ISP009-2018-19: Provision of ongoing technical support for GHHP Data and Information Management System

Final Report

4th July 2019

Author: Eric Lawrey, Murray Logan, Marc Hammerton, Australian Institute of Marine Science

Project team: Eric Lawrey, Aaron Smith, Marc Hammerton, Gael Lafond

Date: July 2019

Version: 1

Final report

The aim of the ISP-009 DIMS 2018-19 maintenance project was to ensure that the GHHP DIMS report card system was maintained and updated to allow the production of the GHHP Gladstone 2018 report card. This included a range of maintenance tasks, one-off tasks to enhance the system and general assistance with the GHHP website. The maintenance tasks were:

- To ensure that the servers kept running and were secure (with Operating System patches). (Task 6.1). This included assisting with outages due to the GHHP website, upgrading the Statamic Content Management System (CMS) and assisting with providing access for the new GHHP web developer contractor.

- Priority bug fixes/changes were made to the system. This included extending the Report Card permission system to allow PCIMP data to be included in cloned report cards, whilst still having restricted download rights to be compliant with the dataset licensing.

- Assistance was provided to data providers, where necessary, including slight adjustments to the seagrass scripts to improve their robustness (Task 2).

The project also included one-off tasks to further expand the capability of the report card scripts. These included:

- Integrating the fish indictor code into the environment script (Task 5).

- Adding support for including mud crab scores in the report card (but not the raw data processing code) (Task 6).

- Adding support to allow individual PCIMP measurements to be removed from the report card (Task 4.3).

- Improving the management of the script code in GitHub to allow code changes to be tested and put in production easily (Task 4.1 and 4.3).

Details of each of these tasks is outlined below.

Background

The GHHP Data and Information System is a system for managing the generation of the GHHP Gladstone Report Card. It aims to fully automate the processing of raw or near raw data through to the final scores and grades of the Report Card. This automation helps ensure that the provenience of the data and subsequent processing is fully recorded making it possible to track back from the Report Card scores and grades to the original data.

The initial Gladstone Report Card System was developed from 2014 – 2016. During this time the system and the Gladstone report card indicator structure and logic was also developed. A pilot report card was released in 2014, followed by a more complete version in 2015. Each year there have been new indicators added to the report card and the DIMS Report Card System. In 2016 the DIMS Report Card System and associated scripts were complete and integrated, but not in time for the published report card. In 2017 the Gladstone report card was produced from the results generated by the DIMS – Report Card System for the first time.

The GHHP DIMS is now a maintenance project with only small aspects being adjusted each year, typically to include new indicators that have been developed.

The GHHP DIMS consists of off-the-shelf software components (Document Repository) and customs pieces of software (DIMS Report Card System, the analysis scripts and the reporting scripts). The structure of this system is shown in Figure 1.

Data Providers upload their data into the Report Card System, these are then processed by the analysis scripts to calculate the scores and grades from the raw data. These scores and grades are then converted to a series of intermediate reports (in Word format) to be reviewed by the GHHP science team and the Independent Science Panel (ISP). Once reviewed and any issues with the data are resolved, the final documents are sent to the communication team to be presented via the public website. In addition to this a Technical report is produced by the GHHP science team. To assist in this process the Report Card System generates a template of the Technical report (in Word format), with pre-populated tables and graphs. The GHHP science team then add synthesis text explaining the results and the report card process in detail.

