



Solutions Architecture: Baseline version, 2020

Project Details

Contracting organisation: Queensland Cyber Infrastructure Foundation (QCIF)

Primary investor: the NCRIS-funded Australian Research Data Commons (ARDC)

Other organisations involved in project: Atlas of Living Australia (ALA), CSIRO, Centre of Excellence for Biosecurity Risk Analysis (CEBRA), Griffith University, Macquarie University, TERN, University of NSW

Technical Lead: Arve Sølland

Technical Lead phone & email: 0450 742 677 | a.solland@griffith.edu.au

Program manager name: Elisa Bayraktarov

Program manager phone & email: 0401 676 881 | e.bayraktarov@griffith.edu.au

Revision History

Version	Date	Editor	Summary of changes
0.1	29/06/2020	Arve Sølland	Initial Document
0.2	03/07/2020	Arve Sølland / Elisa Bayraktarov	Revised based on feedback from team.
0.3	21/7/2020	Arve Sølland	Revised based on feedback from team and GU stakeholders.
0.4	23/7/2020	Arve Sølland	Revised based on feedback.
0.5	30/7/2020	Arve Sølland	Revised based on feedback
0.6	30/8/2020	Arve Sølland	<ul style="list-style-type: none">Added more details to how existing systems will be transferred into EcoCommons.Added links to CoESRA from EcoCloud/Analysis Sandbox.Added integration/migration plan overview diagramRevised Security parametersAdded mention of tracking stats and analyticsRemoved reference to stakeholder list, and rather link to Implementation document for this.Added information around development process and pipeline
1.0	26/10/2020	Arve Sølland	<ul style="list-style-type: none">Integrated feedback received during two national consultations with technical representatives from the Australian Research Data Commons and the Terrestrial Ecosystem Research Network as well as two international consultations with Lancaster University 'Data Labs' and Microsoft - Higher Education IndustryIntegrated feedback from consultations with members of the EcoCommons interim Technical Reference Group

			<ul style="list-style-type: none">• Fixed typos• Removed duplicate diagram• Removed old overview diagram• Added more detail on integration and user acceptance testing• Added more detail on CSDM and its role/integration
--	--	--	--

Supporting Documents

- EcoCommons Implementation Plan
- EcoCommons Milestones Gantt Chart

Requirements for Individual platform components

- Microservice Toolbox Requirements
- Workflow and result management API service Requirements
- Data Explorer Requirements
- Analysis Sandbox Requirements

1. INTRODUCTION

1.1 Background

Now more than ever researchers, industry and governments are being challenged by a growing, interconnected human population, a changing climate, and an increasing demand on natural resources such as land, energy and water. Recent technologies have enabled consistent and continuous collection of ecological data at high resolutions across large spatial extents. The challenge remains, however, to bring this data together and expose it to methods and tools that generate meaningful information about the environment, and lead to solutions to enviro-socio-economic problems.

EcoCommons has been proposed as a way to address these information and analyses gaps and remove the technical barriers of the past. Through a three year (2020-2023), \$5 million investment from nine partner institutions, [EcoCommons](#) will transform ecological and environmental research by creating a trusted single platform for all digital modelling and analysis needs. To do this, it will leverage existing platforms and services, and introduce new user-centric functionality for cross-domain applications.

EcoCommons will build on the work of well-regarded existing platforms like the [Biodiversity and Climate Change Virtual Laboratory](#) (BCCVL), [ecocloud](#), the [Collaborative Species Distribution Modelling](#) (CSDM) and the training portal [ecoEd](#), which together support more than 7000 researchers based at over 400 different organisations in more than 35 countries worldwide. These platforms have had a significant impact on enabling access to ecological modelling tools, pre-configured environments and cloud compute power for a wide range of users such as students, teachers, researchers and key decision makers at state, national and international levels. The platforms have demonstrated the tremendous potential that unified tools, models and infrastructure can have on increasing efficiency, collaboration and consistency of scientific analysis, modelling and the sharing of outputs; and EcoCommons has been initiated to fulfill this.

