ISP009-2017-18: Provision of ongoing technical support for GHHP Data and Information Management System

Final Report

11th September 2018

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

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

Date: September 2018

Version: 1

Final report

The aim of the ISP-009 DIMS 2017-18 maintenance project was to ensure that the GHHP DIMS report card system was maintained and updated to allow the production of the GHHP Gladstone 2017 report card. This included a range of maintenance tasks and one-off tasks to enhance the system. The maintenance tasks were:

- To ensure that the servers kept running and were secure (with Operating System patches). This included upgrading the file sharing software of the DIMS (Pydio) to enable the latest security patches to be applied. (Task 1.1 and 1.2)

- Priority bug fixes were made to the system and that the reporting scripts were updated to bring them in-line with reporting needs of the GHHP Science team (Task 1.3). Additional time (approximately 1 week) above the original schedule was invested in this task to ensure the ISP reports generated matched changes in the reporting documents.

- 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 a number of 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 a number of off-the-shelf software components (Document Repository and the Content Management System) and a number of 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.
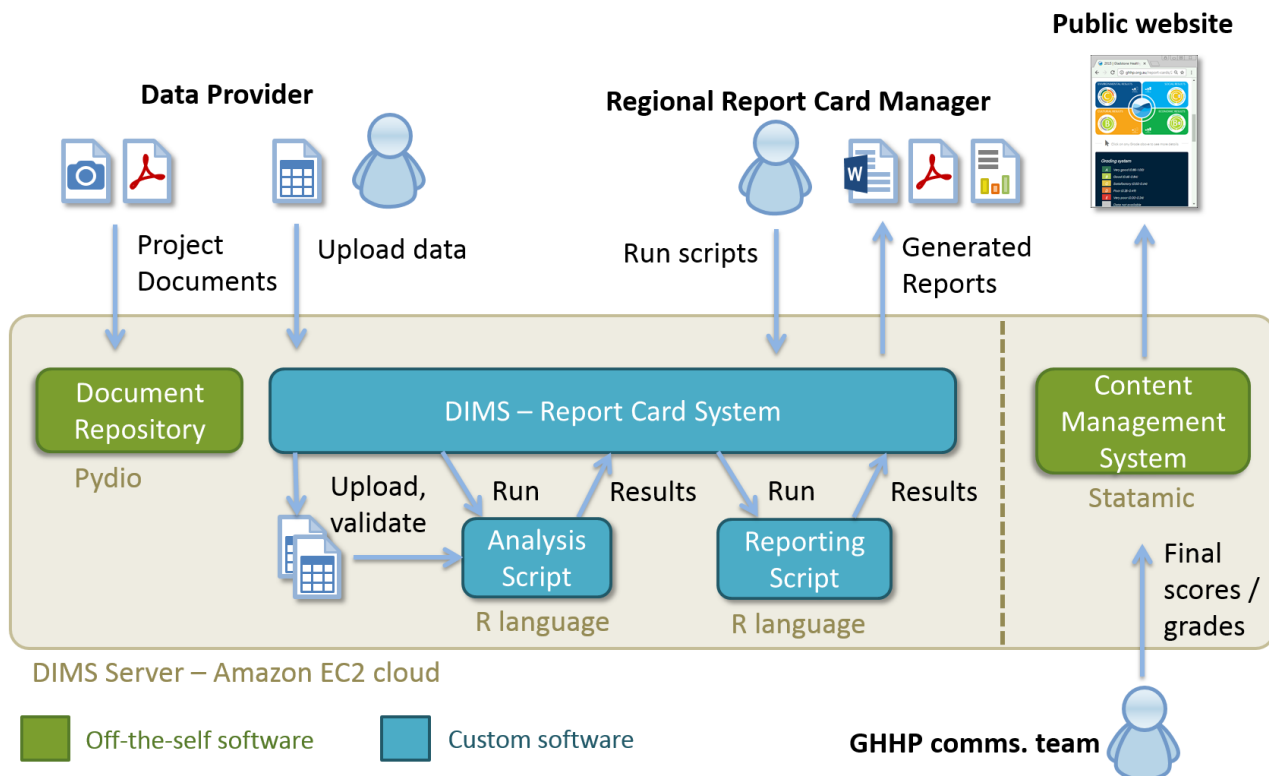

Figure 1. Structure of the GHHP Data and Information Management System.