While the *Content Management System* for the public website is hosted on the same server as the *DIMS – Report Card System* and the *Document Repository* the two sections (as indicated by the dashed line in Figure 1) are managed by different teams. AIMS manages the server itself, the *Document Repository*, the *DIMS Report Card System* and its associated scripts. The *Content Management System* is managed by a third party contracted by GHHP.
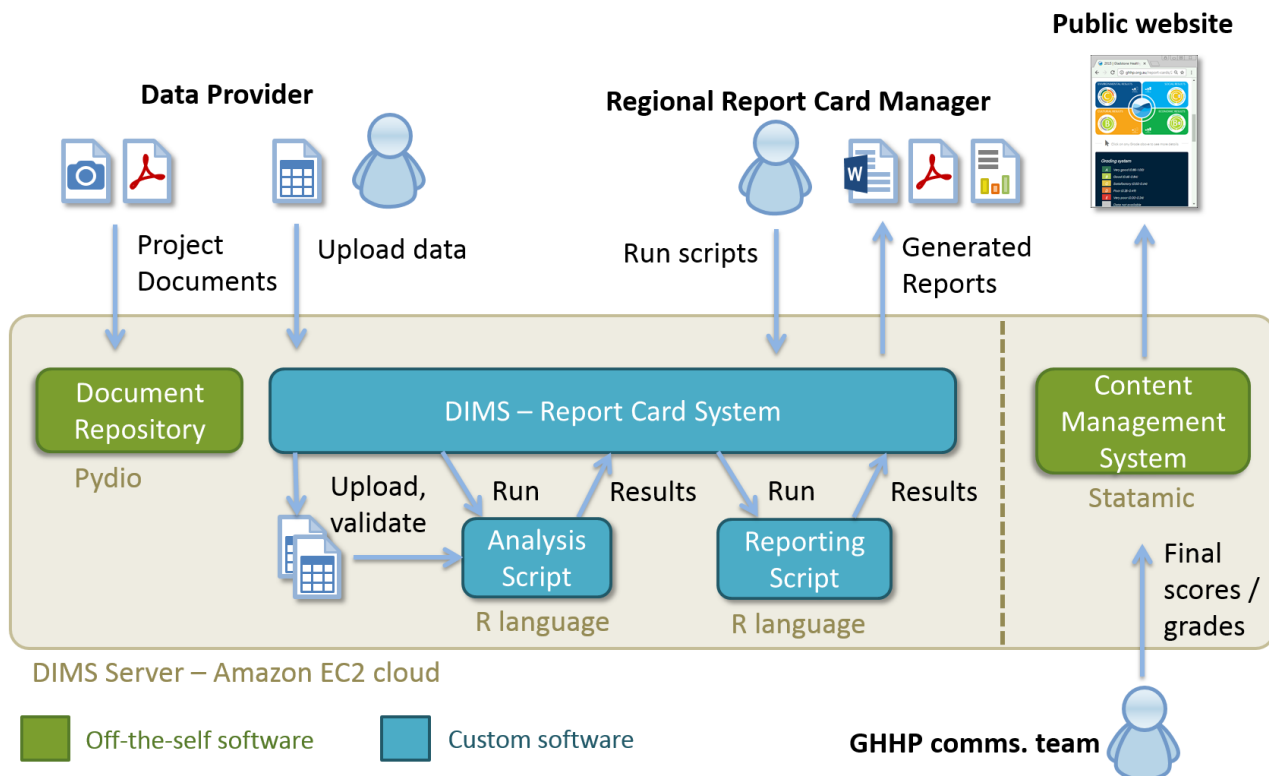

Figure 1. Structure of the GHHP Data and Information Management System (DIMS). Note: The Content Management System is run on the same server as the DIMS but is not part of the DIMS.

Routine tasks

Task 6.1: Server Administration / Bug fixes

The aim of this task is to maintain a secure server for the GHHP report card products. This task will involve:

•   Ensuring that the DIMS server remains online by promptly resolving any issues that lead to an outage. The team is responsible for the server, operating system and the DIMS system (Report Card System, Pydio and http://data.ghhp.org.au/), but not the Content Management System hosting the GHHP website (http://ghhp.org.au) as this is managed by a third party. Where there is an outage of the website we will assist where possible.

•   Ensuring that the DIMS server is backed up and security patches are routinely applied to the Operating System and other system software including Apache, Tomcat, Pydio, etc. This includes testing to ensure that updates and patches do not affect the performance and operation of the server.

Completed April 2019

Routine Operating System Patches

Routine Operating System security patches were applied to the server on the Monday of each week. These ensure that all the packages on the server are up to date with the latest security patches. The automated backup system of the server (snapshots in Amazon EC2) was tested 3 times during the year by starting one of the backup images to ensure that they would run.

Improved backup schedule

The length of time the backups are retained was extended from one month to 2 years. This was done because some parts of the DIMS system are infrequently updated (one per year) and so an issue may not be noticed for more than 12 months. Extending the backups to 2 years will allow comparisons between the current state and the previous state which may be key to resolving an issue. A review of the Amazon costings for snapshot storage (which is used for the backups) indicates that this increase in the number of backups should not significantly increase the backup storage cost. This is because the backups are stored as a difference between each backup and the previous backup. If nothing has changed on the server then new backups take close to zero storage and thus minimal additional expense.

