![Gladstone Healthy Harbour Partnership DIMS — Healthy Harbour, Healthy Future]

ISP009-2016-17: Provision of ongoing technical support for GHHP Data and Information Management System

October 2017

Screenshot of the Report Card System running the Environmental scripts for the 2017 Report Card

Eric Lawrey, Murray Logan, Lyndon Llewellyn, Gael Lafond, Aaron Smith

Australian Institute of Marine Science, Townsville

Executive summary

The GHHP Data and Information Management System (DIMS) was established in 2016 to help GHHP manage and host its website, the production of the annual GHHP Report Card using an automated Report Card System, and host legacy data information products. This project was to maintain and support the DIMS system, support the production of the 2017 Report Card and add trend plotting capability to the Report Card System.

The server administration during the year was routine with no significant outages or security issues. Operating system updates were applied weekly and a trial run of restoring from the backup system worked as expected.

The Report Card System performed well. There were no modifications or bug fixes that were necessary to the system itself. All changes were to the system configuration and scripts to make adjustments for updated indicators in the 2017 Report Card.

The DIMS team worked with the coral indicator and the social and economic component teams to integrate changes to these parts of the production Report Card.

A new trend plots script and feature was added to the Report Card System. This script generates trend plots of the Report Card scores and grades. This tool is aimed at allowing the GHHP science team to review changes over time, allowing potential issues and problems to be identified.

The addition of the trend plot feature required that the published 2016 Report Card was reproduced in the Report Card System so that trend plots would include 2016 in its plots. The official 2016 Report Card was produced manually outside the Report Card System. Recreating it required reverse engineering to track down the correct manual overrides, scripts and data.

The Report Card System is now well established for future runs of the Report Card with it being used for the 2017 Report Card.
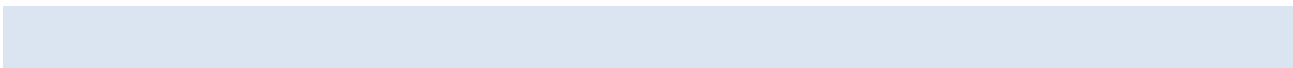
Work Completed

Task 2. Server Administration

1. *Operating System security patches are applied within 1 week of them being available. Security updates for other software on the server (GeoNetwork, Tomcat, Pydio, etc.) is monitored and security patches are applied with 1 week where there is a possible problem. Any issues arising from these patches are resolved to maintain the server's online presence.*
2. *Backup copies of the server (daily, weekly, monthly) are maintained.*
3. *System maintains a high availability (> 99%).*

The DIMS server is hosted on the Amazon Cloud using a server running the Linux operating system. Patches for the Linux operating system are released a few times each week. Each Monday these patches are applied to the DIMS server, bringing the server up to date. In addition the team subscribe to a number of security news feeds to monitor for serious issues that require immediate attention. During the 2016-17 year there were security issues that required deviations from the standard maintenance schedule.

The DIMS server has maintained a very high level of availability with no known outages of the service during the 2016-17 period.

The DIMS server is automatically backed up daily and weekly to allow the server to be restored should there be any issues. The backup system is implemented using a regular script on the server that instructs Amazon EC2 cloud to take a snap shot of the server each day. These snapshots can then be used to start a duplicate of the production DIMS server. The backup system was tested by starting a duplicate DIMS server from the daily backup. The restored backup was found to operate as expected.

No priority bug fixes were identified as part of this project and so the time associated with this subtask was used to cover the additional time required to prepare the server for the trend plots (see Task 7)

Task 3: Getting data into DIMS

1. *Provide assistance to Data Providers (including training, training material, walk-through, etc.).*
2. *Adjust input validation rules and improve validation API to improve robustness.*
3. *Update DIMS manual to reflect improvements.*

The aim of this task was to cover the time associated with assisting data providers with uploading their data into the DIMS system. This involved resolving issues associated with ensuring the data was added to the system for the 2017 Report Card.

No changes to the data upload process was necessary, however a new simplified data upload guide was developed and made available to data providers (see Appendix 1). This is an extension to the existing DIMS manual.

Note: Most of the datasets (all except seagrass) were available after the official end of this project (June 2017), however the project had budgeted time towards this activity and so this time was put toward ensuring the data was available for the 2017 Report Card. The following outlines the issues associated with each data source:

- Seagrass – No issues with data upload.

