ARDC Platforms Project Progress Report – 1 June to 13 Aug 2021

Project Tracking Information

1. Project Details:

Project title: GMRT-AusSeabed

Lead organisation: Geoscience Australia

Other organisations involved in project: Bureau of Meteorology, Deakin University, James Cook University, Lamont-Doherty Earth Observatory, CSIRO, Australian Antarctic Division

Contact person name: Kim Picard

Reporting period: **1 June to 13 Aug 2021**

2. Project Aims:

This platform will provide seamless, high-quality elevation and bathymetry data to the consistent standards required for oceanographic models and predictions. This work will accelerate research by reducing manual effort, avoiding duplication and removing the barrier of specialised skills needed for accurate bathymetry data.

The outputs, suitable for high-performance computing or desktop environments, ensure the service is scalable and accessible to the broader research community. The consistency, resolution, and processing capability delivered by this platform improve the accuracy, validity and reliability of modelling. As such, this platform will benefit coastal communities nationally by informing environmental management and risk mitigation.

3. Summary of Project Expenditure:

*Note that the Description of expenditure has been modified from the original project plan. The Project management cost have been separated from Component 3 to ease tracking of expenditure.

Description of expenditure	Budgeted expenditure of ARDC investment	Actual expenditure of ARDC investment
Component 1: End-user analysis and platform requirements setting	**$73,421**	**$43,816**
Component 2: Bathymetry subset dataset	**$53,202**	**$18,271**
Component 3: GMRT adoption and Expansion	**$199,203**	**$27,360**
Project Management*	**51,880**	**$18,311**
Total	**377,707**	**$108,028**

4. Progress against Project Milestones and Deliverables:

Progress are included in the detailed project plan under the Status column of ARDC GMRT AusSeabed_PL019_Project Plan - Progress Aug 2021.xlsx

Core design decisions have been made since the development of the project plan and have resulted in a reordering and slight modification of activities that make up Milestone's 3,4,6,7 and 8. Milestone 9 will remain on schedule as per the original plan. A revised work plan is attached in the appendix and has been translated into the original detailed project plan.

Milestone 1 and 2: Completed. Final report attached. Publication online only awaiting final step by GA publication team.

Milestone 3: Workshop 1 and 2 report completed. Final reports attached.

The decision to move forward with TileDB has been tabled through workshop 1 and 2 and is supported by the users engaged. With this in mind, workshop #3 has been changed to a document/review process rather than a workshop/document/review process. (Workshop #3 activity removed, but consultative review of documentation remains).

Milestone 4: Uptake of the TileDB services modifies this milestone into a point cloud conversion to a metadata harvesting activity.

Milestone 5: Well underway. Subset datasets have been identified and include datasets within the inner red boundary in Figure 1. The goal of this subset dataset is to test the platform. Requirements identified for selection included sensor type, year of collection, resolution, quality, point cloud vs grid. Dataset coverage that are not presented but are considered include LiDAR along Victoria coast, Satellite-derived bathy (SDB).

Due to a synergy with another GA led project, where 3D seismic-derived bathymetry datasets will be produced this FY, we agreed to leverage of this opportunity and have the relevant 3D-seismic data delivered in Jan-Feb 2022. This will not impact testing, we consider it a bonus.

Prioritised datasets are being processed and should be delivered in-time, noting that the team agreed at our July monthly meeting to have only one delivery phase rather than the two in the original plan (line 37-39 in detailed plan). This was deemed more appropriate because processing of the datasets is done mostly all at once in a software project and it won't impact the overall project.