GHHP website outages and Statamic CMS upgrade

This year saw several issues with the GHHP website that required assistance from the DIMS team to resolve. The GHHP website runs on the same server as the DIMS Report Card System but is not one of the systems that the DIMS team is responsible for. GHHP contracts out the maintenance and updating of this website to external contractors. These contractors are provided access to the server to do their work. Their role is to update the website each year with the new Report Card content.

Unfortunately changes made to the website in 2017 led to some instability issues with the CMS during the year. This became increasingly worse with each new update to the content, with the site failing to allow content for the 2018 Report Card to be added without the site crashing. Initial reports of the problem were that the site was running out of memory and slow. While GHHP looked to find a Statamic developer that could resolve the underlying problem a temporary work around was put in place by the DIMS team. The memory allocated to the website was increased 20x above the normal amount and the server size was boosted to compensate for the slow performance of the website caused by the underlying website issue.

Outages were still occurring taking the GHHP website and the DIMS offline regularly. GHHP had trouble sourcing an available Statmic contractor to resolve the issue and so the DIMS team investigated the issue to assist GHHP. Brad Hansen, BitPlex, suggested that the issue might be resolved by upgrading Statamic, but he was unable to commit the time to test and implement this upgrade.

A development server (duplicate of the GHHP website) was setup and the DIMS team determined a set of actions that would reproduce the problem so that we would know when the problem was fixed. Once we could repeatedly cause the failure, an upgrade to Statamic (2.1.19 to 2.11.2) was performed. The site was then retested and found to no longer have the issue. This change was then applied to the production server.

Part of this process required the team to research and understand how Statamic works to sufficient depth to safely perform the upgrade procedure and in the process understand what changes need to be made to the website in each Report Cards each year. To assist those working with the GHHP Statamic in the future we developed two videos explaining what needs to be done to setup new Report Cards in Statamic each year. See Task 7 for more details.

Assistance was also provided for the new web developer (Brad Hansen, BitPlex) to allow him appropriate access to the server so that he could work on improving the GHHP website. This included creating a server account and improving the internal security on the server so that the web developer could not access for delete parts of the Report Card system or data. (Nov 2018)

All changes to the server were recorded in the Admin-log.txt file on the server (see https://dims.ghhp.org.au/repo/ws-documentation-adminstration/Unix administration/Admin-log.txt)

Other maintenance and server tasks

- The digital certificate for dims.ghhp.org.au was converted over from GoDaddy to Let's Encrypt. This will eventually save approximately $300 pa on certificate renewals and eliminate having to renew the certificate each year.
- Murray Logan was given access to push code changes into the production report card. This was done to reduce the delay between script changes and having them in production during the rush because the finalisation of the report card. As part of this the process for pushing script changes to production was documented (See https://dims.ghhp.org.au/repo/ws-documentation-1/DIMS-Developer-Updating-production-report-card-2019-06-26.pdf) (August 2018)

Task 6.2: Report Card System priority bug fixes and improvements

This task involves minor adjustments to existing scripts and the Report Card System to improve its overall robustness and streamlined generation of the report card. The particular bugs and improvements to be resolved will be negotiated with GHHP during the project. As the full list of potential issues is not known at the start of the project this task is arranges as a time budget. The team will work with GHHP to determine which priority modifications can be made within the task time budget.

This task involves making minor adjustments to existing scripts and the Report Card System to improve its overall robustness and streamlined generation of the report card.

Task 6.2.1: Assisting with integrating changes in the social and economic scripts into production

Completed September 2018

There were a range of changes to be made to the social and economic scripts. Dr Jeremy De Valck provided a list of changes to the R scripts as a MS Word document. The team reviewed all the changes from the document and adjusted them to be compliant with the production code. After integrating the changes they were tested and pushed into production.

