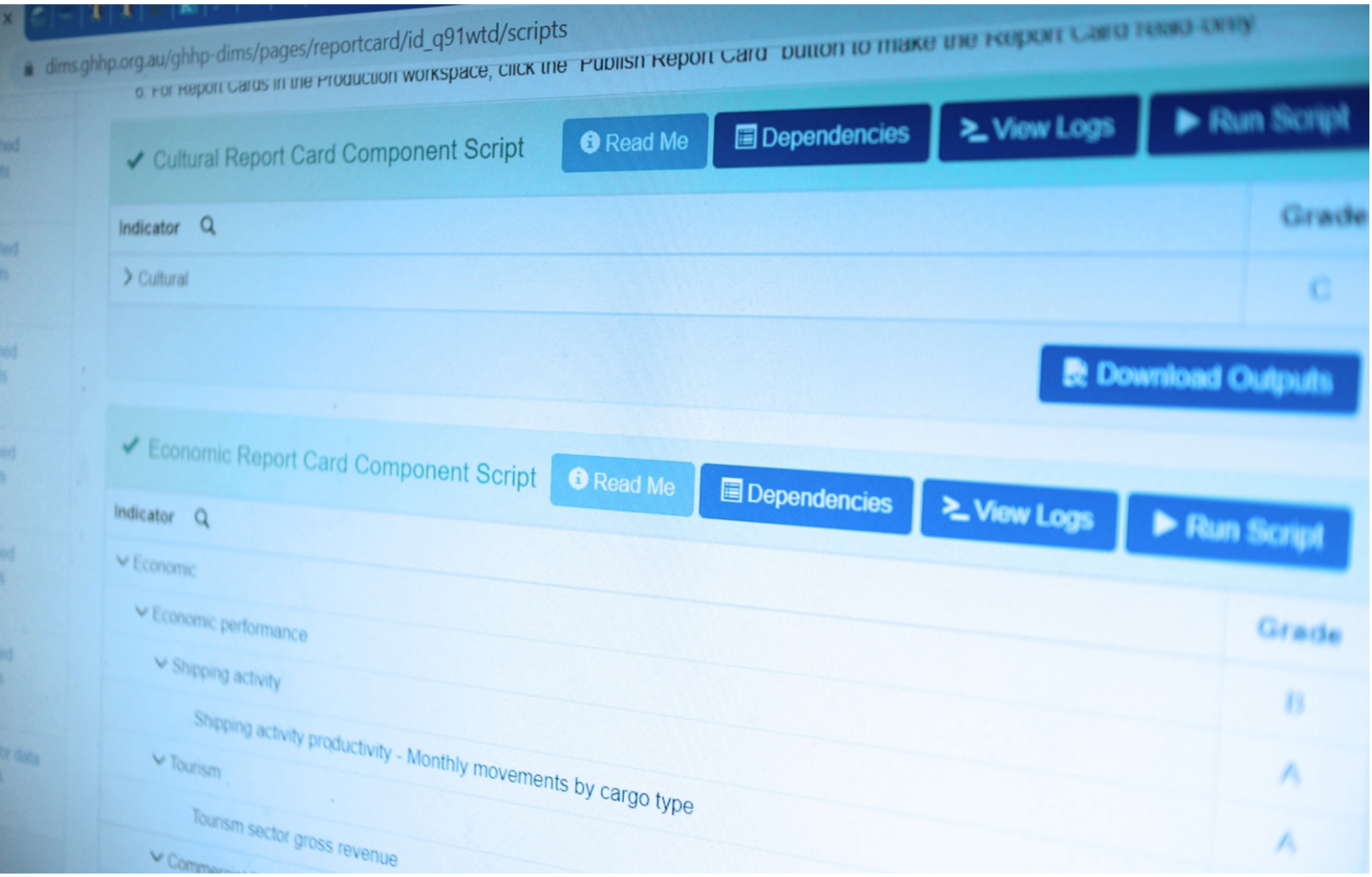
Australian Government

Australian Institute of Marine Science

50 YEARS 1972 - 2022 Science for tomorrow's oceans

FINAL REPORT ISP009: 2023–24: PROVISION OF ONGOING TECHNICAL SUPPORT FOR GHHP DATA AND INFORMATION MANAGEMENT SYSTEM (DIMS)

Marc Hammerton, Gael Lafond, Eric Lawrey



Final project report prepared for Gladstone Healthy Harbour Partnership.

Australian Institute of Marine Science

PMB No 3
Townsville MC Qld 4810

PO Box 41775
Casuarina NT 0811

Indian Ocean Marine Research Centre
University of Western Australia, M096
Crawley WA 6009

This report should be cited as:
M Hammerton, G Lafond, E Lawrey (2024) Final Report ISP009: 2023–24: Provision of Ongoing Technical Support for GHHP Data and Information Management System (DIMS). Report prepared for Gladstone Healthy Harbour Partnership. Australian Institute of Marine Science, Townsville.

