Data Standards Workshop 22-23 May 2019

Summary

The AusSeabed Data Hub project team are in the process of developing the bathymetry and backscatter data processing pipeline and data storage requirements. The processing pipeline aims to take data, archive and process it to consistent quality standards, then deliver the data products to a discoverable database. The data will progress through a series of levels (from 'L0: raw' to L3: 'final product'), each of which requiring a definition of standards and specifications on formats, processing outputs and associated metadata.

This workshop aimed to leverage the large cumulative knowledge of the seabed mapping community to adopt, or establish, the best-known practices to future proof the AusSeabed Data Hub.

The objectives of this meeting were to:

- Establish consensus on the remit of the data hub (*raised and achieved during meeting*)
- Define data levels (*delivered*)
- Define and prioritise data formats, and their standards and specifications (*progressed*)
- Define and prioritise metadata fields (*progressed*)
- Define quality standards for both incoming and outgoing data (*deferred)*
- Understanding on the level of provenance required (*deferred*)
- Discuss processing pipeline requirements (*agreement reached*).

A number of actions arose during the meeting these have been taken out of the workshop notes and summarised in Table 1.

Table 1 Actions from workshop

Action	Description	Responsible party	Due date
1	Develop an extended business case promoting the economic benefit to collaborators. Include a *proof of concept* (GA, AHO, Defence) to show that while individual workflows are possible, they aren't efficient—a centralised data hub is a better solution.	ASB Steering Committee	TBC
2	Revise level definitions and remove any reference to proprietary software.	GA	Completed

3	Lobby the software companies (Eiva, Highpack, and QPS) to release their datagrams alongside their native formats (DB).	TBC	TBC
4	Choose a version for GSF.	ASB Data Hub Working Group	TBC
5	Discuss whether AusSeabed want to request a licence on a condition for [Native] formats so that we can put them into an open source environment.	ASB Steering Committee	TBC
6	Prove the ASB Data Hub concept initially with a decentralised network between GA, CSIRO, and AHO.	ASB Data Hub working Group	6 months
7	Discuss and propose funding model in time for AMSA.	Commonwealth collaborators	July 2019
8	Update common vision of AusSeabed data hub on website.	GA	July 2019
9	Tools group – develop scripting to automatically collect some of these metadata fields.	Guardian Geomatics, GA, FrontierSI	TBC
10	Strawman metadata need to be in line with ISO19115 and the AHO metadata requirements to form a minimum required list with the possibility to add some extra attributes to the metadata list.	GA	TBC
11	Collaborators to inform GA of the size of their bathymetry holdings to assist with the business case estimates.	All collaborators	TBC
12	Investigate the potential of the Discreet Global Grid System (DGGS). It manages, stores and helps visualise big data (petabyte scale). The Open Geospatial Consortium (OGC) officially adopted the DGGS as a new standard (OGC 2017c)	TBC	TBC
13	Investigate potential to apply for AODN/IMOS research Infrastructure funding ($ 20 M)	ASB Steering Committee	19/20 FY
14	Investigate Commonwealth proposal to get buy in for AusSeabed similar to the Copernicus data hub.	ASB Steering Committee	TBC
15	Establish a management/governance requirement of the open tool repository by creating an operation handbook (Processing pipeline and workflow discussion)	ASB Steering Committee	19/20 FY
16	Systematic approach to requirements formalised, software requirements specification to send out to group for comment	Natalie Lennard	TBC

17	Provide systems architecture to group (Just for knowledge, not comment)	Natalie Lennard	TBC
18	Notify AusSeabed if you/your entity would like to join the Data Hub working group for an update (via video conference) every 2 months.	Working Group	July 2019
19	Investigate whether JIRA through GovTeams is an option for the working group	Aero Leplastrier	July 2019
20	Minimum viable product of the Survey planning tool to be presented at AMSA.	FrontierSI/GA	July 2019

Workshop notes

Wednesday 22 May (day 1)

9:00 Welcome and housekeeping—Kim Picard (GA)

9:05 Data hub overview and the context of the processing pipeline—Kim Picard

9:15 Earth Observations Metadata Lessons Learned—Simon Oliver (GA)

Questions:

Simon Harries (AIMS)—How well does a [Metadata] system like this handle different levels; is the system able to cope with data that would be delivered at say level 3?

