to start the accreditation process with Core Trust Seal (if not already started) and support these organisations with the process.  | Closed – on agenda |
| 11.10 | Clare Postlethwaite to liaise with Core Trust Seal to establish whether MEDIN will pay the umbrella fee for accrediting MEDIN DACs.  | Closed – on agenda |
| 11.11 | Clare Postlethwaite to check with Gaynor Evans what the process is for removing metadata records and feedback to the group.Update: Records are removed from the portal on Sundays when there is a complete harvest and reingest of all metadata from suppliers. On a daily basis, new or amended records are harvested and published onto the portal at 6:00 and 13:00.  | Done |
| **Actions outstanding from earlier DAC WG meetings** |  |
| 10.7 | Robin McCandliss/Jon Parr to ask MEDIN Executive whether to drop reporting on INSPIRE compliance. | Closed, moving to OGC compliance reporting |
| 9.3 | Chelsea Bradbury to discuss with Peter Thijsse issue around MDE records being validated on maestro but not fully appearing on the portal.Update: TCE had added some errant code. GE is waiting for them to demo a change to see if my theory was correct. | Closed. Issue was identified and awaiting confirmation from GE that update has worked |
| 7.9 | **Amended action:** Gaynor Evans to follow up with Tim Evans the development of WAF for harvesting ADS records to the portal | Closed – TE has agreed workflow with GE |

**3. DAC network status**

3(a) Size of DAC network

i) Bird, Underwater Noise, Socio-economic DAC

**Seabirds:** CP approached JNCC during 2019 to see if they would consider being the MEDIN DAC for seabird data, as they hold a national database. JNCC are currently updating their systems and were not able to commit the resource that would be required. At the last MEDIN Executive meeting, JR raised that Crown Estate Scotland were looking for an archive for their bird data. DASSH have confirmed that they can provide an archive for seabird data straight away; as a species and biodiversity DAC they are set up to do this. If the data are provided in standard format then DASSH can take them free of charge. DASSH will provide the long term archive for seabird data as part of their role as a MEDIN DAC. A statement is needed on the MEDIN website to show that seabird data are in scope for DASSH. DASSH would not be providing a duplicate of the data flows that JNCC already have in place for seabird data.

*Action 12.1: DL to provide CMH some text for MEDIN website to state that DASSH remit includes seabird data.*

**Underwater noise:** BODC has previously expressed an interest in becoming a DAC for underwater noise and has recently provided an update on this aspiration. For passive acoustic data, BODC expects to be in a position to archive this type of data within the next year, as work takes place amongst the NERC data centres to establish a high-volume data archive using JASMIN storage at CEDA. Both Cefas and MSS may be interested in exploring the archival of underwater noise data at BODC.

**Socio-economic data:** There is a joint NERC-EPSRC funding call Sustainable Management of UK Marine Resources (SMMR). CP has spoken to the SMMR champions about the lack of a MEDIN DAC for socio-economic data. The champions are considering that this could be a useful legacy of the SMMR programme.