- Coral indicators – Code and data changes in 2017 covered in task 11. No issues with the data upload.

- PCIMP – Coordinated with Greg Coleman and the ISP recommendations provided by GHHP science team to upload the data into the system.

- Connectivity – This data was provided as spreadsheets and not in the required form. The team provided recommendations for the required data format to be provided and for how they could be added into the system without them contributing to the final scores (using the weights files). The team did not receive the connectivity data in the correct form by the end of the project. The final data set for connectivity has been uploaded by the science team.

- Social and economic data – There were code changes and additional data files that needed to be added to the system. This took the bulk of the allocated time on this task. Further details are provided below.

Social and Economic code and data

An updated version of the code and data for the Social and Economic components of report card were supplied to the DIMS team for integration as part of the 2017 report card. Changes to the existing code base were integrated and tested with the supplied data. The final integrated version was provided back to the Social and Economic team.

The following provides detail on this activity:

A review of the updated version of the social and economic code supplied by Jeremy (31 Aug 2017) was performed. This assessment and integration into the Report Card System was made more difficult because the code changes were made to an old version of the code prior to its integration into the DIMS. Numerous changes to the plotting code and functional changes to E3_Economic_Performance were identified. The changes to the E3_Economic_Performance included a new dataset: GPC_revenue2017.csv. This dataset was used to remove some of the hard coded constants in the original code base, making them driven by the new dataset rather than constants in the code. The new dataset was added into the 'shipping activities' dataset MANIFEST and validation rules setup.

After the code changes were integrated a test version of the social and economic scripts was setup on the Report Card System and the 2017 data uploaded and tested. Adjustments were needed to most of the dataset files to include Date and Year columns. These adjustments were made to the file formats in 2016 to allow the Report Card System to provide validation on the age of the data. This was to ensure that old data cannot be accidentally uploaded from year to year. These date fields also allow the data from multiple years to be combined together in the future if needed. The supplied 2017 data was sent through without these modifications and so were rejected by the Report Card System. The data files were corrected and notes provided to the social and economic team so that future submissions will not require these adjustments.

Adjustments were then made to the scripts to correct for incompatibilities between the version of R used by the social and economic team and the version available on the DIMS. This included adjustments to the plotting code, specifically changing units() to grid::units().

Once the new code was integrated and tested on the Report Card System, it was further adjusted so that it could also be run on Windows (the DIMS runs on Linux) and so each of the individual script parts (S1, S2, S3, S4, E1, E2, E3, E4) could be run independently of each other. This was not necessary for running the code on the Report Card System, but was done so that the Social and Economic team could run the new code with ease on their systems, allowing them to make future adjustments to the latest code.

Task 15: Project management and reporting

1. *Inception meeting with GHHP*
2. *Project management*
3. *Final report*

This project was a continuation of the first DIMS project, except this time it was to maintain and support the system. The project consists of a range of routine maintenance tasks, and one-off tasks to improve the system. Each of these tasks were split across the team members based on their previous experience and matching skills. The project team leadership met with GHHP via teleconference on the 2nd May 2017 to clarify the tasks that were needed over the next year.

Due to the late start to of some of the project work due to contract delays (see lessons learnt) additional tasks outside the scope of this project arrived to support the production of the 2017 Report Card. These included improvements to the developer workflow (using GIT directly on the server) to facilitate rapid turn-around of script changes, integration of changes to the fish indicators and integration of mud crabs. These tasks were performed to ensure there were no delays to the 2017 Report Card. These tasks will be included in the next maintenance project.

Lessons learnt

Contracting for this maintenance project dragged on for 11 months with significant back and forth between the parties and a significant restructure of the contract 2 weeks prior to the end of the new project end-date. The original proposal was requested by GHHP to be split from calendar years to financial years, splitting the original work into two contracts instead of the original proposal. This meant that the official end date of the project ended up only a few weeks from when the proposal was restructured and as a result the final contract was signed after the official end of this project.

This delay in contract signing also meant that most of the one-off tasks were only started after the official end of the project. While there was significant disconnect between the timing of the project work and the contracting AIMS appreciated the importance that the service was maintained and that the Report Card needed to be completed on-schedule. Most of the issues with contracting stemmed from details of one-off tasks, rather than maintenance tasks.