Simon Oliver—Yes, the metadata system can ingest metadata from any level.

Paul Kennedy (Guardian Geomatics)—Is the metadata held in the NetCDF or in a database?

Simon Oliver: Both, the database is for management and the NetCDF is for tools that work with the NetCDF data.

9:30 Activity 1 Data level definitions—Simon Oliver

OBJECTIVE: Define the data levels used to categorise data in the AusSeabed Data Hub.

OUTCOME: The group agreed on the data levels (Table 2) with the recommendation that any reference to proprietary software be removed from examples and definitions.

ACTION: 2

Table 2 AusSeabed Data level definitions

Level	Definition	Examples		
		MBES	Delayed Heave/ Ellipsoid/Nav	SVP
L0	**Unprocessed instrument data** At full resolution as received from the sensor. Includes MBES and ancillary files as well as any and all artefacts.	Observed by sensor *.all	Observed by sensor *.000	Observed *.raw
L1	**Data processed with ancillary information** Reconstructed L0 MBES data undergoes correction with ancillary information either from within the L0 data itself or the separately calculated ancillary files collected by the ancillary system (e.g. delayed heave and svp). This level may include radiometric and geometric correction and calibration, but not cleaning	Processed depth Integration of L1 ancillary information (uncleaned and unfiltered)	N/A data proceeds straight to L2	
L2	**Derived geophysical/ georeferenced variables** L1 data undergoes cleaning and filtering to create the first 'usable' multibeam data.	Bathymetry product Cleaned & filtered	Processed to SBET	Processed to *.txt
L3	**Variables mapped on a grid** L2 data undergoes additional processing/value-adding to create L3 products. Variables mapped on uniform grid scales, with some consistency to produce charts/gridded products etc. L3 products cannot be	Additional value added, or data sampled (e.g. chart, slope map, geomorphology)	N/A final format of data is L2 L2 is the final 'product' for ancillary data types, and not all ancillary data have a L1. For the majority of commercial software available, backscatter data is progressed automatically through the L1 and L2 stages and saved directly as an L3 final product.	

backwards engineered
into L2.

10:30 Activity 2 Data format prioritisation—breakout groups

OBJECTIVE: Develop and prioritise the list of formats at all levels for submission to the Data Hub and product delivery.

OUTCOME: Progress made by most groups.

ACTIONS: 3, 4, 5

Group 1 (Level 1)

MBES

Priority 1—XTF, GDF, (LiDAR, LAS/LAZ), NMEA (Open source)

Priority 2—GeoAcoustic SWF, GS Plus Flagged ROX, PDS2000, EIVA, HighPack HS2x, Qimera QPD (Proprietary)

Priority 3—ASCII Point cloud.

LiDAR

Priority 1—LAS/LAZ, GSF, XTF (Open source)

Priority 2—HDCS, Qimera (Proprietary)

Priority 3—ASCII XYZ.

SVP\CTD\XBT

No prioritisation chosen and need to have an exact match with AODN.

CTD casting: (FV00 and FV01)

FV00: raw data or data straight out from the instrument.

FV01: a pre-processed data with QC flags (no bad data has been discarded yet) for maximum transparency.

FV02: Good data only.

Calculated SVP from CTD (.csv)

Tide

Priority 1—Text Files: (ASCII .txt, NetCDF, csv) (Open source)

Priority 2—WSL, CARIS (proprietary).

Group 2 (Level 2)

This group discussed that the levels may be marginally different for satellite derived bathymetry compared to multi-beam.
Participants did not get around to providing specifications or prioritisations for the formats provided but are happy to do so in the future.

Group 3 (Level 3)

Specifications needed for ASCII include metadata for xyz data, header information in metadata, positive/negative depth field, and vertical datum.