ii) Proportion of data in MEDIN portal from DACs

There are 9904 records out of 15861 (63%) where MEDIN DAC is custodian.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Metric** | **ADS** | **BGS** | **BODC** | **Cefas** | **DASSH** | **HES** | **MSS** | **Met O.** | **RCAMHW** | **UKHO** | **Comments** |
| 1) Total records per DAC | 74 | 857 | 1107 | 2058 | 710 | 47 | 282 | 7 | 26 | 4736 | Where DAC is custodian |
| 2) New/updated records in year | 74 | 22 | 45 | 536 | 496 | 25 | 54 | 1 | 13 | 0 | Where DAC is custodian |
| 3) Records with online access to data | 74 | 855 | 1056 | 1914 | 631 | 17 | 240 | 2 | 0 | 4736 | Where url is not null |
| 4) Records with doi per DAC | 73 | 0 | 54 | 101 | 1 | 12 | 49 | 0 | 0 | 0 | Where url contains ‘doi’ |
| 5) Records with ‘2 clicks’ to data | 73 | 694 | 997 | 1914 | 165 | 12 | 57 | 1 | 0 | 0 |  |
| 6) Total MSCC records per DAC | 0 | 151 | 16 | 0 | 15 | 0 | 213 | 5 | 0 | 34 | Where DAC is custodian |
| 7) New MSCC records in year | 0 | 0 | 4 | 0 | 9 | 0 | 21 | 0 | 0 | 0 | Where DAC is custodian |
| 8) List of countries by DAC |  | Country where data collected from |
| England | 74 | 615 | 574 | 1890 | 555 | 11 | 182 | 5 | 24 | 3856 |
| Northern Ireland | 19 | 145 | 272 | 341 | 118 | 7 | 143 | 3 | 0 | 86 |
| Scotland | 24 | 478 | 631 | 714 | 249 | 45 | 278 | 3 | 0 | 991 |
| Wales | 29 | 157 | 300 | 524 | 150 | 0 | 44 | 4 | 22 | 1740 |

The portal contains lots of variations in organisation names. Some examples:

|  |  |  |  |
| --- | --- | --- | --- |
| BGS | -BGS-British Geological Survey-British Geological Survey (BGS)-British Geological Survey, Keyworth | Cefas | -CEFAS-Centre for Environment Fisheries and Aquaculture Science-Centre for Environment, Fisheries and Aquaculture Science-Centre for Environment, Fisheries & Aquaculture Science-Centre for Environment, Fisheries and Aquaculture Science, Lowestoft Laboratory-Centre for Environment, Fisheries and Aquaculture Science, Weymouth Laboratory |
| DASSH | -DASSH, The Archive for Marine Species and Habitat Data-Data Archive for Seabed Species and Habitats (DASSH) |
| UKHO | -Data Archive Centre |

* The numbers above will be presented as percentages when reporting in future.
* The issue with multiple names for a DAC has arisen previously during hackathons, a clean-up would be good. OW will look to rationalise the Cefas names.
* For 2 clicks to data, CP provided MARIS with different strings to look for, for the different DACs. We need to share this with the DACs.

*Action 12.2: CP/RM to provide DACs with portal database search strings for DAC metrics. [Done]*

3(b) Update from DACs

DACs each presented a brief summary of progress, see accompanying slides appended at the end of this document. Some key points included:

* ADS achieved Core Trust Seal accreditation in spring 2020.
* RCAMHW are moving away from the data platform shared with HES and are developing their own. Recruitment of maritime staff has been delayed by Covid-19 and this will affect the amount of maritime work that can take place this year.
* DASSH are receiving support from EU FAIRsFAIR with their Core Trust Seal application and expect to be completed at end of this year.
* Met Office. The World Meteorological Congress has instigated the biggest shift in meteorological metadata for 50 years. Five digit unique codes are being replaced with 30 digit codes, whilst providing continuity.
* Suggestion for MEDIN to put a statement on the MEDIN website to say that we are operating as normal during Covid-19.

*Action 12.3: CP to include statement on MEDIN website and social media to say that MEDIN is continuing to operate as normal during Covid-19.*

**4. DAC Work Plan 2020-21**

4(a) DAC Work Plan discussion

1. Activities at risk/postponed

RM went through the work plan tasks in the timeline from Paper P3. Task 11.7 to develop a DAC-wide approach to provenance tracking has been carried forward into the work plan for this year, as it was not completed last year. Three tasks were pushed back into later years in order to allow other related work to happen first or to redistribute the balance of work across the years. These include:

* 11.5 Ensure frictionless recovery of data by streamlining processes and promoting adoption of APIs across DACs and help them implement it.
* 1.10 Identify master datasets and deliver them through the MEDIN infrastructure, making all necessary changes as required.
* 2.10 Provide a data ingestion model, covering costs, timescales and capabilities for all DACs.