The functionality of the EcoCommons platform will be designed to meet user needs. These include:

- access to thousands of otherwise disparate global and national spatial data layers (climatic, environmental and urban/social)
- easy visualisation and query of spatial data
- structured repeatable workflows
- support through point-and-click and command-line interface,
- and extensive compute power to run high-performance tasks.

This functionality will be reinforced and underpinned through scientific and technical user support and training and engagement activities, including partnerships with key international initiatives to ensure interoperability beyond the Australian border.

Ultimately, EcoCommons aims to significantly reduce the time and wrangling needed to get from data to decisions, and set best practice for modelling across many scientific domains and jurisdictions.

1.2 Purpose of this document

The purpose of this document is to provide a proposed baseline architecture for the EcoCommons Platform. This document outlines the architectural decisions resulting from the EcoCommons Australia Program initiative.

1.3 Program overview

EcoCommons was officially initiated in March 2020 as a three-year \$5 million partnerships program of:

- the [NCRIS-funded Australian Research Data Commons \(ARDC\)](#)
- [Atlas of Living Australia \(ALA\)](#)
- [CSIRO](#)
- [Centre of Excellence for Biosecurity Risk Analysis \(CEBRA\)](#), the University of Melbourne
- [Griffith University](#)
- [Macquarie University](#)
- [Qld Cyber Infrastructure Foundation \(QCIF\)](#)
- [TERN](#)
- [University of NSW](#)

With a mission to be the portal of choice to analyse and model ecological and environmental scenarios, EcoCommons will provide a service-based framework for deploying, orchestrating and re-using science-centric services, such as species distribution models and climate change projections. The project will provide a shared, stable and regulated commons that will offer enhanced computing and analytical capability through:

- A curated analytics toolbox
- Workflow and result management service
- Scientific and data workflow
- An integrated Data Explorer.

The program will build on the work of well-regarded platforms like the [Biodiversity and Climate Change Virtual Laboratory](#) (BCCVL), [ecocloud](#), the [Collaborative Species Distribution Modelling](#) (CSDM) and the training portal [ecoEd](#), which together support more than 7000 researchers based at over 400 different organisations in more than 35 countries worldwide. For more information on EcoCommons visit the [program page](#).

1.4 Stakeholders

The EcoCommons Implementation Plan lists relevant stakeholders.

1.5 Scope

In Scope	NOT in Scope
Platform Architecture	Deployment Pipeline
Re-development/Re-Engineering of current services to integrate into the new platform.	Requirements for individual components, as these are done in separate documents as referenced above.

1.6 Constraints

- No additional funding is currently available for the development of the strategy towards sustainability beyond the end of funding in March 2023
- Availability and quality of the identified data to be used in the platform.
- The drivers for the program are closely linked to funding constraints, obliging the program to concentrate on the delivery of high priority capabilities – those that maximise the benefits such as dedicated strategic funds to leverage further funds to go into sustainability or expand to an amount of collaboration with new interested parties.
- There are a finite number of people with specific domain and subject-matter expertise to design, build and deploy capabilities to serve the EcoCommons community.
- The platform needs to be developed to run on the available allocation of compute resources from project partners to run development, staging and production environments. ARDC is actively looking at resource requirements to support projects.

2. BASELINE SOLUTION ARCHITECTURE

2.1 Current services overview

There is currently no EcoCommons platform, but there are some separate services that are currently running and that will be re-engineered so that the new platform can support the functionality of those current services as one common system.