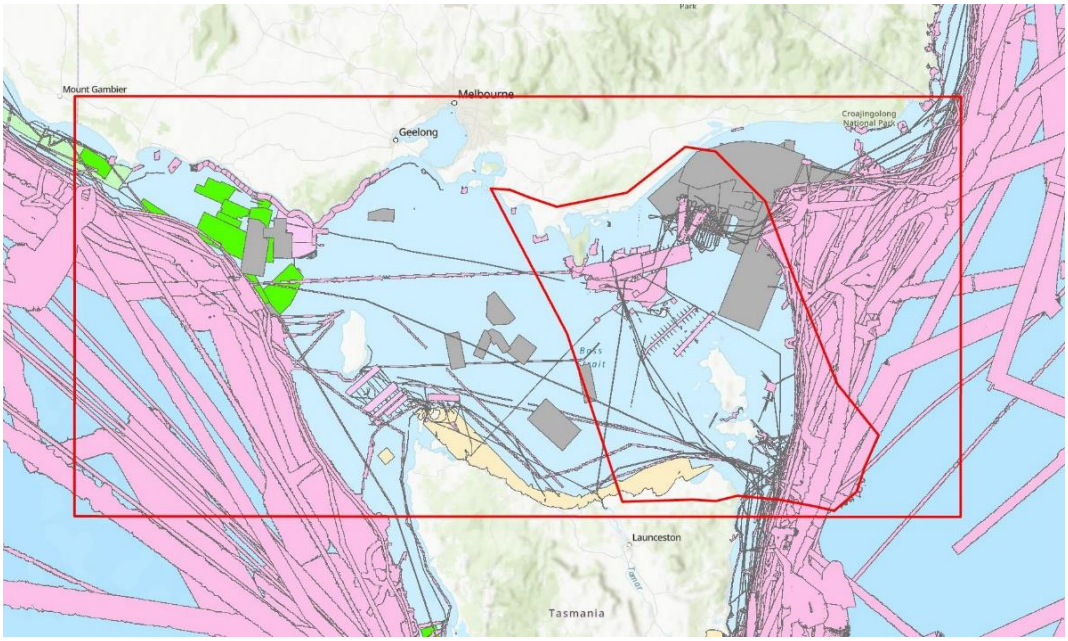

Figure 1 Coverage for bathymetry data existing within the region of interest (inner red polygon). These include bathymetry from global model SRTM, multibeam and singlebeam surveys (pink), 3D seismic (grey).

Milestone 6: We have moved the setup of the Jupyter hub to later in the year to allow for both the data delivery and for the ARDC Jupyter Hub to be established for assessment. Should we be unable to use the ARDC Jupyter Hub, this work has been timetabled for November 2021, and will use the DEA model of installation in AWS. When the ARDC NECTAR infrastructure comes online, we will:

a. Ensure that we are actively engaging with the nectar team to inform requirements.
b. Look at migrating any notebooks/services into this space.

Milestone 8: We have made the decision to work forward in AWS rather than developing on NCI/Nectar/On-prem infrastructure - therefore there is no requirement for capacity to be assigned for Cloud Infrastructure setup as was originally planned.

Milestone 9: We have assumed a hard date of MS9 – GMRT grid composer modified to enable user-control setting as this is the point where LDEO will need to be very active in the development process.

Progress against our KPIs are below:

KPI 1: Subset test datasets delivered to GA in a timely fashion for model development

As per variation of MS 5 below, dataset delivery and publication has been combined into 1 delivery mid-September.

KPI 2: Projects artefacts published on AusSeabed in a timely fashion

Our first three reports (User need analysis, workshop #1 and #2) have been published on AusSeabed website on 15 Aug. These have been delayed by 1 month from original plan due to a late decision to publish through the GA process.

KPI 3: Platform tested by key users returning positive and constructive feedbacks

n.a.

KPI 4: Platform code published to GitHub in a timely fashion

n.a

5. Current Project Constraints & dependencies

Are there any project constraints or dependencies that are outside the control of the project team and may impact on project delivery?

None to report at this point in time

6. Risks

Are there any changes in risks for the project and relevant mitigations and timeframes?

None to report at this point in time, but this next quarter is critical to having all pieces in place.

7. Issues/Challenges

Are there any issues or challenges that you have encountered during the reporting period?

None to report at this point in time

9. Anything else you would like to tell us about the project:

Impact information – Section UNCHANGED

Please not that we are not reporting on this information for this Interim report. We will include in the December report

Impact metrics

Impact metric (*add relevant details/links in next section)	Interim report 1 June	Interim report 2 Dec	Interim report 3	Interim report 4
Number of datasets published (preferably with record in Research Data Australia)*	0			
Number of publications using data generated by platform*	0			
Number of publications that cite /reference the platform*	0			
Social media mentions of platform*	7			
Blog posts/news articles that reference platform*	1			
Number of users*	0			
Number of active users (over 90-day period)	0			
Number of jobs run (i.e. number of processes submitted through the Platform)	0			
Instruments supported	0			
Number of engagement/outreach activities (e.g. seminars, presentations,	2+ (Tbcounted)			

demos, meetings with new stakeholders)*				
Number of researchers trained (online or face-to-face)*	0			
Availability of the platform (days in operation vs days offline)	0			

Links to publications and other communications

Published datasets generated by platform

N/A

Publications using data generated by platform

N/A

Publications that cite/reference the platform

N/A

Social media mentions of the platform

Blog posts/news articles that reference the platform

Details on engagement, outreach and training activities

We have established a communications plan for the project, in collaboration with the ARDC communications team. The communications plan is nearing finalisation and the communications schedule can be found (Comm_schedule_simplified.xlxs in shared folder). The communications campaign includes various media formats including video, infographics, articles, webinars and presentations across a wide range of audiences. In short, our plan is to highlight he project launch, the data subset, project mid-point, developed user tools and project end-point.