Routine tasks

Task 1: Server Administration / Bug fixes

Task 1.1 Server Maintenance

Completed April 2018

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.

On two occasions the team provided support for outages of the GHHP website. This website is maintained by a separate communications team, and thus under their responsibility for issues that arise. The AIMS team did however resolve two issues with the website to bring the site (http://ghhp.org.au) back online as soon as possible. On the 9th Feb 2018 and 30th April 2018 the GHHP website went offline. In both cases the cause was determined to be due to a lack of memory in the PHP settings of the Statamic Content Management System. In the first case it was determined that there was a typo in the .htaccess file (PHP settings in the Statamic Content Management System hosting) that stopped the memory settings (256 MB) from working correctly. In the second case the memory was doubled in size by adjusting the .htacess file settings. In both cases no adjustments to the server configuration was necessary and the adjustments could have been made by the website maintenance team.

Task 1.2 Pydio upgrade

Completed June 2018

Pydio was upgraded from version 6 to version 8 in order to further receive security patches. This task included migrating the content first to version 7 and then to version 8 and upgrading the PHP version on the server. To avoid down time of the live server, a test server was set up and the necessary steps were performed, tested and documented before being transferred to the live server. After the upgrade the system was tested ensuring everything worked as it did before.

Upgrading to a new version implied some major changes to the user interface and workflows and so all documentation containing references to Pydio were updated.

Task 1.3 Priority bug fixes and improvements

Completed September 2018

The following list of priority bug fixes and improvements were implemented according to the additional document 'DIMS amendments to outputs v2.docx' provided by GHHP as the list of priority improvements. These included:

1. Amendments to the outputs of the ISP script to better align with the reporting requirements, based on feedback provided by the GHHP science team.
2. Adjustments to the trend plots based on feedback provided by the GHHP science team.
3. Adjustments to the GHHP Short Report based on feedback provided by the GHHP science team.

Completing all these adjustments exceeded the original project schedule by approximately a week of time, however they bring the Report Card System reporting scripts much more in-line with the current reporting needs of the GHHP science team.

Amendment description	Comment
General Comments	
It is easier for us to follow if all DIMS outputs are organized into three folders 'Website material', 'Technical report', 'Report Card' and 'GHHP Short report & CSV score tables'	After consultation it was decided to remove the folders completely and only keep the resulting Word documents. The Word documents were renamed to have a prefix indicating their intention: - report-card - technical-report - web-material
In all tables the order of the harbour zones should be from 1 to 13	Fixed
For all graphs, keep the caption closer to the top border of the image, so we can crop just the graph.	Done
The scale of all graphs-starts with 0, tick mark for 0.2, 0.4, 0.6, 0.8 and ends with 1.	For the ISP report this was already the case, but it was fixed for the trend plots.
All colours used in graphs and tables of the grades should be in line with the following colour scheme we provided in 2015.	Fixed
1. Location: connectivity folder (ISP report)	
No changes are required	
2. Location: coral folder (ISP report)	
2.1 All graphs and tables need to be updated by including *hard coral cover change* indicator.	Done
3. Location: report-card folder (ISP report)	
No changes are required	
4. Location: social-cultural-economics folder (ISP report)	
4.1 'economicvalue.plot.png' needs to be updated with **Land-based recreation** column first, **Recreational fishing** second and **Beach recreation** third	Done

4.2 'senseofplace.plot.png'> 'self esteem' should be 'self-esteem', 'values of the harbour' should be 'values of harbour'	Done

5. Location: technical folder (ISP report)

5.1 This folder has .PNG graphs similar to 'social-cultural-economic-folder'. Perhaps, this could be included in the 'social-cultural-economics' folder.	Not applicable anymore as the folders have been removed (see general comments)

6.Location: water_and_sediment folder (ISP report)

No changes are required

7. Location: whole_harbour folder (ISP report)

7.1 Remove connectivity from the 'indicator groups-whole harbour' graph	Done

8. Location: 'connectivity-2017-created-xxx.docx' file (ISP report)

No changes are required

9. Location: 'coral-2017-created-xxx.docx' file (ISP report)

9.1 Coral cover change indicator needs to be added to the graph and table.	Done
9.2 In the tables, the score column label needs right justification to align with the right column.	The row has been split into two columns so that the score column label is always above the score column.

