MEDIN DAC WG Meeting 10

11:00-16:00 Monday 29th April 2019, Medical Research Council, London

|  |
| --- |
| **Attending** |
| Jon Parr – Chair (JP) | MBA | Jens Rasmussen (JR) | MSS |
| James Ayliffe (JA) | BODC | Callum Stone | Met Office |
| Dan Lear | DASSH | Oliver Williams (OW) | Cefas |
| Chelsea Bradbury (CB) | The Crown Estate | Garry Baker (GB) | BGS |
| Helen Wojcik (HW) | UKHO | Clare Postlethwaite (CP) | MEDIN |
| Gareth Edwards (GE) | RCAHMW | Robin McCandliss – (RM) | MEDIN |
| Gaynor Evans (GEV) | MEDIN |  |  |
| **Apologies** |
| Peter McKeague (PMcK) | HES | Tim Evans (TE) | ADS |
| Mary Mowat (MM) | BGS |  |  |

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

**Table 1:** Actions from DAC WG meeting on 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. | Closed - Nothing received. If anything is required, should raise with Portal Steering Committee.  |
| 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. | Closed – on agenda. |
| 9.3 | Chelsea Bradbury to discuss with Peter Thijsse issue around MDE records being validated on maestro but not fully appearing on the portal. | Ongoing – action is now with GE. |
| 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. | Closed. ADS have a short-medium term fix in place. |
| 9.5 | Peter Thijsse to reinstate the metadata owner field filter. | Closed – GE has asked for this to happen. |
| 9.6 | Clare Postlethwaite to identify critical fields to test for metadata validation and to identify how many records would fail at present. | Closed – on agenda. |
| 9.7 | Jon Parr and Matt Arnold to discuss pinning down a date for hackathon with Dan Lear. | Complete: The Hackathon took place 7th March.  Was 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. | Closed – UUIDs do not seem to be fluctuating any longer. |
| 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. | Closed – on agenda to discuss metadata validation and versioning. |
| 6.1 | **Clare Postlethwaite** to ensure UKDMOS/MEDIN document is made available on MEDIN website once finalised. | 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. | Closed – TCE holds metadata record for full dataset, BGS has a metadata record for the subset that it holds. |
| 2.18 | **Clare Postlethwaite** to speak to JNCC about possibility of them becoming a Bird DAC. | Ongoing – will be taken forward new Business Plan |

**3. DAC network status**

3(a) Brief report from DACs

The DACs each presented 1 slide summarising progress.

*Action 10.1: RM to circulate DAC update slides to DAC Working Group*

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

4(a) Progress with DAC Work Plan

RM went through the DAC work plan for 2019/20 (Paper P3). The work plan is based on the new MEDIN Business Plan (2019-2024) and has resulted in some new activities for the DAC work plan, as well as the ongoing operational activities.

Items with an asterisk\* have been identified as being at risk due to funding uncertainties. Sponsorship income is not increasing, therefore we need to become more efficient and aim to deliver more for less. There is also the possibility that work at the DAC organisations may already be aligned with the work plan and could be used to delivered elements of the plan.

New items include:

* Review implications of cloud adoption in the DAC network i.e. revise MEDIN accreditation should a DAC use cloud technology.\*
	+ Output is an updated DAC accreditation template.
* Develop a DAC-wide approach to provenance tracking.
	+ The Geo-Spatial Commission is looking into the issue of provenance tracking and is putting out a report on this with recommendations in September 2019. It would make sense to use the report to inform our activity on provenance tracking.
	+ Envri-Plus has also been looking at the issue of provenance and tracking.
	+ Suggestion that BGS and DASSH lead on this activity.
* Proactively identify, target and engage with new suppliers to get their data on the MEDIN portal
	+ Issue around customers wanting to engage and archive data, but being put off by the costs.
* Develop a data ingestion model, covering costs, timescales and capabilities for all DACs.\*
	+ Output for year 1 is a scoping study. This will be a template of what should be covered by the data ingestion model.
	+ Cost Benefit Analysis can provide some of the information towards scoping study.
	+ It was recognised that this will be complicated as each DAC may have a different cost model and there may be different costs within DACs, depending on data type.
* Horizon scanning for new technology approaches e.g. schema.org, so that MEDIN is prepared to adapt to advances in technology.
	+ This can be included as a standing item on the DAC Working Group agenda.

*Action 10.2: CP and RM to draft a MEDIN definition of provenance tracking.*

*Action 10.3: BGS and DASSH to lead on provenance tracking activity and to discuss GeoSpatial commission and Envri-Plus work on provenance tracking at next DAC WG meeting.*

Small data archiving projects were also identified as being at risk; however, the plan is to run a single funding call for projects across the Resources and Applications work stream as well as the DACs.

Regarding NI joining the FishDAC, the MEDIN DAC fund of £11k would then need to be split across four organisations (currently split across three), therefore reducing incentive at the DAC level to push for NI to join. DACs will continue to receive £11k per annum to put towards delivering the Work Plan. CP would like the DACs to provide a breakdown of spend for the £11k in their annual report.

