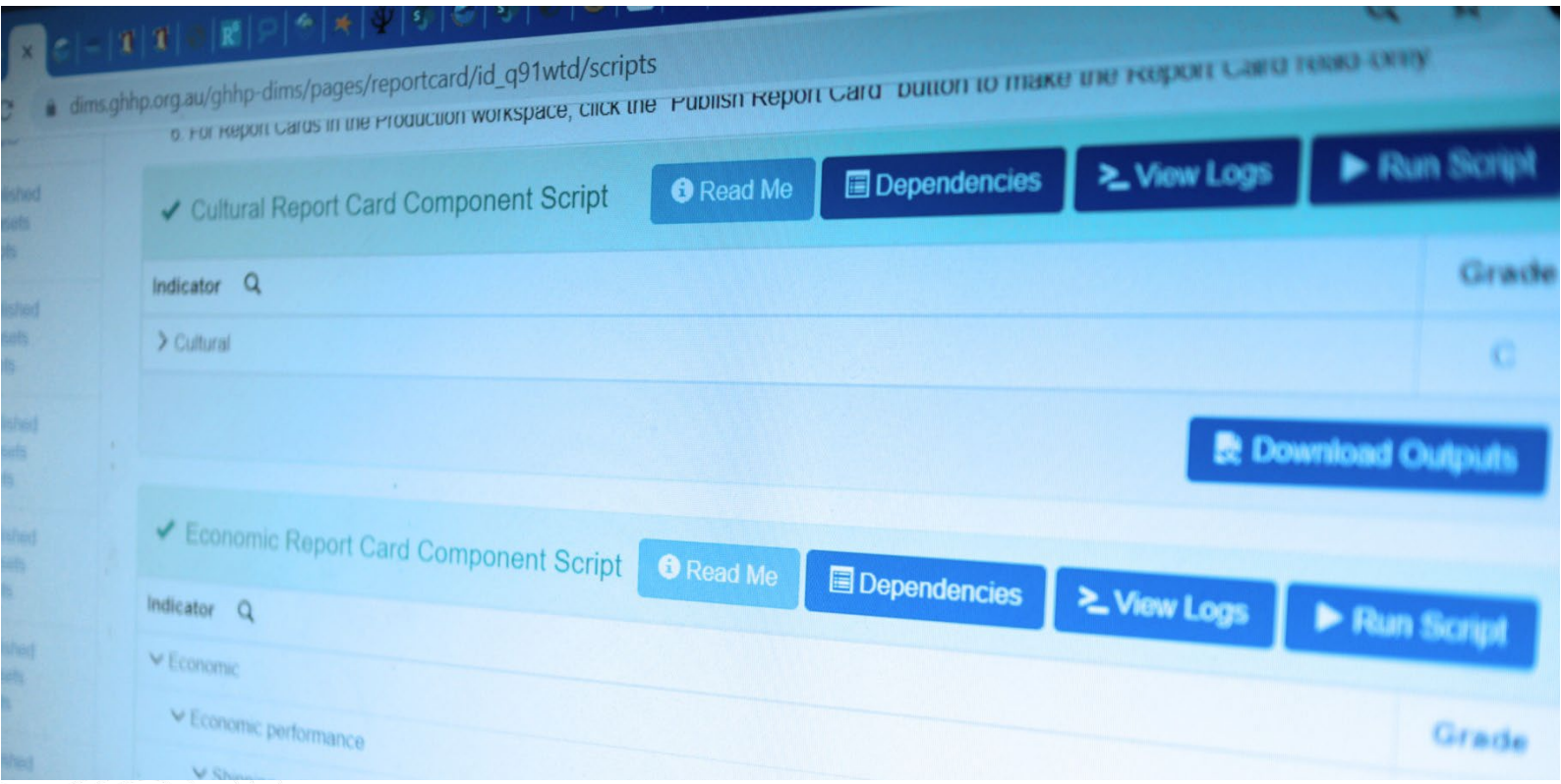
Australian Government

AUSTRALIAN INSTITUTE OF MARINE SCIENCE

ISP009-2020-2021: Provision of ongoing technical support for GHHP Data and Information Management System

Eric Lawrey, Murray Logan, Gael Lafond



Final project report prepared for Gladstone Healthy Harbour Partnership.

AIMS: Australia's tropical marine research agency.
www.aims.gov.au

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:
Eric Lawrey, Murray Logan, Gael Lafond (2022) ISP009-2020-2021: Provision of ongoing technical support for GHHP Data and Information Management System. Report prepared for Gladstone Heathy Harbour Partnership. Australian Institute of Marine Science, Location. (8 pp)

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:	Eric Lawrey	*22/04/2022*	
	Reviewed by:	Lyndon Llewellyn	*27/4/2022*	
	Approved by:	*Insert Program Director's name*	*Insert date*	

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

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 (http://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 focused on major upgrades to the user management and file sharing component (*Pydio*) and the Operating System (OS) of the hosting server. These components had reached their product end-of-life, and so were no longer receiving security patches. Upgrading these components was necessary to ensure the cyber security of the service.