© Copyright: Australian Institute of Marine Science (AIMS) 2023
All rights are reserved, and no part of this document may be reproduced, stored, or copied in any form or by any means whatsoever except with the prior written permission of AIMS

Where the disclaimer and/or copyright are joint, the other parties need to be added accordingly and the tense changed.

DISCLAIMER
While reasonable efforts have been made to ensure that the contents of this document are factually correct, AIMS does not make any representation or give any warranty regarding the accuracy, completeness, currency or suitability for any particular purpose of the information or statements contained in this document. To the extent permitted by law AIMS shall not be liable for any loss, damage, cost or expense that may be occasioned directly or indirectly through the use of or reliance on the contents of this document.

Project Leader shall ensure that documents have been fully checked and approved prior to submittal to client.				
REVISION HISTORY:		Name	Date	Comments
1	Prepared by:	Marc Hammerton	10/06/2024	
	Reviewed by:	Eric Lawrey	11/6/2024	
	Approved by:	Aleksandar Zivaljevic	11/06/2024	

Cover photo:
Screenshot of the GHHP DIMS report card system web interface. Image: E. Lawrey

Contents

1 SUMMARY

The GHHP Data and Information Management System (DIMS) is a web-based system for managing the data and scripts used to generate the Gladstone Harbour Report Card (https://www.ghhp.org.au/). The Australian Institute of Marine Science manages the hosting and maintenance of this system, upgrading software components as needed and making enhancements or adjustments to the report card scripts as requested by GHHP. This project covered one-year of the on-going maintenance of this system and decommissioning of a development server.

2 PROJECT BACKGROUND

The Gladstone Healthy Harbour Partnership (GHHP) Data and Information Management System (DIMS) is a web-based system that helps coordinate the submission and processing of monitoring data associated with the annual GHHP environmental report card for the Gladstone region. This system allows groups performing the monitoring to upload their data into the system. The system performs a range of checks on the monitoring data to ensure consistency and compatibility with the logic of the report card scripts. The algorithms that implement the logic of the report card are implemented as a range of R scripts. The DIMS system allows the GHHP science team to run these scripts on the monitoring data to produce a range of products (trend plots, preview of the report card scores and prepopulated report card templates) that are used to produce the final report card.

The DIMS system was actively developed from 2014 – 2016, with further refinements and additional report card indicators added to the R scripts from 2017 – 2020. This project covers the tasks associated with the maintenance, support, adjustments, and necessary software upgrades of the GHHP Data and Information Management System (DIMS) over the 2023-2024 financial year.

The source code for the DIMS system software and the scripts are stored in private repositories on GitHub at the address: https://github.com/ghhp-dims. The server associated with the DIMS is hosted on the Amazon cloud and managed by the Knowledge System team at AIMS.

3 PROJECT DELIVERABLES

3.1 Task 1: Project inception meeting

Completed September 2023

This meeting was conducted on 19 September 2023 with John Rolfe, Kirsten McMahon, and Marc Hammerton.

3.2 Task 2: Server Administration

Completed June 2024

This task maintains the on-going robust operation of the DIMS. It includes keeping the services online, assisting and investigating issues with uploading datasets, but also developing bug fixes for issues identified in the report card scripts or the DIMS software. The following is a summary of these activities for the year:

3.2.1 Routine Operating System Patches

Completed June 2024

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 during the year by starting one of the backup images to ensure that they would run.

3.2.2 Server storage increase

Completed September 2023

The GHHP team attempted to upload a large dataset to NextCloud but the disk was close to full capacity and could not handle the new dataset. We added an extra 100 GB storage capacity to the NextCloud disk to resolve the issue and provide more space for future datasets.

3.2.3 Restructure of user and roles

Completed September 2023

The DIMS user and roles required a restructure in order to reflect the new organisation structure as several members left the organisation, new members joined, and some groups were split. Several emails were exchanged between Kirsten McMahon (GHHP) and Gael Lafond (AIMS) to find out what changes needed to be made to the DIMS user structure. After this the changes were applied to the system and verified that they reflected the requirements.

3.2.4 Investigation of server inaccessibility issue

Completed October 2023

On the morning of the 9th of October 2023, the server became inaccessible and had to be manually rebooted. An investigation of the server logs did not show any noticeable cause for the issue. It was believed to have been caused by some unusual AWS hardware failure. Some work had to be done to restore all the services, such as re-attaching the SSL certificates to the website and re-creating the NextCloud docker container.

3.2.5 Encrypt historic PCIMP data on the server

Completed October 2023

The report card scripts require sensitive data provided by the Port Curtis Integrated Monitoring Program (PCIMP). This data is not public and needs to be handled with care. The server is secured and only a small team of users have access to the data. But still, all those protections cannot fully protect the data from a server breach.

To mitigate this issue, we decided to encrypt the PCIMP data for previous report card years where the report card scripts have been executed and the scores generated. We should not need the data again unless we are required to re-generate the scores. In this case, we can decrypt the data using the secure password (stored in our local password database) and make it again available to the scripts.