Hang up points: specification/standards, quality of the data, authority to sign off on data, classification of the survey in the metadata, how to ensure quality of the data of the originator

Priority 1—ASCII XYZ (default delivery) and Coverage Shapefile (use open source geojson)

Priority 2—Contours (HOB), soundings (HOB, csv), point cloud (LiDAR, las laz, xyz, xyzt)

Priority 3—CAD

Priority 4—BAGS

Priority 5—Geotiff

Group 4 (Level 0)

MBES

Priority 1—All proprietary formats from big 3 sonar makers; Kongsberg (.all .kmall).

Reson, & R2Sonic. Appropriate datagrams as in strawman + backscatter + WCD.

R2Sonic (maybe via appropriately exported GSF file format).

Navigation and IMU (motion data) need to be added to the ancillaries' section.

Priority 1—Open source/format navigation data. (Unfortunately, none of this is really possible now) so priority 2 best option for now.

Priority 2—000 raw navigation and IMU files for post-processing to SBET.

Ancillary Data (Tides, SVP cast, CTD)

Priority 1—ASCII (txt, csv) of raw observations including georeferencing and time.

Backscatter Group (All levels)

Open data formats with documentation are preferred. Data Hub should be engineered to deliver multiple formats.

L0—add kmall and SAS data formats (kraken, Raytheon, HISAS, SAMs)

L1—add kmall, gsf, all, las, hsx. Discussion about auxiliary data and the need to include (e.g. beam form pattern files).

L2—add sonar scope

L3 Priority 1—BAG format (v1.6.4)

- Ensure floating geotiff.tiff is 32bit.
- NETCDF could provide some good options and alignment with IMOS.
- Remove hosting floating raster (*.csar) as it is proprietary.
- Recommend that we aim for OGC data standards or alignment. OGC geopackaging provides options to package both rasters and vectors in the one package.

11:30 Group discussion centred largely on the GSF format and concerns about format drip. Should review formats strawman in line with ISO 19115, the AHO minimum requirements and the formats document that Paul Kennedy and Wendy Stewart delivered at the workshop.

13:30 Remit of the AusSeabed Data Hub

OBJECTIVE: Define the vision for the AusSeabed Data Hub.

OUTCOME: Consensus achieved, and vision developed.

ACTIONS: 1,6,7,8

During the morning session it became apparent that there wasn't group consensus on the purpose of the AusSeabed Data Hub. This session was included to discuss the topic and bring clarity to the vision going forward.

The workshop described the Data Hub as:

A centralised server with link to contributors, with the future view of a centralised repository, governed by the data providers and owned by the community that delivers a standard suite of products and tools based on open source format.

The discussion was very positive with people advocating for a solution in perpetuity to future-proof against machinery of government changes that could see the AusSeabed platform dropped. Decision was made to prove concept initially with a decentralised data hub (between GA, CSIRO, and AHO). Once we have proved that it works, we can look to a more efficient and nationwide solution and on this basis move to a centralised system. This will be easier once we have shown the capacity to meet community expectations and people have seen the benefits and we have developed trust within the community.

Common vision

- All data levels should be deliverable
- Standard suite of products based on open source format
- Collect once, use many times (also the goal of HIPP, from L0 to L3)
- Data governance/ownership remains with the data providers

- Data Hub Governance and ownership: governed and owned by the community
- Both open data and moratorium data need to be included
- The data hub will also enable open tools to deliver standardised quality assured data.
- The Data Hub is required to have a backup plan, (login to be able to be swapped from state to federal?), use similar architecture for each

Other discussion points:

- WA will act as a distributed hub like CSIRO and AHO.
- The Copernicus Hub is a shared asset with a shared funding model ~$100k per organisation, is there the possibility of funding the AusSeabed Data Hub in a similar way?
- What are the benefits for private entities and state governments to encourage data submission?
    - Can provide data warehouse and remove costs on individual agencies on storing and delivering data
    - Empower more small enterprises to go out and use the data and gain local financial benefits
    - It will encourage more research and drive research grants
    - Benefits of independent validation by incorporating data with other standardised data

14:00 Activity 2 Metadata—Breakout groups

OBJECTIVE: Define a full list of metadata fields for each data level and agree on a minimum compulsory subset, with a prioritised list of the other fields based on the end user benefits.