Task 7: Trend plots

1. *Creating/adjusting scripts to collate relevant data and provide user with means to access the results.*

This task involved adding the ability to view a trend plot of scores of the current and historic report card years. Preliminary work was done on presenting trend plots in 2016, during which Murray Logan developed a prototype R script for generating the trend plot graphs and the DIMS team added support in the Report Card System to collate scores and grades over multiple years to allow trend plots.

This task involved extending on the existing work to integrate the trend plot script into the Report Card System so that the administrator could run the script and review the graphs.
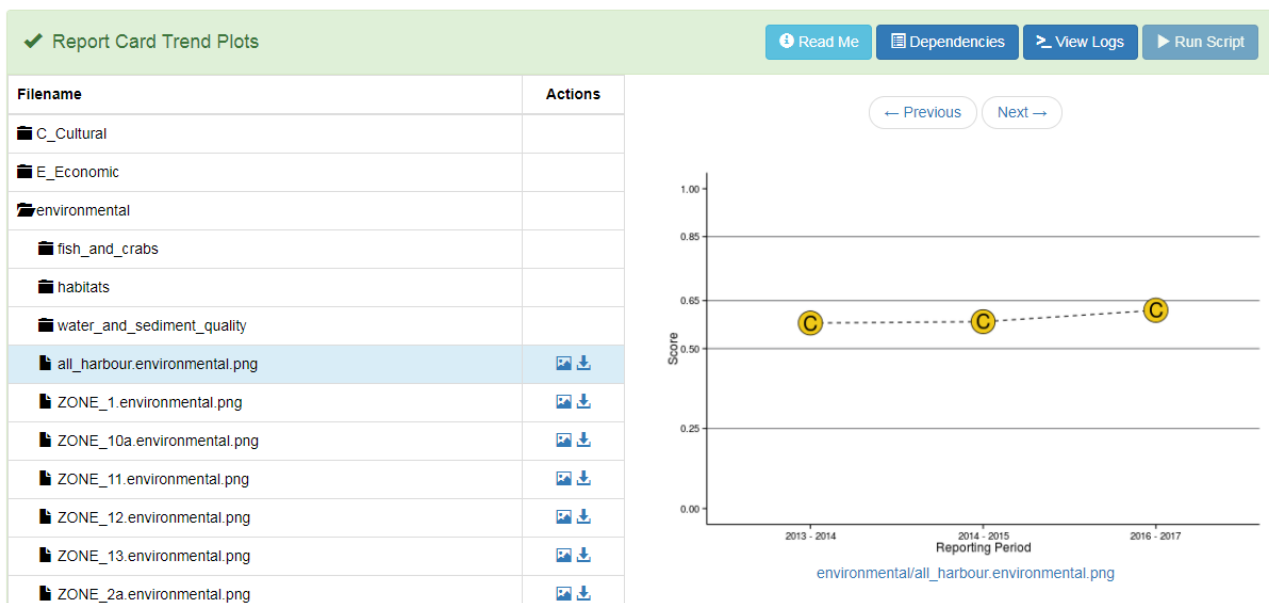

Figure 1. Final implemented trend-plots script as run on the 2017 report card.

The trend-plot script generates plots for all indicators in the system (approximately 2000) that have more than one year of data available. It presents them to the user as a directory tree structure that is organised to match the structure of the Report Card. It was implemented using a directory tree, rather than a specialised user interface to reduce the development time, as the system already supported the ability to browse the files and directories generated by a script.

The Report Card System largely treats each report card year independently from each other allowing the system to be modified over the years without having to worry about breaking previous report cards in the system. The system does however track the report card components and indicators across years by assigning a unique identification to each of them. These IDs are then used to collate all the related scores and grades over time allowing the production of trend plots.

The Trend Plot script module uses the collation feature of the Report Card System to request all the scores and grades from the current and historic production report cards and make them available as an input dataset to the script. The trend plot script then uses these scores and grades to generate a plot for each component and indicator.

The Trend Plot script module can only produce plots for Report Cards that have been run or are recorded in the Report Card System. This presented a problem because the official 2016 Report Card was not run in the Report Card System but instead run external to the system. Additionally some of the older (2014) social and economic indicators were not entered into the system using the same IDs as later report cards. It was therefore vital that these issues were resolved, otherwise the new trend plot feature would have no proper time sequence of data to plot.