We choose 7zip to encrypt the data as it uses the very secure AES256 encryption algorithm. The encrypted file is portable, making it easy to recover from a file backup. The encryption and decryption process is documented in the private GitHub repository https://github.com/ghhp-dims/pcimp-encryption . We will apply the encryption manually each year for new PCIMP data after the publication of the report card.

3.2.6 Upgrade R libraries

Completed October 2023

Kirsten McMahon notified the Knowledge System team of some errors appearing while running scripts on the latest production report card. The errors were caused by old R libraries running in a newer R environment.

We updated the libraries in a staging environment to make sure the new versions were producing the same output. After confirming the scores, we updated the libraries on the live server, and tested the scripts using a cloned report card.

3.2.7 Migrate social-economic scripts to Docker

Completed October 2023

The three social-economic scripts (cultural, economic, and social) were using the version of R installed on the host. After a recent update to the host which included an update to R, the scripts stopped working. To solve the problem and ensure it does not occur again, we migrated the three scripts to a Docker environment with a fixed version of R and its libraries.

3.2.8 Migrate ISP report and Trend-Plots to Docker

Completed November 2023

Similar to the previous section (3.2.7), the ISP report script was using the version of R installed on the host and after a recent update, the script stopped working. To solve the problem and ensure it does not occur again, we migrated the script to a Docker environment with a fixed version of R and its libraries. The Trend-Plots script did not seem to have any issues running with the latest version of R but to prevent future problems, we also migrated the script to a Docker environment.

3.2.9 Services upgrade

Completed December 2023

Several services were outdated and required upgrades to ensure the server stays secure.

- **JkMount**: Some Apache configuration files were still using the deprecated JkMount module. We replaced them with the Apache Proxy module.
- **GeoNetwork**: The GeoNetwork associated with the e-Portal was running version 2.10.3. The latest stable version is 3.12.11. Upgrading from version 2 to version 3 is not straightforward and would require more time than we had allocated for this upgrade. To protect the server, we isolated the GeoNetwork webapp in a Docker container.
- **DIMS**: The DIMS has been compiled with Java libraries which are now at risk. We upgraded all the libraries and deployed the new version of to the server.

- **KeyCloak**: KeyCloak is the system that manages user credentials. It is very important to protect this information. We needed to upgrade to the latest version as it had several new security features and patches.

3.2.10 NextCloud upgrade

Completed January 2024

While most service upgrades are straight forward, upgrading NextCloud is a bit more challenging. In January 2024 we finished upgrading the system to version 28.0.5.

3.2.11 KeyCloak certificate renewal

Completed January 2024

KeyCloak's internal SSL certificates expired in January 2024 and we created a new SSL certificate with a long validity period to avoid this issue happening again in the near future.

3.2.12 Library 'Glibc' vulnerability

Completed May 2024

A new vulnerability was discovered in 'Glibc' which is a library that can be used by PHP. In order to protect the server, we:

1. Converted the GeoNetwork associated with the e-Portal into a site static.

2. Removed PHP from the server.

3. Removed the old Statamic report card website.

4. Moved NextCloud to a PHP FPM container based on Alpine Linux (Alpine Linux uses 'musl' instead of Glibc).

3.2.13 GeoNetwork static website

Completed May 2024

The server was hosting an old version of GeoNetwork (see 3.2.9) as part of the e-Portal data and literature review developed in 2013. The old version of GeoNetwork exposes several vulnerabilities, and so we decided to replace GeoNetwork with a static website which serves the metadata records as PDF, XML and ZIP archive. This was achieved by converting each of the metadata records in the system into PDFs, then creating a navigation page listing all the records in the site. To not break the links, we kept the URL at https://dims.ghhp.org.au/geonetwork/ .

3.2.14 Rebuild DIMS Docker containers used to run R scripts

Completed May 2024

The DIMS R scripts run in Docker to be resistant against side-effects of un-attended updates. Work on the server in January 2024 removed what was thought to be unused Docker images, however this accidentally removed the Docker images needed to run the R scripts.

While rebuilding most of the Docker images was straight forward, rebuilding the Docker image for the environmental script was not possible as it was based on versions of R libraries which were no longer accessible.

We recovered the Docker image from a backup and uploaded it to AWS ECR, an internal AWS Docker image repository. We also uploaded all the other R script docker images to avoid a replication of this issue in the future. With this in place we can always restore the Docker image directly on the server without needing to restore a backup.

3.3 Task 3: Decommission of development server

Completed October 2023

In previous years the Bitplex team was using a development server to test their modifications to the GHHP website before applying them to the live server. This development server is no longer required since the Bitplex team no longer maintain the website, and the GHHP website is no longer hosted on the same server as the DIMS. The development server was terminated, and its resources deleted.

3.4 Task 4: Updating the training manual and final report

Completed October 2023

We created a guide for publishing report cards and added it to the existing user guides on NextCloud.

4 CONCLUSIONS

The goal of this year's project was to maintain the server and its software systems and to decommission the development server as it was not needed anymore. Both goals were achieved successfully without any major issues. For the future it is highly recommended to consider updating the R scripts and their libraries to a newer version as we noticed it becomes more and more difficult to maintain these scripts.