There was some concern about postponing the ingestion cost model work, as it is useful for potential users to have this information upfront, especially where data need to be sent to more than one DAC, e.g. for multidisciplinary data sets. The group agreed to do something very basic towards this activity this year as a start, e.g. each DAC could list what data they will take in free of charge.

1. New Development tasks

**Provenance tracking discussion:** RM presented preliminary findings from the provenance tracking questionnaire that was sent to the DACs in April 2020. JR raised that most of what is coming out in the results is to do with whether EDMO is being used or not. MSS can’t use EDMO as it would put an obstacle in the way of getting people to populate the metadata and it is difficult enough already. EDMO is updated by MEDIN core team and updates overnight. MSS catalogue is self-service and has no direct access to EDMO or RORs, it is not a priority for them at the moment.

MEDIN wants to improve the consistency of approach to provenance and for DACs to move away from free text and more towards using controlled vocabularies.

Suggestions for next steps included:

* To look at each element and say what could be done to move away from free text and towards controlled vocabulary.
* Post processing to match EDMO.
* Use ROR (Research Organisation Registry) identifiers instead of EDMO. ROR is a community-led project to develop an open, sustainable, usable, and unique identifier for every research organisation in the world. EDMO captures the granularity of evolution of an organisation, whereas ROR doesn’t.
* Ultimately looking for DACs to comment on how they can move to using more controlled vocabularies in the provenance fields.

*Action 12.4: RM to complete report from provenance tracking questionnaire.*

*Action 12.5: DACs to provide information on what they might be able to do about moving towards using controlled vocabularies in MEDIN metadata for all relevant fields in the MEDIN discovery metadata standard and use of PIDs for data submissions.*

**OGC compliant view and download services:** MEDIN is asking the DACs to document their OGC compliant services.

*Action 12.6: DACs to include in annual reporting what services they provide that are OGC compliant and to ensure that the services are described using service metadata in the MEDIN Portal.*

**Defining an API structure for DACs to adopt:** CP is aware that OGC are looking at APIs for environmental data retrieval and has had initial discussions with some contacts (Alexandra Kokkinaki – BODC, and Chris Little – Met Office) about relevance to MEDIN. The first version of an environmental data retrieval API is due to be launched in the next couple of weeks. MEDIN has been invited to be part of the process. CP would like to invite Alexandra Kokkinaki to be the liaison between OGC and MEDIN and feed into the DAC Working Group. The DAC WG agreed that this was a good suggestion and that we should be aligning with this emerging standard. MEDIN should look to recommend adoption of this newly emerging standard for APIs for Environmental Data Retrieval, providing it meets all the needs of the broader marine community.

*Action 12.7: CP to invite Alexandra Kokkinaki to next DAC WG meeting to discuss OGC API work.*

**Horizon scanning for new technologies:** It is hoped that the DACs can take a lead on this. JR is chair of the ICES Data Information Group, which does a horizon scanning activity as well and JR can feed that back to MEDIN. It was suggested that a proforma could be developed to capture feedback from sources such as individual DACs, MEDIN Resources and Applications WG and present as part of the annual mega meeting. JR will champion this at the mega meeting.

*Action 12.8: CP to speak to other groups about horizon scanning and provide a proforma to collate information for presentation at MEDIN mega meeting.*

*Action 12.9: JR to lead an item on horizon scanning for new technology at the next MEDIN Mega meeting.*

4(b) MEDIN small funded projects

At the end of 2019, a MEDIN call for small projects (with focus on new technologies) was launched to MEDIN Partners. There was a total pot of £15k available. Six proposals were submitted for review by a panel of three independent reviewers who scored the bids according to relevance to the call, how well defined they were and value for money. There were two projects funded: one from DASSH to develop an image management system, and one from BODC to expose discrete sample data via ERDDAP technology. It was felt that the timing of the call was good as the confirmation of funding coincided with the start of the financial year for most organisations. Another call for small projects will happen towards the end of the year.