Making 2014 social and economic indicators display in the trend-plots

In order for trend plots to be useful they need several years of data. In the Report Card System indicators are tracked across report card years using IDs. Each indicator is assigned a unique ID that should not be changed over time. Unfortunately most of the scores and grades that were developed in the 2014 pilot report card were developed prior to the Report Card System and so these indicators were given temporary IDs. In 2016 most of these indicators were assigned permanent IDs when the scripts were finalised. In order for the trend plots to display the 2014 social and economic report card results this task involved going back through the record of the 2014 Report Card and update the IDs in the Report Card System's scores database.

Making 2016 report card values be included in the trend-plots.

While the DIMS was setup to run the 2016 report card it was completed so close to the cut-off date that the final official Report Card generation was developed external to the Report Card System. As a result at the start of this project there was no official completed Record Card for 2016 in the Report Card System. Unless this was resolved then the trend plots would not include the 2016 scores and grades.

Unfortunately reproducing the published 2016 report card in the Report Card System was more difficult and time consuming than expected because there was no ready record of the exact scripts, manual overrides, PCIMP data exclusions used in the generation of the Report Card external to the Report Card System. To make things even more complicated the final copy of the 2016 PCIMP data had been deleted (as per the data sharing agreement) from the environmental script developer's (Murray Logan) machine. It took significant reverse engineering and discussions with Murray, and Greg Coleman (PCIMP data representative) to reconstruct the 2016 report card in the DIMS.

Task 11: Coral cover indicators

1. *Adopt/modify r-scripts to calculate the coral cover change indicator as required.*
2. *Integrate the r-scripts for the coral sub-indicator with existing DIMS workflows.*
3. *Document details of task in the final report*

This task involved integrating adjustments to the coral cover indicators into the environmental script. The DIMS team worked with Angus Thompson (Coral Indicator team) to incorporate his code changes into the main environmental code base. The setup of the Report Card system (dataset MANIFEST files) was also adjusted to handle and validate the new additional attributes added to the coral.csv data file. The code and MANIFEST changes were then tested in the Report Card System, then once finalised they were migrated into the production version of the 2017 Report Card.

Appendix 1

This guide was developed to assist data providers with uploading data to the Report Card System for the 2016 Report Card.

Quick guide to uploading data into the GHHP Report Card

Eric Lawrey

This is a quick guide for uploading of data into the DIMS. Datasets need to be uploaded and published by each of the data providers before they can be used to calculate the official report card. It is important that data providers upload the data themselves as it ensures that the correct version of the data is uploaded to the system.

1. Go to http://dims.ghhp.org.au and login using your DIMS username and password. If you have forgotten your username or password contact the System Administrator (currently Uthpala Pinto) to reset your password.
2. On the left are the GHHP Report Cards. Find the current Report Card year. In this example we will use the 2017 report card. Click on the Datasets link.
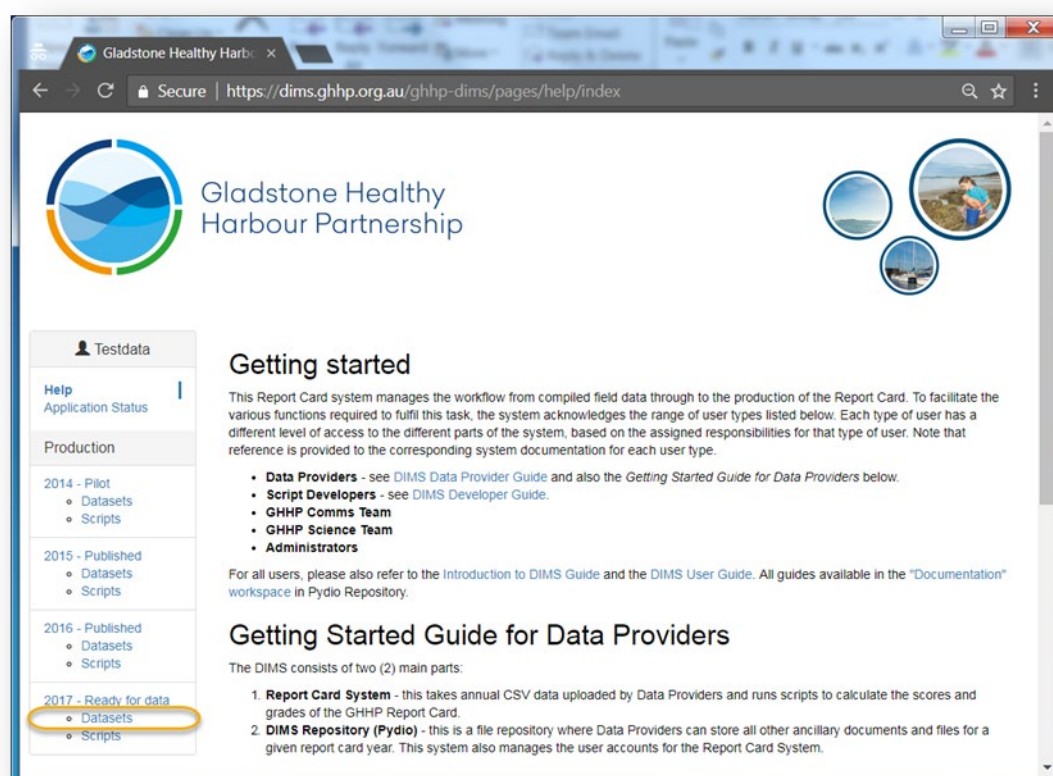