OUTCOME: Metadata fields progressed and refined by groups, final decision that ISO 19115 will be adopted as the standard and then extra fields will be added to meet all requirements using AHO template as a minimum.

ACTIONS: 9, 10, 11

Group 1 (Level 1)

Essential attribute fields

- GPS data used.
- Horizontal Datum
- Vertical Datum
- Lineage
- Tide data (essential in relative referencing and irrelevant with absolute referencing)
- **S**ound **V**elocity **P**rofile
- **S**ound **V**elocity **S**ensor

- SBET
- Description
- Attitude assumed to be included optional (sometimes it is not included in single beam)

Group 2 (Level 2)

General notes

- There is a need to identify which metadata attributes can evolve from L0 (acquisition) to L1/2 (processing) and finally to L3 (production) like for example the vertical/ horizontal datum.
- S100 should be a minimum standard for metadata (*Ed. It was used to develop the strawman*).
- The need to have a quality hierarchy attribute algorithm that upgrade or downgrade for different characterised areas (rocky ground (high) versus sandy areas (low)) in geomorphological studies.
- There needs to be a quality/currency measure for old data that is missing some of the current day standards (i.e. data is worth something – just need to be clear to communicate what it is worth)
- There were identified metadata items that could not be inherited, that would need to be identified for L2 data.
- There was a significant number of items on the list that were non-standard—i.e., GA only activities

Non-essential fields

Number of bathymetry edits (%)

TPU (TPU for clarity should be split into Vertical and Horizontal uncertainty).

Group 3 (Level 3)

Essential attribute fields

- Survey resolution and remove the file locations.
- Survey multibeam plot file locations same as display file, remove one of them.
- Survey track line plot file locations
- Reports directory location/s
- Provenance
- Headers location with explanation/ example (BAG or HTF) format specific
- Visual Enhancement Backscatter?
- Gridding strategy is it the same as governance
- Units
- Date of created

- Remove Geospatial vertical min, Geospatial vertical max, Geospatial vertical positive and replace by bounding box.
- Citation
- Acknowledgement
- Disclaimer also add (Not for Navigational Purposes)
- License
- Provenance (Activity 5)

Extra metadata (mostly covered in Provenance list)

- Lineage
- Description
- Processors
- Role
- Process URL
- Software version
- Processing status
- Source Datasets
- Processed data locations
- Data quality
- Contact
- File format
- Horizontal Datum
- Vertical Datum
- Title
- Abstract
- Lineage/Provenance agreement on the meaning
- Collecting institution
- Investigator/Chief scientist
- Custodian
- System used
- Ship/Vessel
- System Beam width (1°X 1°) etc.

- Units (metre, feet)
- Type of data (Multibeam, single beam, Lidar, etc.)
- Feature type (grid or point cloud)
- File Format/type (ASCII XYZ, Geotiff, raster, etc.)
- Data Quality (Same as Provenance)

Non-essential fields

- Navigation file locations
- Ends-and-bends file locations.

The **Extra** (e.g., Horizontal and Vertical Datum, Lineage etc.) metadata attributes are not a duplicate in L3 but are considered essential in L3 products for stakeholders to consume. These metadata can change from the acquisition, processing to the production.

Group 4 (Level 0):

Group 4 accepted all the metadata of the strawman with no additions.

Backscatter Group (All levels):

- Look at using ISO-19115 (v3) as defining minimum standards of metadata fields and add field that will/can be requested in the future.
- KM: compile the methodology fields together, rather than split them up (system automatically generated).

Essential

L0

- Backscatter Data Type: Beam average or Time Series or both
- Backscatter Data Unit: dB or bytes or unknown
- Data type: Bathymetry/Backscatter/both
- Data File Name: Name of the data file that this metadata relates to

L1

- Backscatter Data Type: Time Varied Gain Boolean.
- Backscatter Data Unit: Absorption correction method (combo box to select).

