

GMRT-AusSeabed Steering Committee Meeting Minutes

Meeting no. 3: 17th november 2021, 11:00-12:30

Notes prepared by Kimberlee Baldry and Kim Picard

Attendees: Kim Picard (Chair, Geoscience Australia), Robert Kay (Committee member, Geoscience Australia), Paul Branson (Committee member, University of Western Australia /CSIRO), Eric Schulz (Committee member, Bureau of Meteorology), Kerry Levett (Committee member, Australian Research Data Commons), Kimberlee Baldry (Minute taker, Geoscience Australia).

Invited guests: Joshua Sixsmith (Geoscience Australia), Natalie Lenard (Geoscience Australia).

Apologies:

Actions list

Table 1. Action list with new actions raised during this Steering Committee meeting (SC3)

	Action	Responsible Party	Due date	Comments/Status
SC3.1	Put link to detailed plan in the ARDC progress report document	KB	December	Completed
SC3.2	KB to finish the ARDC Progress report and deliver to ARDC.	KB	December	Completed
SC3.3	KL to advise on how to report the delay of publishing the datasets to the Data Portal in MS5.	KB	December	
SC3.4	Push data providers to provide datasets by mid-December and close MS5.	KP, JS	Mid-December	In JIRA
SC3.5	KB send GMRT-AusSeabed video to ARDC comms to include in next video	KB	November	Completed
SC3.6	KP to contact Michelle about hosting an end-user workshop in the IMOS planning meeting in March 2022.	KP, KB	December	In JIRA

SC3.7	KP and others to prepare application for FOO spotlight presentation, or consider another form of participation (working group, last minute presentation cancellation?).	KP, KB	November	Completed. Spotlight will be presented at FOO at the Waves and Currents session. AusSeabed will also attend the Climate, MHW and temperature session.
SC3.8	JS talk to discuss with Diana see if there is anything worth presenting at this point in time.	JS	November	
SC3.9	NL and JS get a good understanding on infrastructure costs and services from TileDB in the future.	NL, JS	March	In JIRA
SC2.3	The technical team should define the prototype/MVP (MS7)	JS, KP	10 Oct	Completed
SC2.4	The technical team should provide justification the SC on metadata decisions.	JS	10 Oct	Completed

Introduction

- Meeting opened 1105
- The SC endorsed the agenda with no additional business raised

Minutes and Actions

- Minutes from SC Meeting #2 were endorsed out of session.
- SC minutes will be published on the new GMRT-AusSeabed webpage (<http://www.ausseabed.gov.au/gmrt>)

Progress Update

- MS1, MS2, MS3 are delivered. MS5 (datasets) are experiencing a minor delay
 - Most datasets have now been received by GA. GA is still gathering and Quality assuring data.
 - Data ingestion has not commenced, some data translation and GMRT requirements are still being implemented in the back-end of the platform. Team is close to beginning ingestion
 - Initially publication of datasets to the AusSeabed Data Portal will be delayed into March, to ease pressure on AusSeabed publication team who are experiencing a backlog. No objections reported from the SC on this change.

- There are 18 datasets that are preselected for ingestion to make up the “data subset”.
- The reasons MS5 is marked as delayed
 - Receiving the data from the data has been delayed but is underway. We are still waiting on datasets from providers.
 - Some delay has been caused by GA not providing clear expectations on how datasets should be provided and data specifications, and miscommunications within the instruction on how to provide metadata to GA. Time has been spent producing templates and resolving inconsistencies in the metadata templates received. Getting data specifications right has been important and knowing how to communicate these specifications to data providers has been challenging.
 - Some delay also due to receiving some datasets on hard drives, and minimal access to physical office space (COVID19 restrictions) to collect and upload this data.
 - Publishing datasets to the data portal will be delayed until March.
 - It is unknown as to how much the incoming datasets will influence the platform development. This is because it depends on the format and content of datasets received, which is currently unknown for datasets not yet received.
 - The result of the delay may result in not all datasets ingested in GMRT, some actions may need to be resolved at the end of the project.
 - ES commented that the team should get onto Data providers now before mid-December.
- Question from KL
 - What is the difference between having data on the portal vs in the GMRT-AusSeabed platform. This project will not merge the Data Portal and the GMRT-AusSeabed Platform, but will focus on the functionality of the platform. It is planned that the Data Portal will be merged with the GMRT-AusSeabed platform to ingest datasets in directly in the future of AusSeabed.
- MS4 and MS7 are delivered in draft format and are available for review by the ARDC SC. They will be complete by 01 December when the progress report is delivered.
- KP noted that a lessons learned document will be produced for the future on data collections and specifications.
 - KL would like this document shared with the public or published as an article.
- Webpage and video has been made live! <http://www.ausseabed.gov.au/gmrt> comms campaign to come
- An end-user workshop at the IMOS planning meeting is being considered for the GMRT-AusSeabed platform.
- A spotlight presentation at the Forum for Operational Oceanography, is being considered – deadline is in 2 days!
<https://www.foo.org.au/forum/foo-2021-registration/spotlight/>

-

Action SC3.1: Put link to detailed plan in the ARDC progress report document

Action SC3.2: KB to finish the ARDC Progress report and deliver to ARDC.