These upgrades were more complex and time consuming than planned. We had planned on upgrading *Pydio* (file management and sharing software) to *Pydio cells* as this was the recommended path from the original supplier of *Pydio*. Unfortunately, after testing we found it to be unsuitable as a replacement due to had poor performance and reliability. As a result we needed to source a replacement for *Pydio*. After a review process the team determined that *Pydio* could be replaced with two open-source software products: *NextCloud* (file sharing) and *KeyCloak* (user management). An advantage of these new packages is that they have a much larger community adoption and development team than *Pydio* and so should hopefully have good long-term support. These were configured and integrated with the DIMS and the existing file shares and users were migrated the new system. In the original DIMS, both the user management and file sharing features were performed by *Pydio*. In the new design user management is implemented using *KeyCloak*, while the file sharing is performed using *NextCloud*. This was done to increase the modularity of the design. These new software packages should serve the DIMS well for many years.

As part of the maintenance of the DIMS server regular weekly security patches are applied to its Operating System. These security patches apply small changes to the Operating System kernel. Every 4 years a major version upgrade of the Operating System is required to continue to receive these security patches. During this year the server running the DIMS and the GHHP website needed a major version upgrade from Ubuntu 16.04 LTS to Ubuntu 20.04 LTS.

Unfortunately, this Operating System upgrade triggered numerous additional upgrades in software packages on the server that resulted in slight behaviour changes in various tools and libraries that led to complications that took significant time to resolve. This resulted in issues with the website and the report card scripts that needed to be resolved. As a result the Operating System upgrade took significantly more time than anticipated. As part of this fix the report card scripts were changed so that they now run within a fixed containerised environment (Docker). This ensures that the scripts run on the same version of R that they were originally written and tested for, rather than the latest one that gets shipped with the Operating System. This approach ensures that the scripts have a consistent behaviour even across major upgrades of the Operating System.

In addition to these system upgrades an adjustment was made to the scores aggregation hierarchy for mud crabs.

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 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 2020-2021 financial year. This project covers routine maintenance tasks and one-off specified tasks to ensure the continued operation of the system.

The source code for the DIMS system software and the scripts are stored on 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. This server also hosts the GHHP website (http://ghhp.org.au) along with the legacy GHHP e-Portal (http://data.ghhp.org.au/) that hosts a catalogue of 350 metadata records. Additionally, a development server containing a mirror of the GHHP website is maintained to allow the website developer (Bitplex) to develop changes to the website safely prior to putting them on the production website.

3 PROJECT DELIVERABLES

This section outlines each of the project scheduled tasks and the results of what happened during the year.

3.1 Task 1: Project inception meeting

This meeting was conducted on 1 Oct 2020.

3.2 Task 2. Upgrade Pydio 8 to Pydio Cells.

- Install Pydio Cells and migrate existing configuration and files.
- Update connector used by DIMS web application for username/password login.

The server that hosts the GHHP DIMS also hosts the GHHP website and a legacy metadata website (e-Portal). *Pydio* is a user management and file sharing web application that is used by both the DIMS and for sharing documents on the website. It has been in use since the initial development of the DIMS in 2015. Unfortunately, the *Pydio* is no longer supported by the original developer and so reached its end of life as it no longer received security patches. The recommend replacement for *Pydio* was *Pydio cells*, which was a rewrite of *Pydio* using a more modern programming language (Go).

During this project we reviewed *Pydio* cells to determine its robustness a suitability as a replacement for *Pydio*. Unfortunately, it was found to have numerous performance and stability issues that meant it was not suitable as a replacement. This meant that we needed to find a similar product to act as the replacement and to integrate it into the system.

After reviewing a suite of filesharing applications we settled on using *NextCloud* (https://nextcloud.com/) for file sharing and *KeyCloak* (https://www.keycloak.org/) for user management.

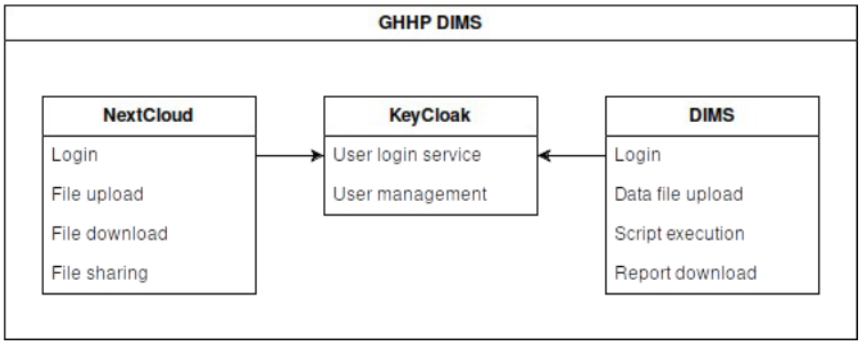


Figure 1. Overview of the roles of NextCloud, KeyCloak and the DIMS.