- Data Type: The corrections associated with the area insonified by the beam.
- Data File Name: Was a correction applied to align the pattern of the beam to the predicted theoretical pattern.
- Backscatter Data Type: The method used to remove the angle of dependence.
- Backscatter Data Unit What statistical operator was used for the angle dependence removal.
- Data type: The number of pings used for the sliding window method.
- Provenance: The number of pings used for the sliding window method.

L2

- TPU: Total propagated uncertainty.
- Provenance: This field will capture all the information required to track the evolution of the raw data through to product delivery. For examples see activity 5 worksheet provenance.
- Visual Enhancement: Have you edited the image properties.

L3

- Provenance: This field will capture all the information required to track the evolution of the raw data through to product delivery. For examples see activity 5 worksheet provenance.
- Visual Enhancement: Have you edited the image properties.
- Gridding Strategy: What blending method was used to produce each cell.
- Units: The units of the mosaic.

15:30 Recap on Metadata—Group discussion

- Post process navigation (should be the name used for SBET) is actually a redundant field
- TPU was suggested to be separated out into TVU and THU.
- CSIRO have 15-20 TB Multibeam L0 data

Thursday 23 May (day 2)

9:00 Welcome and Day 1 review—Kim Picard

- Data Hub Group (AHO, GA, CSIRO, AIMS) will meet in late June to develop a business case and a presentation for the senior managers.
- Once this is developed it will be reviewed by the steering committee and then given to the senior managers.
- With the metadata there was a discussion about automated tools to harvest metadata. Without funding it will be difficult to get that automation developed.

RESPONSE: Paul Kennedy already has a line metadata tool that he is happy to share.

- Money has been spent on the processing pipeline – Simon is looking at metadata – his team has some commitment with a developer to investigate harvesting automatically as it has already been done with the Copernicus group.

RESPONSE: Dan Ierodiaconou—$ 20 M increase in research infrastructure to AODN/IMOS, can we apply for some of that grant?

ACTIONS: 9, 13, 14

09:15 Activity 5 Requirements for Processing Pipeline and Workflow— Simon Oliver

OBJECTIVE: Refine the requirements, workflow, and choice in processing system for the automated processing pipeline.

OUTCOME: The requirements for the Processing Pipeline and Workflow were met with general agreement by the working group, two other suggestions were made and appended to the table (9 and 10).

ACTIONS: 15, 16

This activity began with an overview of the requirements identified by GA for the processing pipeline (Table 3), the organiser then facilitated an open discussion session based on the following questions:

- Are there any other requirements that the processing pipeline should meet?
- Does the assessment of the processing systems acknowledge their various strengths and weaknesses?
- To get list of requirements regarding processing pipeline, e.g. develop tools that allow for processing to happen (open source).

Table 3. List of the processing pipeline and workflow requirements

ID	Business and high-Level Functional Requirements	Description	Suggester/Remarks
1	Output must have a measurable, definable uncertainty estimate.	The system should provide for knowledge around the following parameters: • standard error of ancillary inputs, • observed measurements Options considered (so far) 1. TPU	Put forward by GA. Kim: this point is captured by having a QC component in the pipeline. Need to mention inputs here, they need to be measurable etc. - PK

		2. CUBE	
2	Interoperable data output format	Data output format post-processing, but pre-product generation (e.g. gridding) needs to be interoperable, ideally open source and without any loss to quality (e.g. GSF) or ancillary data.	Put forward by GA. Supported without comment.
3	Automatable	• To ensure consistency • To deliver scalability • To automate corrections such as refraction correction that is currently manual	Put forward by GA. Supported without comment.
4	The process should be editable, scriptable, parallel-able, and customisable.	Individual preferences are to be catered for (where possible) in the processing pipeline “decision tree” to allow for different user scenarios	Put forward by GA. Supported without comment.
5	The process needs to cater for both new data forms and legacy data	For example: Modern day data sets include GPS Tide/Ellipsoid Height, while older data sets do not. The system needs to be able to bring these two disparate approaches together within the control of the user.	Put forward by GA. Supported without comment.
6	The system needs to be capable of assessing large areas of interest without the need to process segments individually	For example, A large area of interest that is lacking GPS tide/Ellipsoid Height may be best processed using multiple zone tide measurements. The system should be capable of choosing the appropriate zone tide measurement for a portion of the area of interest without the need to run individual processes. The same concept would apply to Zone SVP.	Put forward by GA. Supported without comment.
6	Report generation	This function gives a metadata summary post-processing. Useful for documentation purposes, if there is an automated tool already in existence, it should be used to properly document statistics.	Put forward by GA. Supported without comment.
7	The system needs to cater for variable types of corrections to achieve a common standard.	For example: 1. A common standard (IHO S44) can be achieved using either real time or delayed heave. The system should be able to cater for both scenarios (depending upon the	Put forward by GA. Supported without comment.