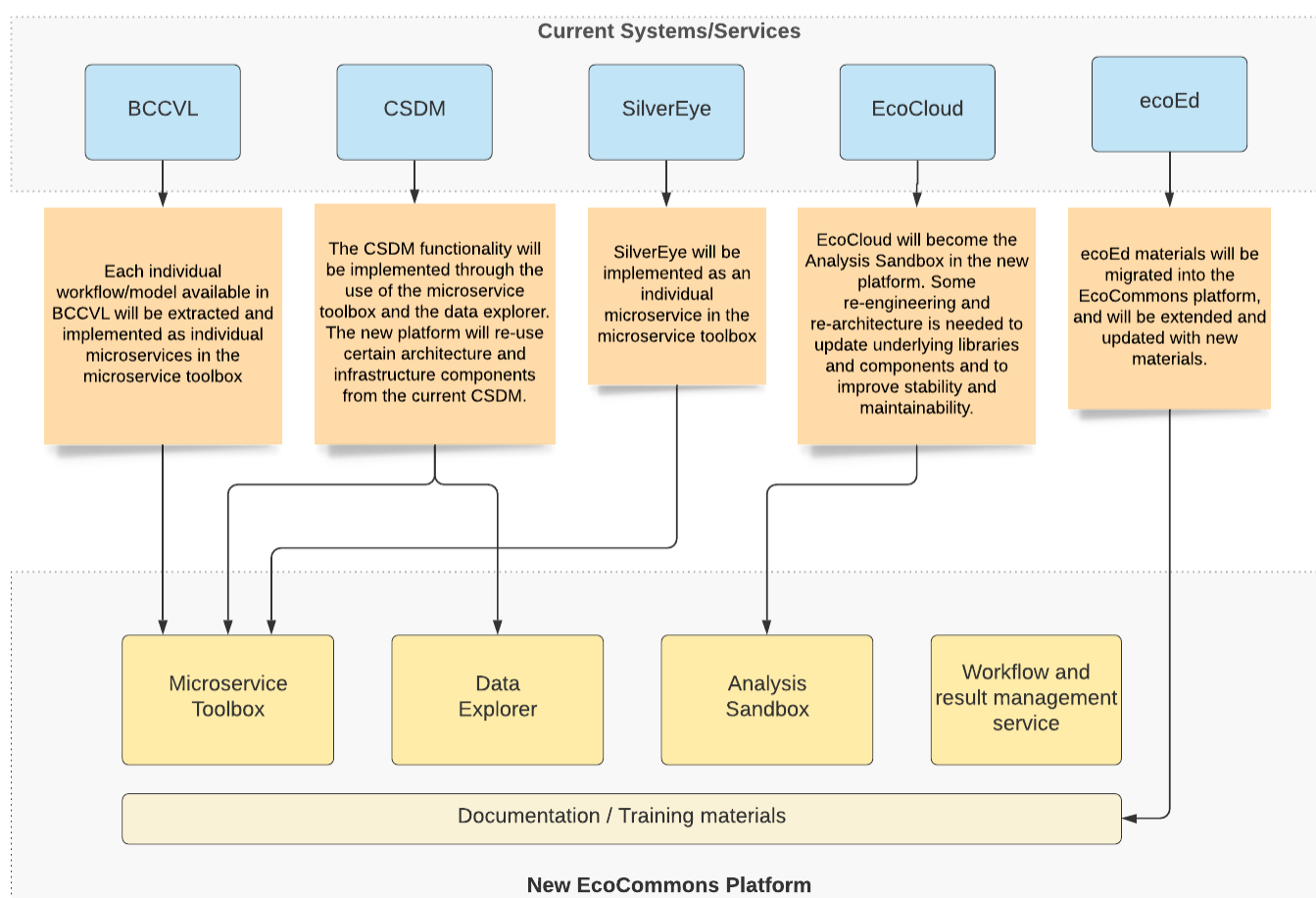
The current services are:

- **Biodiversity and Climate Change Virtual Laboratory (BCCVL):** bccvl.org.au
- **EcoCloud:** ecocloud.org.au
 - Including link to CoESRA Virtual Desktop - <https://coesra.tern.org.au/#/tern-landingpage>
- **ecoEd:** ecoed.org.au
- **Collaborative Species Distribution Modelling Portal (CSDM)** (prototype): csdm-test.bccvl.org.au
- **SilverEye:** prototype of cloud-based micro-service that allows spatio-temporal analytics. client.silvereye.bccvl.org.au

2.2 Integration / Migration Overview

Below is a high-level overview of current systems and what will be migrated into the new platform.