Action SC3.3: KL to advise on how to report the delay of publishing the datasets to the Data Portal in MS5.

Action SC3.4: Push data providers to provide datasets by mid-December and close MS5.

Action SC3.5: KB send GMRT-AusSeabed video to ARDC comms to include in next video

Action SC3.6: KP to contact Michelle about hosting an end-user workshop in the IMOS planning meeting in March 2022.

Action SC3.7: KP and others to prepare application for FOO spotlight presentation, or consider another form of participation (working group, last minute presentation cancellation?).

Action SC3.8: JS talk to discuss with Diana see if there is anything worth presenting at this point in time.

KPI and Risk Reporting

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

- As per variation of MS5, 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

- GitHub repositories are sitting within the AusSeabed organisation:
 - <https://github.com/ausseabed/reap-gsf> (reading the GSF)
 - <https://github.com/ausseabed/bathy-datasets> (dataset construction, metadata, geometry, tileDB schema outputs)

Risks

- Data Delivery and Publication delayed - Minor restructure of the development activities to allow more time for the data delivery to progress. As stated in MS5 of the progress report table, publication is not critical to the outcome of the project, thus delaying action is deemed reasonable.
- Lack of technical design clarity - Technical design document is being circulated and presented through to the AusSeabed and ARDC Steering committee for oversight and review.

Technical Architecture

NL and JS presented the technical architecture which has been developed for Component 3 – the GMRT-AusSeabed platform (MS4 and MS7). Slides are appended and available here. The technical architecture of the GMRT-AusSeabed platform has been provided to the ARDC SC for feedback in the version of a draft document. The SC endorse the technical architecture and the progress of the software development of the GMRT-AusSeabed platform.

General questions from the SC:

- TileDB - Software as a service. Is there a cost to the project using this? Will there be a cost beyond this project to use TileDB?
 - Yes, there will be a cost associated with the continuation of the work/platform
 - Avoids the cost of developing the infrastructure (which also needs to be maintained and has a maintenance cost)
 - We are looking to put an operationalising paper as an output for the project.
- Have I understood correctly, that even if you deployed and managed TileDB yourself, you'd still have the AWS compute costs
 - No AWS and compute costs – they are included in TileDB operational costs, because the TileDB provides the compute within this project.
 - An individual can deploy TileDB to your own infrastructure if they want, and then you would have the AWS compute costs.
 - TileDB could be deployed on the Nectar cloud – KL will pass onto her cloud team.
 - A conversation with TileDB will be had in the future. It is possible that they may develop infrastructure for us.
- Are we locked into a particular company to provide this and how would that interact with cloud services in the future?
 - No, you can always choose to develop your own cloud technology stack.
- User requirements – Were user requirements narrowed down based on feasibility?
 - Requirements were looked at broadly, not as specific features. The project has focused on requirements where the team were not sure if they could be met.

- Will there be un-met user requirements?
 - Broadly covered requirements detailed from the community. Specific requirements on features may not be met.
 - An operationalisation document is being put together on what needs to be done after the project to operationalise the platform.
 - The edit spurious data will be the feature that will be dropped from the project if it's too big, and may not be what the user is expected regarding editing data.
 - The thinking on this feature is more around providing metadata attributes to work with and not directly editing the data.
- In which AWS region is the SaaS running?
 - In the USA
 - Also have an Asia/Pacific region
 - TileDB are aware that we would require a technical stack in the AUS region and are open to standing up the service in our region
 - PB acknowledges that the chosen use of TileDB is a good decision for the project, and to avoid building new infrastructure. RK in agreement.
- Metadata harvesting – how are you handling different formats and structures with a metadata harvesting tool?
 - Metadata collection is being facilitated with an excel spreadsheet
 - There have been some problems with inconsistencies in this spreadsheet that is being populated by the various data providers.
- What about communicating the stac infrastructure to the data providers?
 - Good idea. This could possibly be done by providing some toolkits to the data provider to make them aware.
- How are you tracking the merging of the edges between datasets and to ensure there are no step stages?
 - It is on the team's mind to gather information on feathering and decision making on the algorithms used within the community.
 - It is a critical piece of software development required for data compilation

Action SC3.9: NL and JS get a good understanding on future infrastructure costs and services from TileDB.

Next meeting

- If anyone knows of a national platform to pitch phase two of this project for FY 2022/23, please let KP know

The next meeting will be in February 2022.

Meeting closed at 1230

Appendix 1: Agenda

Date: 17 November 2021

Venue: Online

Time: 11am ADST

Coordinators: Kim Picard, Kimberlee Baldry

Papers: Shared folder QuarterlyMeetings/Meeting3_17112021

Agenda

Time	Item	Duration	Paper
11:00	Welcome and Action	10 mins	Action list below
11:10	Progress Update	20 mins	Yes Progress Report Nov (doc and xlsx)
11:30 – 12:20	Technical Architecture presentation	50 mins	Yes
12:20	Other	10 mins	No
12:30	Close		