3. You will see the datasets that you are responsible for providing. The list of datasets will be different for each data provider.

Note that in this example there is a dataset (Index of Economic Resources) that already has some data setup. This is because this dataset has baseline data automatically copied from the previous Report Card year.

4. If we click on one of the datasets (Commercial fishing in this example), we can see the list of files that need to be uploaded for that dataset.

5. For each file in the dataset upload the latest version of the data into the system. You can use drag-and-drop to put the file in, or use the Upload button.
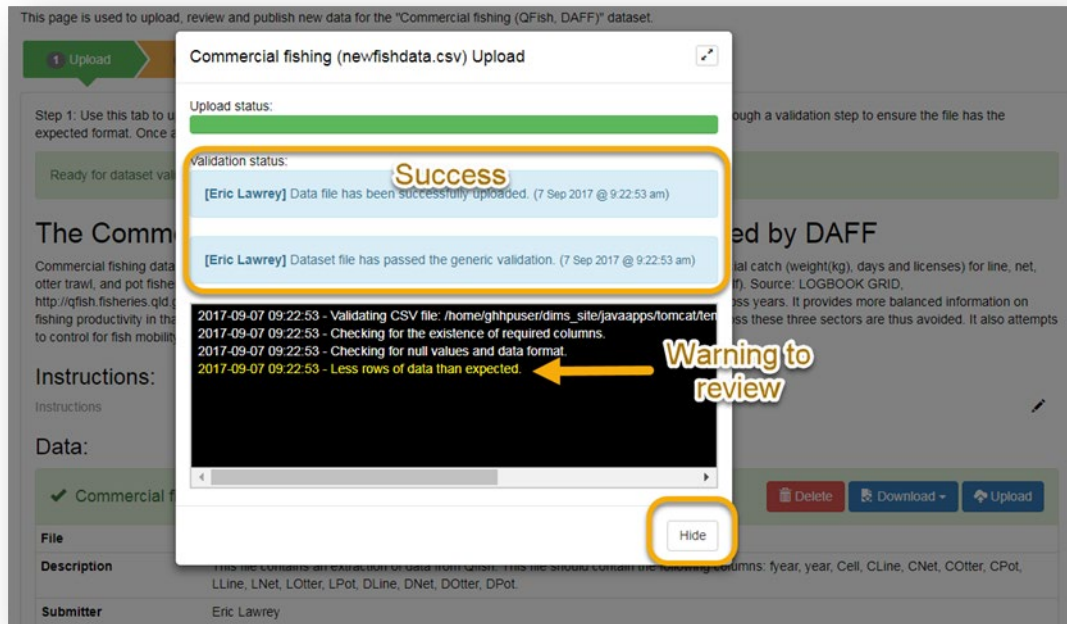
As each file is uploaded it is checked against a number of validation rules that confirm the format of the file. These rules have been based on data from previous years and so any changes to the file structures may result in warnings or errors. If there is a warning then you must verify that it will not cause a problem for the processing scripts. If there is an error then you have most likely created a significant change for the file format and the scripts will need adjusting or you have uploaded the wrong file.

6. Once all the files for a dataset have been successfully uploaded, as indicated by the "Upload" tag changing to green you can move to the "Validate" step by clicking on this stage.