EcoCommons Integration / Migration Overview



2.3 Solution Description / Context

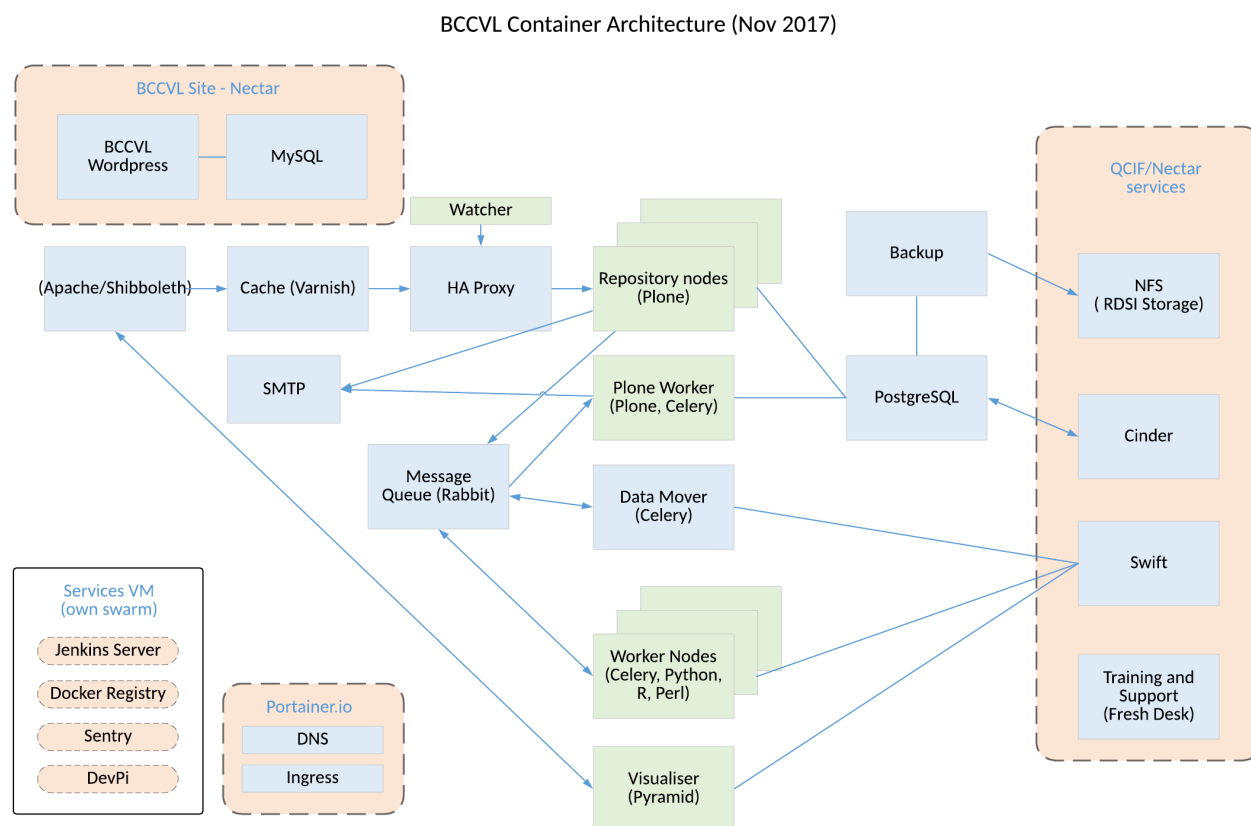
The following sections will provide an overview of each of the current services, and the technical plan on how to integrate them into the new EcoCommons platform.

2.3.1 Biodiversity and Climate Change Virtual Laboratory (BCCVL)

Description

With a full range of species and climate impact experiments the BCCVL offers researchers access to data, experiments, advanced configuration options and all new in-depth experiment outputs allowing users to fully investigate model outputs.

BCCVL Current Architecture



BCCVL Integration Plan

The current functionality in BCCVL will be made available in the new EcoCommons platform through the microservice toolbox component of the platform. The functionality of each of the scientific workflows currently available in BCCVL will be extracted out to individual, containerised, functions that will be made available in the platform. Individual microservices may contain several functions that are logically contained in the microservice, but all functions will be available to be re-used in other workflows. Microservices will be composed with cognitive complexity in mind, where balancing simplicity with communication and integration overhead, to create services that strikes the right balance.