Task 6.2.2: Better aligning the generated outputs of the ISP scripts with the documents that need to be sent to the ISP, web team and technical report

Completed March 2019

The output files of the ISP scripts were compared to the published technical report 2018, report card 2018, and the GHHP website (http://ghhp.org.au). To better align the generated documents the following adjustments to the ISP script were made:

- Headlines added and renamed
- Missing tables and figures added
- Table and figure captions modified
- Table column order adjusted
- Table column headings and row labels renamed
- Table and figure numbers adjusted
- Content rearranged

Task 6.2.3: Minor change to aggregating the seagrass scores

A minor change is required to aggregating the seagrass scores. Dr Alex Carter will be available to assist Dr Murray Logan during the rule update. An example of the application of this rule for 2017 report card is in Appendix-B. This update is required for the 2018 report card and should be completed by 31 of August 2018.

"If the lowest of the 3 indicator scores is the area or biomass score, then the overall score = the lowest of the 3 indicator scores (the current system). If the lowest of the 3 indicator scores is the species composition score, then the overall score = the average of the species composition and the next lowest score (area or biomass)"

Minor changes to the seagrass scores code were conducted in consultation with Dr Alex Carter.

Task 6.2.4: List of priority minor changes

Completed March 2019

A list of priority minor changes to the DIMS were provided by GHHP to the team in 'Minor issues to be updated after the release of 2019 report card.docx'

These included:

1. Minor adjusting to the GHHP short report generated by the environmental script including: changing the copyright notice, adjustments to some tables and figures and adding a table to report on Limits of Reporting (LOR).
2. Fix a missing image in the report card output file.
3. Add tables and figures for the indigenous cultural heritage indicators and updating the statistics table in the technical report document.
4. Add a table with the results for the fish recruitment to the web material output file.
5. Add tables and figures for the indigenous cultural heritage indicators to the web material output file.

Each of these changes were made to the scripts.

GHHP Short Report – 7. For all dot plots in the report 'b' zones need to be removed. Site names should reflect the current site names used by PCIMP.

The Short Report uses the names provided in the data supplied in the upload. If the names are modified the script will work, but the link between sites across years will break. The system therefore needs the following adjustments.

- replace *lookup.csv* with */modules/labels.csv* in the environmental script and remove *lookup.csv*. This is to make it easier to apply new ID aliases.
- GHHP_write_scores to perform inline site name replacements (to allow cross linking of new names to old names)
- *labels.csv* needs to have new_name and old_name columns added to enable GHHP_write_scores to generate the consistent ids.
- Need to get a list of old and new site names from Greg
- Greg to supply PCIMP data with new names from now on.

Improvements to Report Card cloning to allow trial runs with PCIMP data

In addition to the minor changes listed above a request was made to improve the workflow associated with testing new Report Cards due to restrictions associated with the PCIMP dataset.

The DIMS Report Card System allows any report card to be 'cloned' into the user's own workspace on the server. The 'clone' can then be used to perform test runs with the existing uploaded data, or the code modified to test new version of the scripts, all without affecting the production Report Card. They are also useful for performing tests of code changes prior to pushing them into the production report card.

GHHP also found this feature useful for performing trial runs of the report card, particularly with different override settings.

The restrictive licensing of the PCIMP dataset did however introduce a complication in this workflow. PCIMP data requires very tight permissions around access to this dataset. Once uploaded the PCIMP cannot be downloaded from the DIMS Report Card System. To ensure that users could not access the PCIMP data via a clone the original implementation of the permissions system in the DIMS would simply delete the PCIMP data in the clone. While this approach ensured the system met the obligations under the dataset licensing if also prevented the 'clone' feature from being able to be used to test a trial run of the Report Card with the PCIMP data.

To overcome this limitation the permission system in the DIMS was changed slightly so that the 'clone' operation would copy the PCIMP, but that the user could still not download the dataset from the clone.

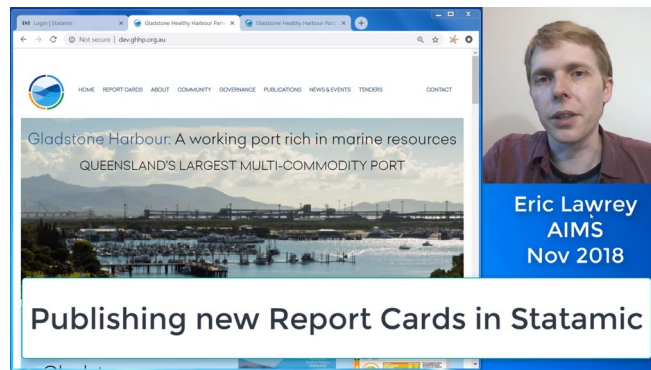## Task 7: Updating the training manual and final report

This task requires documenting all work undertaken during this project and updating the DIMS user manual as necessary to reflect any updates to the system.