10. Location: 'report-card-2017-created-xxx.docx' file (ISP report)

10.1 Table below the map: fix the order of the zones	Fixed

11. Location: 'seagrass-2017-created-….docx' (ISP report)

11.1 The headings of the seagrass data tables need minor amendments. The required amendments are shown in green below for inclusion. - Meadow ID - Overall zone score	Done
11.2 Table for Western Basin> Meadow column: The order should be 4, 5, 6, 7, 8 and 52-57. It is currently presented as 4, 5, 52-57, 6, 7, 8.	Fixed

12. Location: 'technical-2017-created-xxx.docx' file (ISP report)

Australian Government

AUSTRALIAN INSTITUTE OF MARINE SCIENCE

12.1 Table 3 and 4.9: the order of the zones and meadows needs to be updated.	Done (now table 4 and 4.8)
12.2 Table 4 and 4.12 need results for 'change in hard coral cover' indicator.	Done (now table 5 and 4.11)
12.3 Table x- The column headings should be 'Abundance (CPUE)', 'Prevalence of rust lesions', Size (Sex ratio), and pls correct the order of the zones.	Done (now table 7)
12.4 Table 4.13 needs coral cover change indicator added to the table. Also sites under 'Reef' column seems duplicated and this is not required.	Done (now table 4.12)
12.5 Table 4.19: The order of the columns should be 'Water and sediment quality', Habitats (seagrass and corals), 'Fish and Crabs' and the 'Overall zone score'	Done (now table 4.25)
13. Location: 'water_and_sediment-2017-created-xxx.docx' file (ISP report)	
13.1 The first two tables fit well on a landscape page layout.	After consultation it was decided not to do this point as it would imply a disproportionate amount of work.
14. Trend plots	
14.1 NH4 trends plots are presented for report card years that we did not use this as an indicator.	Fixed
14.2 Each plot can have a title, towards the top edge of the image.	Done
14.3 Some data points on trend plots have a CI while this is not consistently appearing on every graph.	Fixed
14.4 'all_harbour_environmental.png' trends plot does not have 2015-16 data point.	Fixed
15. GHHP_ShortReport	
15.1 Revision History Table on P2: This does not seem to be updating correctly.	This part has been removed completely because it is not necessary as this report is for reviewing data and QAQC.
15.2 Table 2: The scripts seem to be not working properly when indicators are removed from the model. This table is very important to us as it shows what measures are included in the analysis.	Fixed

Australian Government

AUSTRALIAN INSTITUTE OF MARINE SCIENCE

15.3 Section '2.5 Sediment Quality PAH' can be removed from the report.	Fixed
15.4 The following key in the report can be improved. E.g. ● Data available ● Data not available	Done
15.5 Figure 8: The x axis of some plots are hard to read.	Fixed
15.6 The fish recruitment scores: Between this report (Table 30) and the once produced by Bill V (below), there is a 0.01 difference for some reporting zones and we believe this is due to the bootstrapping aggregation. Perhaps, Murray can look into this for next year and add a footnote to Table 30 explaining this. E.g. compare scores in Table 30 with Scores from Bill V for Boat Ck (zone 4), Inner Harbour (zone 5), Boyne estuary (zone 10a) and Colosseum Inlet (zone 12)	Fixed
16. Additional (ISP report)	
16. Report a count of the total measures used in the Report Card (dealing with exclusions).	A table with the number of measures per indicator has been added been added to the technical report document.

In addition the technical report and the report card had a general overhaul of the ISP reporting script was done to more align the generated outputs with the published reports. For the technical report the overhaul included

- adding headlines

- adding missing tables and plots

- adjusting table and figure numbers

- adjusting table and figure captions

- rearranging content

For the report card the adjustments included

- rearranging the scores tables (see screenshot below)

- changing the look of the rosettes (see screenshot below)

- removing connectivity from environmental results

- adjusting map and environmental table

Environmental Results

ENVIRONMENTAL	SCORE	GRADE
OVERALL ENVIRONMENTAL	0.63	C
Water and Sediment Quality	0.86	A
- Water Quality	0.77	B
- Sediment Quality	0.94	A
Habitats (Seagrass and Corals)	0.47	D
- Seagrass	0.35	D
- Coral	0.60	C
Fish and Crabs	0.60	C
- Bream	0.66	B
- Mudcrab	0.55	C

Besides working on the amendments, part of this task was to improve the robustness of the ISP report.