4(c) DAC accreditation: adopting Core Trust Seal

ADS has recently gained Core Trust Seal accreditation. HES is in the process of applying for Core Trust Seal accreditation, and UKHO is expecting to start the process, hoping to gain it during 2020-21. DASSH is also in the process of applying and is receiving support from the FAIRsFair project. BGS already has CTS accreditation and will be due to renew it in 2021-22.

Financial support is available through MEDIN to cover the 1000 euro fee. There is £3k built into the yearly MEDIN budget to cover CTS fees. There is also an umbrella discount (25%) available for 10 organisations over a three-year period, and CP is awaiting a response to her enquiry as to whether CTS would consider a longer time period for the umbrella discount, while MEDIN DACs transition to CTS accreditation. If DACs require financial assistance then they need to let CP know.

*Action 12.10: DACs to let CP know if they require MEDIN financial help to cover their CTS application fee.*

5**. Promoting the DAC Network**

5(a) OECD – Valuing marine data

OECD is working with GOOS on a project to estimate the value of marine observations. There is a survey live at the moment, which has been advertised via DACs and MEDIN, as well as other avenues. The survey is looking to target all users (i.e. global, not just UK). There were 78 responses to the survey to date and the DACs were asked to encourage their users to complete the survey.

5(b) Forthcoming international meetings

This will be routinely covered as part of DAC annual reporting, and there was a round table of updates from each DAC. Meetings included: ICES Science and Advisory Committees, Science Industry Sampling Data Workshop, EMODnet Biology network meetings, OBIS meetings, European Marine Board, IMDIS, World Conference on Marine Biodiversity, EMODnet Ingestion, RDA, EOSC, IHO meetings, EMODnet Geology, Digital Pasts, SEADDA .

5(c) DAC webpage updates

Update pages with Core Trust Seal logo for those DACs that have accreditation.

*Action 12.11: RM to ask MEDIN Communications to update BGS and ADS DAC pages to include Core Trust Seal logo.*

**6. Coordinated data archiving**

6(a) Strengthen and formalise archive of multidisciplinary data

In the new MEDIN Business Plan, Task 2.6 is to strengthen and formalise triage process for archiving multidisciplinary datasets. The existing triage process, which is dealt with through the MEDIN helpdesk, has had limited use to date. The main use has been the archive of MCZ data at Cefas and BGS and this is now an established and functional workflow.

DL gave a short presentation to outline the process at DASSH and covered some issues and challenges. While the Crown Estate has many examples of multidisciplinary data deposits, there are very few at the DACs.



Issues and challenges include:

* DAC acceptable formats
* Mechanism of data transfer
* Multiple/unlinked metadata records – want to avoid duplication
* Mixed permissions
* Re-aggregation – not been tested yet
* Transparency/communication with data provider

Pressure on monitoring agencies is increasing, therefore demand will increase. MSS has been trying to engage with The Crown Estate Scotland. One issue is that for the provider we are telling them that we will scatter their data across the DACs. Another issue is how permissions are managed in the different DACs, different DAC cost models, format requirements etc., as well as the requirement for a depositor to have conversations with lots of different DACs instead of one person. We need a report that goes with the re-aggregated data. We are missing an opportunity e.g. to answer the question – how can I find MCZ data in the MEDIN portal?

DL suggested starting with a table of acceptable formats and indicative costs for each DAC. Formats are captured in the MEDIN guidelines. This would need to include some narrative for the end users to inform them why we are doing this to get them into the DAC network and that they can be brought back together.

MSS and Cefas only take data from their own organisations. DASSH provides the third party component of FishDAC. The next steps are for the MEDIN Helpdesk to start to pull together some use cases and requirements for archive of multidisciplinary data.