		information available for processing).	
8	Accessible Support	System design should include consideration for: • The need to remain up to date (version control/release schedule and appropriate associated support) • The need/requirement for day to day support and maintenance	Put forward by GA. Supported without comment.
9	Monitoring of the pipeline at different levels (the completeness/ failure of given tasks – maybe)		Suggestion raised by Guillaume
10	Understanding of the requirements for success at a given stage		Having the ability to see those statistics at a given stage might be a good way of deciding whether data needs to be revisited.

Comments:

- The tool needs sufficient information about the input going in so that the system can make a value assessment of the data. As such the metadata needs to be attached to every single sounding. This is done through a merge (for example).
- A QC tool that checks the completeness of submission (data, metadata and reports) and reporting on the completeness of the inputs.
- Assessment of the quality of what comes in and the capability to improve it.
- List of processing requirements needed to contribute to the processing pipeline.

10:45 Activity 6 Product requirements for gridded output — Kim Picard

OBJECTIVE: Define the workflow and the parameters of producing grids "on-the-fly" (interpolated/compiled or not).

OVERVIEW: Most users will want a custom gridded product. This may involve various types of data with variable extents. Therefore, the methodology surrounding how we integrate and interpolate the information to meet the request will need to be established. In this session, we want to explore the possible workflow and the parameters that we need to consider in building an automated gridding process.

OUTCOME: A good spectrum of responses was put forward by the working groups. These will form the basis for future discussions of gridding-on-the-fly.

ACTION: 12

Breakout group 1

For General user: (Default Attribute Grid) (DAG)/(BAG), computationally efficient and simple.

- Use: high frequency to look at data, quick check of data, provide basic tasks provide good estimates of uncertainty
- Take all point cloud points in a cell, every grid point in a cell treated as a bathy point
- NetCDF output, gridded value as mean, max, min, count in grid, STD of points, no data –equals null field.
- Grid computationally easy to make, no CUBE, no interpolation

For Expert user: scales, functionally adaptable over 10 years (input data sources, apply chain modifiers with default options (e.g. prioritisation module with input data sources, allows/deny acceptance based on criteria, each module to allow for change and input (weighted by user). Results in chained modules that provides a compound choice into surface creation.

- Outputs: grid type of what you need guided by module used
- Module-focussed based on what expert user needs (e.g. output focussed etc.).
- Create open source Application Programming Interface (API), ability to develop modules, upload and share modules on cloud, allow set up of default chains in modules. Ensure modules can be peer-reviewed.
- Caveats: not to be used for navigation

Breakout group 2

- Grid cell size was the number one priority.
- Use of International Zones of Confidence.
- No extrapolation but potential interpolation.
- If you are triangulating use the Delaunay triangulation.
- How to acknowledge/citations for items that have been created by gridding on the fly.
- Citations need to acknowledge both the owners and the funders. This should be added as metadata.
- AIMS (custodian of data) may have to renegotiate the usage constraint with their suppliers or funders (e.g. BHP as funder)
- Scientific citation – The need of (DOI) and persistent identifiers against metadata pages – also have the library providing a frame for citation format (if you are using this data then you have to put this in – still applicable for CC-BY.
- URL to deliver a detailed citation would be the best implementation (aka google maps

User story:

1. I want to see what is in my area
2. As a user what is the minimum quality data that is provided so I can choose
3. User selects the minimum quality of data in here:
   a. IHO Cat 1a – Highest quality followed by most recent
   b. If the user chooses “All” then give listing potential sources and allow the user to interact with these.
   c. Believe that once they see the product will see what works – hence the play with the inputs.
   d. Potentially some limited user choices for the non-expert user – leverage the user stories to define defaults
4. Default options:

a. Grid cell size – no 1 – automate the grid cell size to 1% of water depth as the "best available" grid size
b. Quality as a second tier (probably define a different quality style for each of the thematic styles)
c. Age of the data

- Believe that the b and c should be able to be user control prioritisation

Breakout group 3 (Level3)

Prioritized the grids scenario:

1. Product purpose and survey goal (final product usage, e.g. habitat mapping, seafloor gridding).

   -Two Interface: general/public (simplicity tool) versus expert users (more choice in parameters selection).

   -Explanation & terminology (disclaimers on use).

   Appropriate grid creation with the data used.

2. Time and Date of survey used (e.g. most recent data used to form the grid)
3. Uncertainty: (If uncertainty data is available and fit for purpose, choose the most appropriate datasets)
4. Footprint best resolution: (create a tool based on rules in the background, hidden from the user).

   The resolution is related to the following:

- Depth: (depth of the survey (shallow/ deep))
- Sensor (most appropriate sensor for the survey goal)
- Data type used in the gridding (single beam, multibeam, Lidar, etc…)
- Bedrock /habitat: (soft sediment, seagrass, igneous rocks that affect the return)

- Default should not contain any interpolated product
- Assessment between modelled data and actual data – need caveat.
- Citation: acknowledgement of data to organisations, URL (that includes all contributing parties/ organisations and link to report, unsure of onus of storing datasets that were generated from AusSeabed (on AusSeabed and/or general)

- Publications and data exported from AusSeabed (retained access by tool + metadata for usage statistics and the grids that authors used in their papers).

Breakout group 4

- Gridding on hierarchical basis – value judgement based on incoming data
- Uncertainty differs between organisations and surveys – AusSeabed to validate uncertainty in background
- Users: interpolation, background filler grid (can be misleading to provide modelled data – how should this be handled (source ID: package data, image data with no depth, for report writing purposes).

OTHER ACTIONS: 17, 18, 19, 20

12:30 End of Workshop—Kim Picard

Kim Picard—Thank you all very much for coming, we are happy that we have gotten some really good solid direction from a technical perspective. We have also firmed up the vision for the AusSeabed Data Hub, which is really positive. There are a number of actions that we have identified (action table was recapped).

Some things that will be coming up soon:

- Delivery of a requirements architecture
- System design documents for interest
- Call to add members to the data hub group
- Cloud task management space will be investigated (JIRA through GovTEAMS)

13:30 Optional session: Planning of QCtools Training week (18-21 June)—Kim Picard

As part of the AusSeabed QC Tool development (previously called QA4MBES), GA used this session to develop a workplan for visiting scientists from NOAA and CCOM with input from the community. The workplan was scheduled to include a three-day workshop introducing the various open source software and packages currently available, two weeks of development on an interface with automated QA checks (in collaboration with GA), and time to make up a presentation of the work and future direction for AMSA.

Appendix

List of Attendees

List of Attendees	
Wendy Stewart	AHO
Mark Case	AIMS
Simon Harries	AIMS
Guillaume Galibert	AODN
Dave Watts	CSIRO
Matt Boyd	CSIRO
Stuart Edwards	CSIRO
Magnus Wettle	EOMap
Shereen Sharma	Fugro
Aero Leplastrier	GA
Justy Siwabessy	GA
Kim Picard	GA
Natalie Lennard	GA
Michele Spinoccia	GA
Nick Dando	GA
Robert Parums	GA
Simon Oliver	GA
Wenjun Wu	GA
Adam Steer	GA
Paul Kennedy	Guardian Geomatics
Geoff Lawes	iXblue
Peter Locke	iXblue
Robin Beaman	JCU
Anna Meissner	LINZ
Brad Cooper	LINZ
Owen Cantrill	Maritime Safety Qld
Kevin McKay	NIWA
Neil Hewitt	Precision Hydrographic
Ralph Talbot-Smith	WA DoT