The final report will provide a summary of activities over the life of the project.
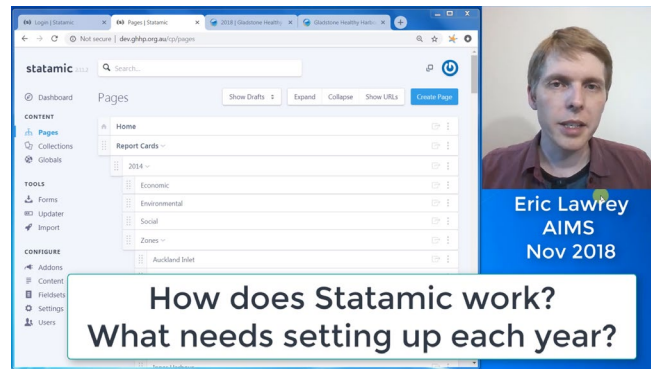
This task also includes time associated with the project management task involves the coordination and scheduling of project team members to achieve the project tasks and occurs over the life of the project.

There were no significant changes to the DIMS Report Card system code during this project and so there were no changes to the existing user and developer guides.

As part of resolving the stability issues with the GHHP website (see Task 6.1) the team gained some knowledge of how the website worked. This was documented in two videos:
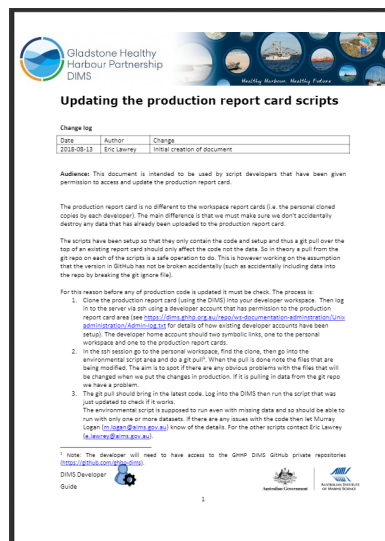

https://dims.ghhp.org.au/repo/ws-documentation-1/Publishing-report-card-website.mp4

https://dims.ghhp.org.au/repo/ws-documentation-1/Publishing-report-card-website-statamic-in-depth.mp4

The process for pushing script changes into production was documented:

All changes to the server administration changes, such as adding developer accounts, resolving the website issues, switching over the digital certificate to Let's Encrypt, etc. were documented in the administration log file:

One-off tasks

Task 3: Integration of scripts related to the cultural component with DIMS

This task involves the integration of calculation scripts related to the Cultural component of the report card. The Cultural component consists of two indictor groups, sense of place and indigenous cultural heritage.

The sense of place related score calculation R scripts have already been developed and integrated into the DIMS by Central Queensland University. A draft version of the Indigenous cultural heritage R scripts is also available and this will be further modified by the GHHP contracted project team in early August and provided to the DIMS team. Upon completion of this task, the DIMS should be able to accept raw field data file for the sense of place and indigenous cultural heritage and generate report card grades and scores.

Working closely with the script custodians (sense of place-Dr Jeremy De Valck, indigenous cultural heritage-TBA) this task may also require:

- *Reviewing R scripts provided for the new indicators and adjusting the scripts to make them DIMS compliant.*

- *Finalizing the data input module for indigenous cultural heritage and sense of place in order to receive the raw data sets.*

- *Conducting necessary updates to the ISP output files.*

Completed March 2019

This task involved finalising the integration of calculation scripts for the Cultural component. The Cultural component consists of two indictor groups, *Sense of place* and *Indigenous cultural heritage*.

The *Sense of place* score calculations were integrated into the DIMS several years ago. This task involved integrating in the *Indigenous cultural heritage* and the overall *Cultural* component score calculations provided by Dr Jeremy De Valck. These scripts needed to be adjusted to make them DIMS compliant. This also involved adjusting the existing *Sense of place* scripts to make them part of the overall cultural component calculations.