Both NextCloud and the DIMS use the KeyCloak services for authentication of users. Replacing Pydio required the following tasks:

- Configuring NextCloud to connect with KeyCloak.

- Configuring KeyCloak to act as a user management system
- Configure branding of NextCloud and KeyCloak to have Gladstone imagery to help users feel comfortable that they are in the right location.
- Developing a new module for NextCloud to allow public links to have meaningful URLs, not just random IDs.
- Developing scripts for migrating existing documents shares.
- Configuring email services on the server to support email password recovering via KeyCloak.
- Setting up all existing users in the system.
- Adjusting the DIMS software to integrate with the KeyCloak authentication services.

The upgrade process took significantly more time than was originally budgeted in the project, highlighting the challenge of upgrading major components of the site as they age.

3.3 Task 3: Server Administration

- Ensure the DIMS server remains online, is backed up and has security patches applied to its software components.
- Priority bug fixes
- Minor upgrades to the existing report card scripts.

This task is a bit of a catch all bucket for activities that maintain 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:

3.3.1 Adjustment to the mud crabs aggregation hierarchy

At the request of GHHP, the aggregation hierarchy for mud crabs was altered such that the Indicator Score for Whole of Harbour was calculated as the mean (both arithmetic and bootstrapped) of Whole Harbour Subindicator Scores which were themselves calculated as the means of the Zone Subindicator Scores. This scheme represents a deviation from the rest of the report card Indicators that calculate Whole Harbour Indicator Scores by averaging the Zone Indicator Scores.
Diagrammatically, this might be summarised as:
Regular aggregation hierachy: Zone/Indicator -> WH/Indicator
Mudcrab aggregation hierarchy: Zone/Subindicator -> WH/Subindicator -> WH/Indicator

From a technical perspective, this involved creating an additional aggregation pathway (WH/Subindicator -> WH/Indicator) just for mudcrabs and substituting this result into the general WH/Indicator database prior to its downstream application in the full aggregation hierarchy.

Murray Logan made the required code changes which were then tested and pushed to the production report card.

3.3.2 Bug - Group write permissions not set correctly on cloned report cards.

After an upgrade of the Tomcat server (it runs the DIMS software) it was noticed that cloned report cards could not be modified by the owner of the cloned report card when accessed directly on the

server (via SSH). This caused development issues for Murray Logan when trying to adjust and test the environmental R scripts.

After investigation it was determined that mechanism for ensuring the group write permissions were set properly had been changed in Tomcat. The Tomcat setup was adjusted to the new approach and the group permissions of existing cloned report cards was manually fixed.

3.3.3 Upgrade of the GHHP Server Operating System

Normal maintenance of the server operating system involves applying weekly security patches and performing Kernel upgrades. These regular patches help ensure that the server is secure. These security patches don't however perform major operating system upgrades. The GHHP Server runs on the Long-Term Support (LTS) version of Ubuntu Linux. Each LTS version of Linux is supported for 5 years. Prior to the upgrade the GHHP server was running the version 16.04, resulting in an end of support in April 2021.
Performing a major version upgrade of the operating system also upgrades all the tools and libraries that make up part of the operating system. This resulted in upgrades to PHP that Statamic runs on those powers the GHHP website. This also upgraded Java that runs the legacy GeoNetwork in the e-Portal website on the server. It also upgraded the version of the R language and its libraries used to run the report card scripts. This upgrade to R was found to slightly change the behaviour of the scripts, leading to a difference in the generated scores of the environmental scripts. This change was due to an error that was triggered somewhere in the script, which the script treated as missing data.

To resolve this problem and ensure that the environmental report card script generated the intended results it was setup to run in a Docker container. This process involves creating an execution environment that has a specified version of R that is known to work with the script.

3.4 Task 4: Project management, training manual update and final report

The file sharing and user management systems were changed significantly with the introduction of *NextCloud* and *KeyCloak*. The following documents were developed or updated to align with the new changes. Note: The links provided require a login to DIMS.

- DIMS Roles Admin – Creating and Maintaining Roles (https://nextcloud.dims.ghhp.org.au/f/162036)
- KeyCloak Admin – Creating and Maintaining Roles (https://nextcloud.dims.ghhp.org.au/f/162078)
- Keycloak NextCloud Setup (https://nextcloud.dims.ghhp.org.au/f/162515)
- DIMS Introduction (https://nextcloud.dims.ghhp.org.au/f/164348)

4 CONCLUSIONS

This year's project was dominated by the work required to upgrade the file sharing and user management systems of the DIMS and the hosting server's Operating System. While they represent a large amount of work, they only need to be undertaken approximately every 5 years and they help ensure that the DIMS will remain modern and secure.