7. The Validate stage generates summary graphs for all the files in the dataset. It can also be used to run more advanced validation scripts on the entire dataset. This feature is however not used in the 2017 report card. To complete the Validation click the "Run validation script" button.

Australian Government

AUSTRALIAN INSTITUTE
OF MARINE SCIENCE
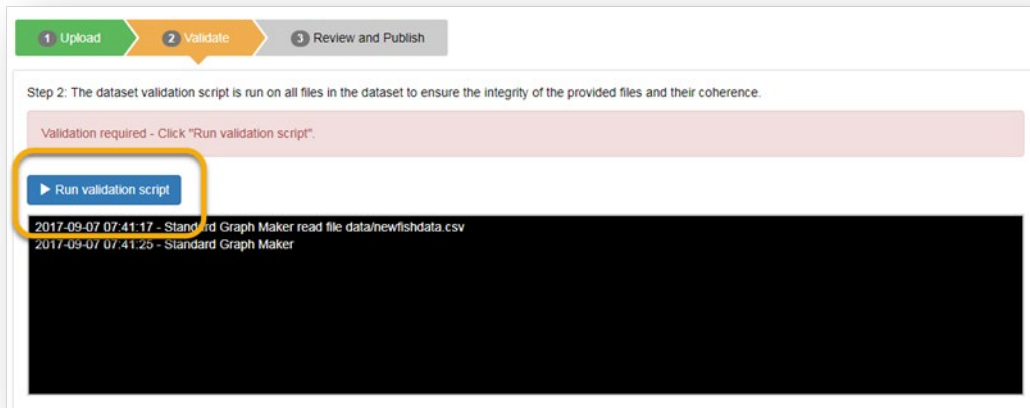
8. Once the Validation has been run successfully we can move onto the Review and Publish step.

9. On the Review and Publish step you can review the graphs generated by the system. These graphs are automatically generated from all attributes in the uploaded files. Some of the graphs will seem a little strangle because the system does a pretty dumb plot of all the variables. The aim of these plots is to review the data for any obvious mistakes.
Once you are happy with the data then click the Publish button. This makes the data available for the Report Card scripting. If any adjustments are necessary later you can always unpublish the data to correct any errors.

This page is used to upload, review and publish new data for the "Commercial fishing (QFish, DAFF)" dataset.

1 Upload → **2** Validate → **3** Review and Publish

Step 3: This page shows the artefacts generated by the Dataset validaton script. You are required to review these artefacts for any unexpected anomalies. The generated files may be downloaded for offline examination, such as by specialised software. When satisfied with the validity of the data, "Publish" the Dataset to make it read-only. Only Published Datasets can be used to generate a Report Card. The Dataset can be un-published if required.

The dataset is not published

Filename	Actions
newfishdata.csv.graphs	
newfishdata.csv.boxPlot.pdf	⬇
newfishdata.csv.histogram.pdf	⬇
newfishdata.csv.scatterPlot.pdf	⬇

← Previous   Next →

page 5 of 14
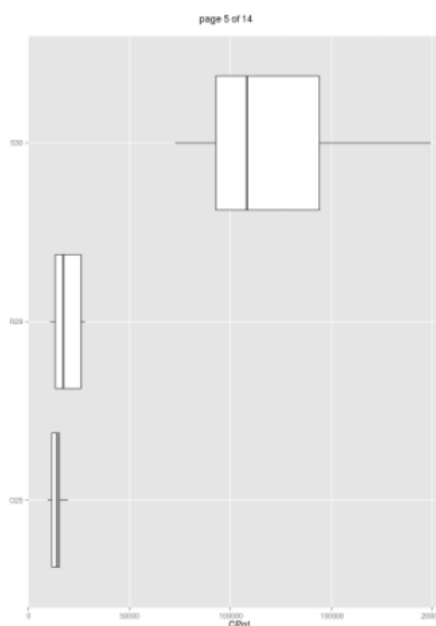
newfishdata.csv.graphs/newfishdata.csv.boxPlot05.png

Publish

Once the dataset is published it will be all green in the dataset list:

1 upload → **2** validate → **3** publish   Commercial fishing (QFish, DAFF)

10. Complete the upload process for all the datasets that you are responsible for.
11. Let GHHP know that you have completed the process.

DIMS Report Card System   2016-17 Final Report

Australian Government

AUSTRALIAN INSTITUTE OF MARINE SCIENCE