This involved adjusting the ISP-Report to tolerate missing labels associated with indicator IDs. This problem was noticed when the new fish indicators code was added. When the new indicators were added, but there matching labels were not provided (in the labels.csv file) the ISP script would fail with a confusing error message. This task involved making the ISP script detect this condition then provide a good error message detailing the problem so that it could be easily fixed. The updated ISP script reports that the labels are missing, but still proceeds to generate the reports instead of failing.

Furthermore the "overwrites" mechanism needed some modifications to improve the usability. This included visual enhancements and the addition of buttons and mouse over effects to make the workflow easier.
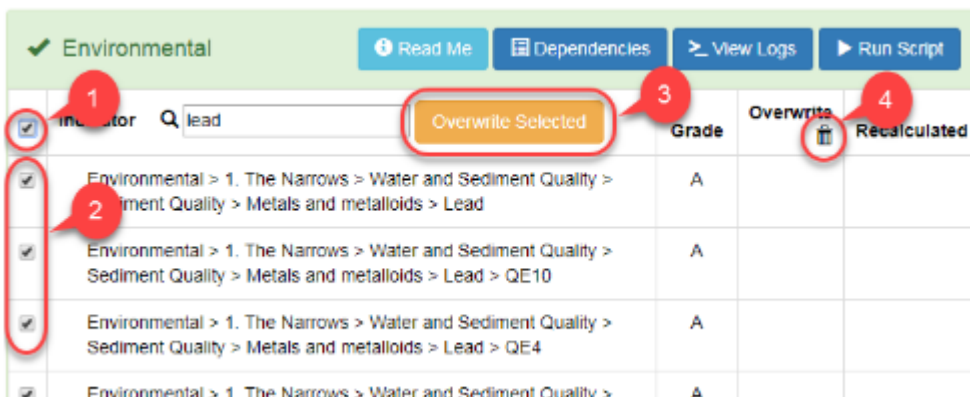
The documentation has been revised and updated to include the changes and to have a detailed description of how the manual overwrites work and how to use them.

https://dims.ghhp.org.au/repo/ws-documentation-1/DIMS-UserGuide-2018-07-20.pdf

Task 2: Getting data into the Report Card System / General support

Completed May 2018

Provided access to the environmental scripts to Stefan Sawynok to assist him understanding the existing Bream fish indicator.

Adjusted the environmental seagrass script to deal with the case when biomass values were null. This was done to allow the system deal with 2018 seagrass data. This involved diagnosing the issues of why the script fails, working with the seagrass team to establish how the scripts should be adjusted to deal with this case, then testing and pushing the changes into production.

Task 3: Project management and reporting

The project management and reporting task included an inception meeting and regular telephone meetings to communicate the status of the project and get feedback on arising questions. In addition the tasks needed to be scheduled and coordinated internally and the final report written. During the project a contract variation was made to remove Task 6.2 (see below for more detail).

One-off tasks

Task 4: Statistical advice and R script maintenance

Task 4.1 Fixing up the Environmental code GIT repository

Completed July 2017

All the scripts and code associated with the DIMS Report Card System are hosted as private GIT repositories on GitHub (https://github.com/ghhp-dims). Having the code in a source control repository (GIT) hosted on GitHub allows a detailed record of all code modifications and allows multiple developers to work on the same code base. It also ensures that a complete master copy of all the code for the system is available for future maintenance of the system.

At the end of the original DIMS project the environmental scripts were setup in a GIT repository, however this repository contained a copy of the PCIMP dataset in its history. To comply with the PCIMP data sharing agreements which required that "All PCIMP data be deleted after use" the entire GIT repository needed to be deleted and a new GIT repository be established. This new GIT repository was then setup to have exclusion rules that prevent the dataset from ever being saved in the repository.

This new version of the repository needed to be established before any new modifications to the code base could be applied.