*Action 12.12: DL and helpdesk to develop initial plan to formalise multidisciplinary archive, e.g. ask DACs what they can accept at no cost; ask DACs to provide indicative costs for archiving data; collate and feedback at next DAC WG meeting.*

**7. DAC Data Access Services**

7(a) Improving user experience

MEDIN is looking to make more progress in DACs offering direct access to data from the MEDIN portal within 2-clicks. This is achieved by a user being able to click on a URL in the MEDIN portal and then download data in two further clicks of the mouse. In 2015 10% of the data in the MEDIN portal offered 2-clicks access to data. In 2020 this is now 35%. This is MEDIN’s response to global initiatives pushing to link data to metadata.

The Crown Estate has links to download the data but this is not yet showing in their metadata records in the portal. They are hoping to have sorted this by the end of the calendar year, which will increase the percentage further.

MEDIN isn’t aiming for 100% of records to offer 2-click access as this is not feasible for some data which have restrictions. UKHO records form a large percentage of DAC records in the portal and they will not be in a position to offer 2-clicks.

*Action 12.13: HW to send CP some narrative to explain the issue around UKHO records and why 2-clicks can’t be progressed.*

If UKHO records are excluded, what is proportion of DAC records offering 2-clicks to data?

*Action 12.14: CP to pull out proportion of records offering 2-clicks to data if UKHO is excluded.*

7(b) QC of existing metadata records

Gaynor Evans provided a report showing the status of metadata validation errors for the DACs. The DACs would find it useful to see validation changes over time.

*Action 12.15: GE to provide DACs with a report of metadata validation changes over time.*

JR reported that for the last 1.5 years he has records that validate without problem in Maestro and then fail validation in the portal.

*Action 12.16: CP to report on when the most recent version of Maestro will be released.*

*Action 12.17: CP to feed back when increased portal compliance checks will be implemented.*

**8. AOB**

Call for Marine Data News articles – deadline 10th July.

Next meeting – will be held in conjunction with mega-meeting and will be held as a remote meeting

**Table 1:** Actions from DAC WG meeting on 11th June 2020

|  |  |
| --- | --- |
| **Action** | **Progress** |
| 12.1 | DL to provide CMH some text for MEDIN website to state that DASSH remit includes seabird data. |  |
| 12.2 | CP/RM to provide DACs with portal database search strings for DAC metrics. | Done |
| 12.3 | CP to include statement on MEDIN website and social media to say that MEDIN is continue to operate as normal during Covid-19. |  |
| 12.4 | RM to complete report from provenance tracking questionnaire. |  |
| 12.5 | DACs to provide information on what might be able to so about moving towards using controlled vocabularies in MEDIN metadata for all relevant fields in the MEDIN discovery metadata standard and use of PIDs for data submissions. |  |
| 12.6 | DACs to include in annual reporting what services they provide that are OGC compliant and to ensure that the services are described using service metadata in the MEDIN Portal *(.* |  |
| 12.7 | CP to invite Alexandra Kokkinaki to next DAC WG meeting to discuss OGC API work. |  |
| 12.8 | CP to speak to other groups about horizon scanning and provide a proforma to collate information for presentation at MEDIN mega meeting. |  |
| 12.9 | JR to lead an item on horizon scanning for new technology at the next MEDIN Mega meeting. |  |
| 12.10 | DACs to let CP know if they require MEDIN financial help to cover their CTS application fee. |  |
| 12.11 | RM to ask MEDIN Communications to update BGS and ADS DAC pages to include Core Trust Seal logo. |  |
| 12.12 | DL and helpdesk to develop initial plan to formalise multidisciplinary archive, e.g. ask DACs what they can accept at no cost; ask DACs to provide indicative costs for archiving data; collate and feedback at next DAC WG meeting. |  |
| 12.13 | HW to send CP some narrative to explain the issue around UKHO records and why 2-clicks can’t be progressed. |  |
| 12.14 | CP to pull out proportion of records offering 2-clicks to data if UKHO is excluded. |  |
| 12.15 | GE to provide DACs with a report of metadata validation changes over time. |  |
| 12.16 | CP to report on when the most recent version of Maestro will be released. |  |
| 12.17 | CP to feed back when increased portal compliance checks will be implemented. |  |