Besides adjusting the supplied and existing cultural component R scripts, the input module was modified to allow the upload of the raw input data sets.

Once the cultural component scripts were complete the ISP reporting script was adjusted to generate the matching tables and figures.

Task 4: Integration of mangrove indicators

This task requires integration of mangrove indicators with the Environmental component of the report card. This will be completed in two stages with only the first stage being completed in this 2018-19 project.

> *Stage 1: Making the DIMS suitable to accept score and grade calculated by the consultant. This will be done to allow the mangrove results to be included in the 2018 Report Card prior to the availability of the R scripts that process the results from raw data. Expect the scores and grades from Mangrove team by 31st Aug 2018.* **(2 days: 1 day ML, 1 day AS, 11 Sept 2018)**

> *Stage 2: Integrating the mangrove indicator scripts (developed by Norm Duke) into the environmental script DIMS to accept raw field data and generate report card grades and scores. This stage is not part of this contract and is expected to start in July – Aug 2019.* **(Not part of this contract)**

During 2018/19 project the GHHP ISP only needs the contractor to address Stage 1.

The DIMS was updated to support the upload of mangrove indicator scores into the Report Card. In this initial stage (stage 1) scores calculated by the Mangrove team are uploaded into the DIMS where they are integrated into the environmental scoring. The processing of the raw data through to the Report Card scores is prepared by the Mangrove team.

As planned, stage 1 of the integration of the mangrove indicator was completed. This involved configuring the DIMS Report Card System to accept the mangrove indicator scores as a CSV data file. The environmental script was then extended to read in, process and integrate mangrove indicator scores into the aggregation and summarization processes.

Task 5: Comparison of scores and grades using two guideline values

The ISP is considering changes to the current guidelines used in Physico-Chemical and Nutrient indicators. To support the ISP decision making process, a comparison of grades and scores for all zones for 2015, 2016, 2017 and 2018 report cards using 50th and 80th percentiles as the guideline (see a sample in Figure-1) is required. We have been using the 50th percentile for all previous report cards (2015-17) and this information is already available in the DIMS. The GHHP will request permission from PCIMP for the project team to access the previous data sets to undertake this work.

*This analysis will be performed only on the measures: **Total-N**, **Total-P**, **Chlorophyll-a** and **Turbidity**. The output will be a table showing the scores and grades for each zone and whole harbour score for both 50th and 80th percentile for each of these measures as well as aggregations up to the Water Quality indictor.*

		2015		2016		2017		2018	
		Score and grade based on (50%) percentile	Score and grade based on (80%) percentile	Score and grade based on (50%) percentile	Score and grade based on (80%) percentile	Score and grade based on (50%) percentile	Score and grade based on (80%) percentile	Score and grade based on (50%) percentile	Score and grade based on (80%) percentile
Total-N	1.The Narrows								
	2. Graham Ck								
	3. Western Basin								
	4.Boat Ck								
	5.Inner Hr								
	6.Calliope Es								
	7.Auckland In								
	8. Mid harbour								
	9.South Trees Inlet								
	10. Boyne Est								
	11.Outer Hr								
	12.Colosseum In								
	13. Rods Bay								
	Whole Harbour Score								
Total-P	Same as above								
Chlorophyll-a	Same as above								
Turbidity	Same as above								

Figure 2. Expected structure of the analysis output.

To simplify the analysis excludes applied in previous years will not be reproduced in this analysis and the 2018 environmental script code base will be used. This will result in minor changes to scores from previous years. The output of this script will inform the ISP decision whether to adopt the 80th percentile for the 2018 report card.

A comparison of alternative guideline values was conducted. See Table 2 on page 7 of the attached ISP-009-DIMS-2018-19-Task5_Percentile-comparison.pdf.

Task 8: Integration of updated Fish Indicator code

This task involves reintegrating changes to the fish indicators code into the main DIMS environmental script code base. Significant changes have been made to the fish indicator code base that require reintegration. These changes were not made to the existing DIMS compliant scripts and so significant extra work is needed to resolve the differences between the code bases and to perform multiple test and feedback cycles with the Fish indicator team.

In collaboration with the fish team, changes in the fish indicator code were incorporated into the main DIMS environmental script code base.