Task 4.2 Improved script developer workflow (GitHub Integration)

Completed August 2017

This task involved improving the workflow associated with synchronising the script code running on the DIMS with the code stored in the GIT repositories hosted on GitHub. This streamlines the migration of code in development to the production report card, reducing errors in managing updates to the code.

Previously code was developed on the script developers own machine (it still can be). They would then commit this code back to the code repository in GitHub. The DIMS team would then pull this code down into a local copy then manually copy the code to the server using Pydio. We would then test the new code. Any changes meant modifying the copy on our local machine then recopying it to the server. Once the testing and adjustments were complete we would repeat this process to put the code in production. The difficultly with the production report card was that we need to preserve any existing data uploads. This meant carefully copying over the code and data settings without affecting any of the existing data. This was a fiddly, time consuming and potentially error prone.

To improve the workflow a new process was setup where the script code on the server was directly linked with the GitHub repositories allowing them to easily pull down new code modifications. The script developers (with sufficient skills) were also provided with server access via SSH. This allowed them to use Git directly on the server, make adjustments to the code, test them on the server then push the adjustments directly back to the repository. Testing directly on the server removed many potential incompatibilities between the local test environment and the production environment.

Adjustments to the code base were made to allow the code to be run and edited directly on the server. This principally involved adjusting the .gitignore file to ensure that no files generated by the scripts or any uploaded data is pushed back into the Git repository and the Git repository must only contain the script code, not generated scores or data.

Task 4.3 Support for selectively removing individual PCIMP data points

Completed August 2018

While the Report Card System allows individual indicators and measures to be removed from the Report Card calculations using its User Interface it does not allow adjustments to individual time samples. Unfortunately in the last several report cards individual samples of the PCIMP dataset have been requested to be removed by the ISP due to data quality issues. Due to the restrictions in the Report Card System User Interface the only mechanism for achieving these exclusions was to delete the data from the dataset prior to uploading it into the system. This made it difficult for the GHHP science team to make these adjustments in a timely manner.

This task was to develop a simple way to allow individual PCIMP data points to be removed from the Report Card System calculations. To allow this level of control to be performed using the Report Card System User Interface would have required significant development time resulting in project that exceeded the budget and so a simple approach was taken where a CSV file is used to specify data points for exclusion.

The CSV is generated by the system on the first run of the environmental script and included in the output. After downloading the output, the user extracts the file from the ZIP folder and marks all data points which should be excluded with an 'x'. The file is then uploaded as a dataset file (PCIMP exclusions). During the next run of the environmental script the uploaded file is processed and the marked data points are ignored.

Task 5: Integration of Fish recruitment indicators

Completed September 2017

An update to the fish indicators was provided to the DIMS team by Bill Venebles. This update included a modified version of the scripts to incorporate changes to the modelling component of the script, and the addition of calculating grades and scores at the site level, rather than just at the zone level. Murray Logan

DIMS Report Card System   2017-18 Final Report

integrated these changes into the existing environmental script and converted the supplied spreadsheet data into the required CSV format. Eric Lawrey then performed additional integration testing, enhancements to the validation rules for the input dataset and adjustments to the system to cater for the new site level indicators. The example CSV was also adapted to the format changes. The updated code and the CSV formatted data was sent back to Bill as a reference for how the data must be supplied in the future.

Task 6: Integration of mud crabs indicators

Completed October 2017

A new set of indictors for mud crabs was developed for the 2017 report card by project ISP015-2017. The results of the project were however only available very close to the finalisation date of the 2017 report card and so the intention was to integrate this new indicator in two steps:

1. The mud crab indicator scores were added directly into the system as a dataset, to be processed by the environmental script. This allowed the scores to be aggregated into the environmental scores, without having to integrate the scripts that process the results from the raw data. This allowed a very quick turn around on the initial integration. This involved creating a MANIFEST file for the new dataset, adjusting the environmental script to pick up the new dataset to include it in the aggregation process, adding support for it in the ShortReport.pdf generated by the environmental script and adding the mud crab scores to the generated technical report in the ISP script. The changes were then tested and migrated to production and the scripts rerun.

2. This task was to integrate the mud crab indicator scripts into the environmental scripts. This task was however removed from the project contract as the script code, required to complete this task, could not be provided to AIMS during the runtime of this project. For more details see the Letter of Variation dated the 9th May 2018.