*Action 10.4: RM to ask for DACs to include a breakdown of the spend of MEDIN DAC £11k funds as part of this year’s annual report (2018-19)***.**

These items should be reviewed at next DAC WG meeting.

*Action 10.5: RM to include new work plan items, small projects, NI FishDAC progress and DAC funding issue on agenda for discussion at next DAC WG meeting.*

It was noted that some of the activities on the work plan are multi-year activities and will continue into 2020-21. This needs to be made clear in the work plan. Moreover, some of the activities in the work plan are beyond MEDIN’s control and in such cases, the working group will report on status.

*Action 10.6: CP to make it clear in the MEDIN work plan where activities will be carried out over several years.*

4(b) Metrics for new Business Plan

The business plan contains 12 High Level Objectives (HLOs). The table below shows these HLOs and which DACs are involved/leading. The DACs need to define KPIs for these activities and agree on metrics to report.

|  |  |  |
| --- | --- | --- |
| **No.** | **HLO** | **Work stream 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 |

The metrics the DACs report allow us to assess progress and are used by the MEDIN Coordinator in many different ways, for example to illustrate MEDIN progress to Sponsors. It was noted that it is desirable for DAC metrics to be consistent, comparable and easy to gather wherever possible.

The DAC metrics that have been in use for the 2014-2019 Business Plan (paper P4) were assessed and given a thumbs up/down on how useful/measurable/consistent they are.

Total number of data sets available - DACs agreed that quality is more important than quantity, high numbers is not necessarily better. Some DACs requested that this metric come straight from the MEDIN portal.

Total number of view services – hard to count and inconsistent approach across DACs.

Total number of download services – hard to count and inconsistent approach across DACs.

Some DACs are uncomfortable reporting whether their view and download services are INSPIRE compliant because it is hard to validate compliance and requested MEDIN Executive considers this reporting can be dropped.

*Action 10.7: RM/JP to ask MEDIN Executive whether to drop reporting on INSPIRE compliance.*

New data sets archived - Varies from DAC to DAC because of different data types. Some DACs add to existing data sets, others increase the number of data sets they hold. Some count number of rows, while others report on volume held. There was a suggestion to index the growth to make it more comparable. For example, DACs could report on percentage increase in their data holdings year on year.

*Action 10.8: Each DAC to suggest a metric for their DAC to cover increase in new data archived, with some narrative to explain by end of July 2019.*

Data sets accessed - Number of DOIs minted and citations suggested, as well as statistics from MEDIN portal. Cefas is involved with the DataCite ‘Make Data Count’ initiative. It is a standard code of practice for scientific data repositories to consistently report hits, to views and downloads of data by human end users.

*Action 10.9: OW to report back on DataCite Make Data Count at next DAC WG meeting.*

There was not enough time to bottom out the KPIs for the new business plan, and it was agreed to revisit this before the next DAC WG meeting.

*Action 10.10: RM to invite further discussion on KPIs for new business plan so that they can be agreed before DAC Working Group November meeting, aiming for September 2019.*

4 (c) DAC accreditation:

At present DACs are accredited using the MEDIN accreditation process. The group discussed potentially adopting a wider accreditation process, such as the Core Trust Seal.

Core Trust Seal is an independent and recognised standard used internationally. The Core Trust Seal process requires a DAC to complete 12 sections of narrative and submit with a self-rating. The application is assessed by 3 reviewers.

Pros and cons include:

* MEDIN accreditation is free, while Core Trust Seal has a financial cost (~1000 euro per DAC covering 3 years). If MEDIN acts as the umbrella organisation for a minimum of 10 accreditations over a 3-year period, then a 25% discount may be obtained.
* MEDIN accreditation lasts for 5 years, while Core Trust Seal accreditation needs to be reapplied for every 3 years.
* MEDIN accreditation takes about 4 weeks, preparation of material for Core Trust Seal takes considerably longer (e.g. BGS).
* Core Trust Seal has access to a wider pool of independent reviewers than MEDIN.
* Some of the DACs have achieved or are in the process of applying for Core Trust Seal already.

The DACs present were broadly supportive of moving to Core Trust Seal accreditation for MEDIN.

*Action 10.11: GB to send information on Core Trust Seal accreditation process.*

*Action 10.12: CP to take DAC WG recommendation to Exec that we move to Core Trust Seal and delay HES reaccreditation until a decision has been reached.*

5**. Promoting the DAC Network**

5(a) Forthcoming international meetings

Each year, CP provides a summary of international meetings attended for publishing on the MEDIN website. The Sponsors Board expressed an interest in knowing about these meetings in advance.

*Action 10.13: DACs to send CP list of international meetings they are planning to attend during 2019-20 within 2 weeks.*

5(b) DAC webpage updates

None required this time.

5(c) Tweets

CMH was interested to find out whether the tweeting of dataset of the week was causing any increase in traffic to DAC web sites/download services. The DACs would need to know when the tweets went out in order to be able to supply the feedback.

*Action 10.14: CMH to let DACs know when tweets went out about dataset of the week for their DAC, so they can check to see if there are signs of increase in traffic/requests.*

**6. Coordinated data archiving**

6(a) coordinated approach to archiving multidisciplinary datasets (DASSH)

Nothing to report. MEDIN Helpdesk has had no enquiries to deposit multi-disciplinary data as yet.