To-date, we have:

- Released a project launch article with the ARDC (https://ardc.edu.au/news/piecing-together-the-puzzle-of-australian-seabed-data/) and circulated through social media.
- The project page has been established by the ARDC (https://ardc.edu.au/project/gmrt-ausseabed/).
- Presented on the project at:
    - the external AusSeabed Quarterly Showcase in March
    - AODN Technical Advisory Group quarterly meeting in May, and
    - Geohab 2021 in May (international conference on Marine benthic habitats)

Coming up shortly, we have:

- A short introductory video featuring two partners (GA, LDEO) and a key end user (UWA) providing video bits and simple animation showing the expected result of the project. Delivery of the video on track for 1 June week.

Implementing FAIR Platforms and Outputs

FAIR project outputs

Describe the project's progress in making the **project** outputs (publications, code, etc) FAIR.

No progress has been made in the early stages of this project. All outputs are intended to be FAIR as outlined below.

FAIR platform outputs

Describe what functions/features are being implemented in the Platform to enable FAIR or FAIR-ready outputs (data, code, visualisations, etc).

- All datasets selected will be published to the AusSeabed data portal and the GMRT synthesis, which serves FAIR and Open data nationally and globally.
- We will be adopting ISO and OGC standards where appropriate to make frameworks interoperable and findable.
- We will be delivering/utilising accessible, open-source tools where possible
- We will be perusing automation and standardisation as a core value to build a reproducible platform.

Details on users of the platform

Provide a brief breakdown of where these users come from; for example, which institutions the core user groups are from, or how many users from each of the following broad categories: .edu.au, .gov.au, .org.au, .edu (international academic institutions), .com (gmails etc.), other.

We expect that we will engage 2900-3600 users per annum, based on AusSeabed and GMRT usage statistics. We expect a range of users from the academic, government and industry sectors.

We cannot provide a specific breakdown on where our users will come from as the platform is not live. However according to the 95 respondents to the end-user survey completed in Feb, users from all around Australia responded with highest percentage from WA, NSW, Vic and Qld, as well as internationally. All sectors were represented, with the most associating with private industry, Federal government, university and Research Institute (85%)

Impact stories

Please detail any research impacts or highlights here. Other impacts could include outcomes of industry collaborations, platform community activities leading to new policy or practice, etc.

No impact stories are available in the early stages of this project.

Appendix – Revision of work plan for Component #3

The revisions have been transferred to the detailed work plan spreadsheet (sheet Revised plan)

Milestone ID	Original Milestone Name	Description	Revised Date	Resources	More detailed
N/A	Not Applicable	Set up AWS ARDC Account	11/08/2021	**NL**/CET	
MS4	Point cloud conversion services	Establish TileDB Test environment/accounts	30/08/2021	**NL**/J6	- TileDB service credits provided - Access user setup (J6, VF, LDEO Dev)
		Dataset Conversion	22/10/2021	**J6**/LDEO/NL	- TileDB Training/Learning/research time - Point Cloud conversion/development - Metadata Harvesting/development - Testing
		(Potential) API/Services Development	12/11/2021	**J6**/LDEO	- Development/configuration of services/API - Testing
		Showcase of MS4 at End of Year (End of PI4 showcase/PI5 Showcase)	*06/12/2021*		- *(Note date selected based on sprint plans)*
		**May want to prep something for AUS/US workshop*	*13/14 October*		
MS6	Establishment of Jupyter Hub	Assessment of ARDC Jupyter Hub Project	19/11/2021		- Consideration of the capability of the program/limitations etc.
		Establishment of Jupyter Hub	10/12/2021	**J6**	- Only progresses should ARDC Jupyter Hub not fit the requirements - Pull DEA Jupyter config and standup (1 sprint)
MS7	Future state architecture and workflow process	Conceptual Model	Delivered	N/A	- (initial document – will be carried forward and expanded)

		Business process/context models	15/10/2021	**NL**/J6	- Standard business process, identification of edge cases/out of scope
		Data Flow Models	12/11/2021	**NL**/J6	- Data flow models for data provided, identification of edge cases/out of scope
		Final POC Architecture	01/12/2021	**NL**/J6	- Documentation of POC "As-Is" to support future state activities
MS8	Cloud infrastructure set up locally for data ingestion into GMRT and subset datasets ingested	Not applicable – development will be occurring in cloud, no additional effort required to "move".			
MS9	GMRT grid composer modified to enable user-control setting	GMRT grid composer modified to enable user-control setting	**Start 05/01/2022 -** 25/03/2022		
MS10	Connection to Ocean Model and delivery of GMRT-AusSeabed Platform PoC	Connection to Ocean Model and delivery of GMRT-AusSeabed Platform PoC	27/05/2022		
MS11	Additional capabilities identified added to the Grid Composer	Additional capabilities identified added to the Grid Composer	30/06/2022		