**7. DAC Data Access Services**

7(a) Metadata validation and versioning

Version 3 of the MEDIN Metadata Standard has been released. This is in response to updates to INSPIRE and GEMINI. Most DACs are currently on version 2.3 onwards. The Defra and GEMINI approach is to accept all metadata of any version with the proviso that once past the end of the 2019 calendar year, any newly created or edited metadata records need to be made/updated to GEMINI 2.3.

DACs agreed to accept this approach such that:

* New metadata records created after 31st December 2019 will be done using Version 3.
* Updates to existing metadata records after 31st December 2019 will entail updating the record to Version 3.
* Legacy records can remain as they are.

MSS and UKHO noted their support and will aspire to this, but may not be able to commit resource to meeting this deadline at present, meaning that they may not be able to add new metadata records (or update existing records) to the portal.

7(b) QC of existing metadata records

There are just under 15,000 metadata records in the portal. At the moment, all the validation filters are off. If they were to be switched on, ~50% of the records would fail validation and the number of records appearing in the portal would substantially decrease until the errors were addressed by the DACs.

DACs agreed to address metadata errors by the end of the 2019 calendar year.

MSS and UKHO noted their support and will aspire to both correction of errors and updating records to version 3 of the metadata standard, but may not be able to commit resource to meeting this deadline at present, meaning that they will not be able to add new metadata records (or update existing records) to the portal.

7(c) Codefest

The Codefest took place in March 2019, with 10 participants from five organisations (MEDIN, DAERA, BODC, JNCC, and MBA). The focus was on UKDMOS and outputs include a ‘heat map’ of UKDMOS coverage (temporal and spatial), as well as word maps of platform frequency and legislative drivers.

There were some lessons learnt from the third MEDIN hackathon. It is better to have the data prepared and ready for attendees to use, rather than them having to spend time interrogating the database themselves. The same DACs participated each time, and it would be good to be able to open up to wider participation.

TCE reported having a hack day targeting digital data, which was geared as a networking event. The main cost involved was the catering. TCE required people to apply beforehand, submitting information about their background, which ensured TCE could get a good mix of people. They advertised the event on Linkd-In, TCE website, and in universities.

It was noted that future MEDIN hackathons are at risk, due to funding constraints and CP suggested that a DAC might wish to run one. UKHO are planning a hackathon on wrecks.

*Action 10.15: HW to provide CP with contact details for organisers of the UKHO hackathon.*

**9. AOB**

9(a) Date of next meeting

The next DAC WG meeting will be held alongside next mega-meeting in Q3 of 2019-20.

**Table 1:** Actions from DAC WG meeting on 29th April 2019

|  |  |
| --- | --- |
| **Action** | **Progress** |
| 10.1 | RM to circulate DAC update slides to DAC Working Group | Done |
| 10.2 |  CP and RM to draft a MEDIN definition of provenance tracking. | Done |
| 10.3 |  BGS and DASSH to lead on provenance tracking activity and to discuss Geospatial Commission and Envri-Plus work on provenance tracking at next DAC WG meeting. |  |
| 10.4 | RM to ask for DACs to include a breakdown of the spend of MEDIN DAC £11k funds as part of this year’s annual report (2018-19)**.**  | Done |
| 10.5 | RM to include new work plan items, small projects, NI FishDAC progress and DAC funding issue on agenda for discussion at next DAC WG meeting. |  |
| 10.6 | CP to make it clear in the MEDIN work plan where activities will be carried out over several years. |  |
| 10.7 | RM/JP to ask MEDIN Executive whether to drop reporting on INSPIRE compliance. |  |
| 10.8 | Each DAC to suggest a metric for their DAC to cover increase in new data archived, with some narrative to explain by end of July 2019. |  |
| 10.9 | OW to report back on DataCite Make Data Count at next DAC WG meeting. |  |
| 10.10 | RM to invite further discussion on KPIs for new business plan so that they can be agreed before DAC Working Group November meeting, aiming for September 2019. |  |
| 10.11 | GB to send information on Core Trust Seal accreditation process. | Done |
| 10.12 | CP to take DAC WG recommendation to Exec that we move to Core Trust Seal and delay HES reaccreditation until a decision has been reached. | Done |
| 10.13 | DACs to send CP list of international meetings they are planning to attend during 2019-20, complete this by mid June. | Done |
| 10.14 | CMH to let DACs know when tweets went out about dataset of the week for their DAC, so they can check to see if there are signs of increase in traffic/requests. |  |
| 10.15 | HW to provide CP with contact details for organisers of the UKHO hackathon. |  |
| **Actions outstanding from earlier DAC WG meetings** |  |
| 9.3 | Chelsea Bradbury to discuss with Peter Thijsse issue around MDE records being validated on maestro but not fully appearing on the portal. | Ongoing – action is now with GE. |
| 7.9 | **Amended action:** GE to follow up with Tim Evans the development of WAF for harvesting ADS records to the portal | On hold until resources are available. ADS records are in the portal as interim measure taken to bulk export records via DASSH. |