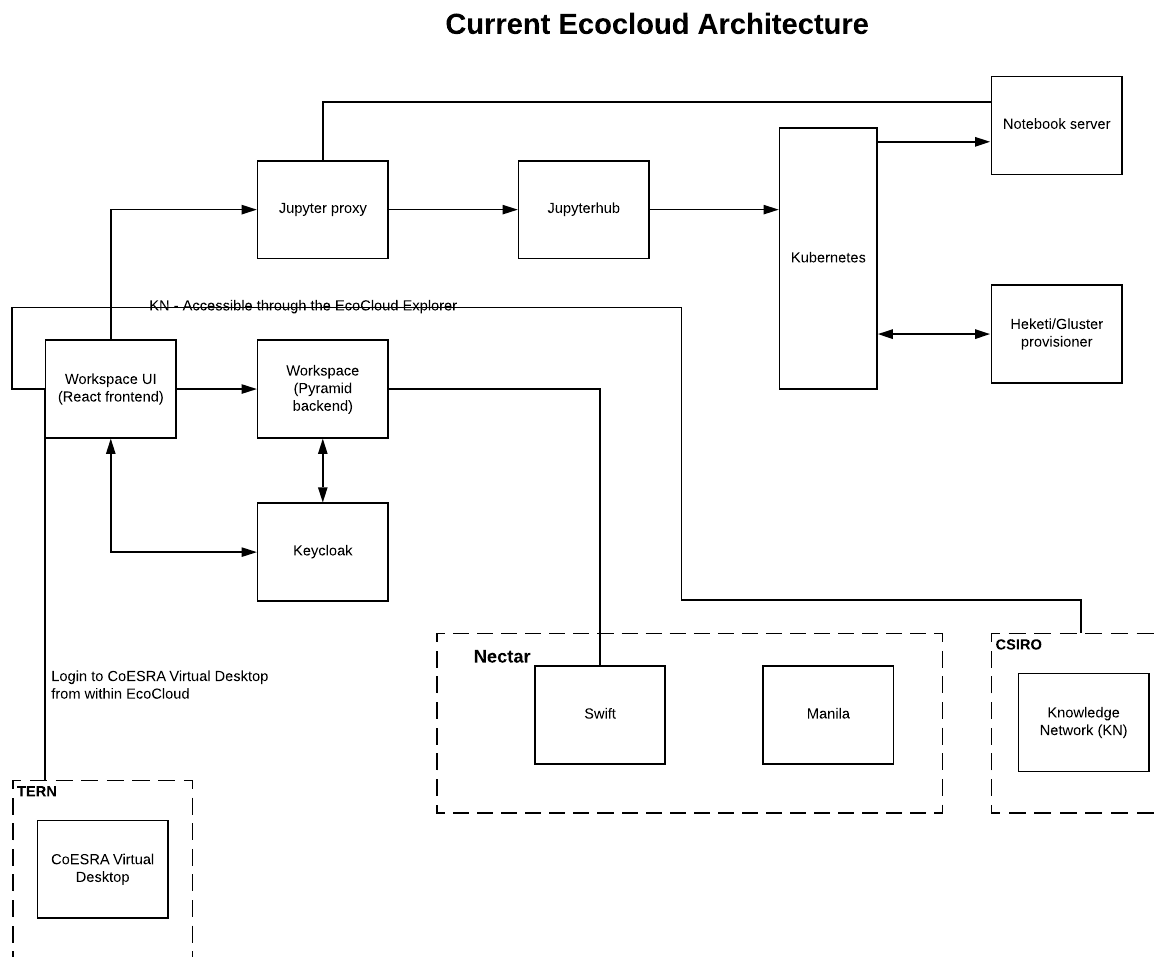
During the extraction process, each service will be reviewed to ensure that the best applicable standards are used for the models/workflows, and updates will be made wherever necessary. Each microservice function will have their own WPS API interface, in addition to a point and click interface, in which users of the platform will be able to interact with them.

2.3.2 EcoCloud

Description

EcoCloud delivers cloud based computing tailored to ecological data and researchers. EcoCloud provides a platform that brings together servers, storage, databases, coding languages, training, analytics and more – over the internet ('the cloud') to offer faster innovative solutions, flexible resources and ongoing support.

EcoCloud Current Architecture



EcoCloud Integration Plan

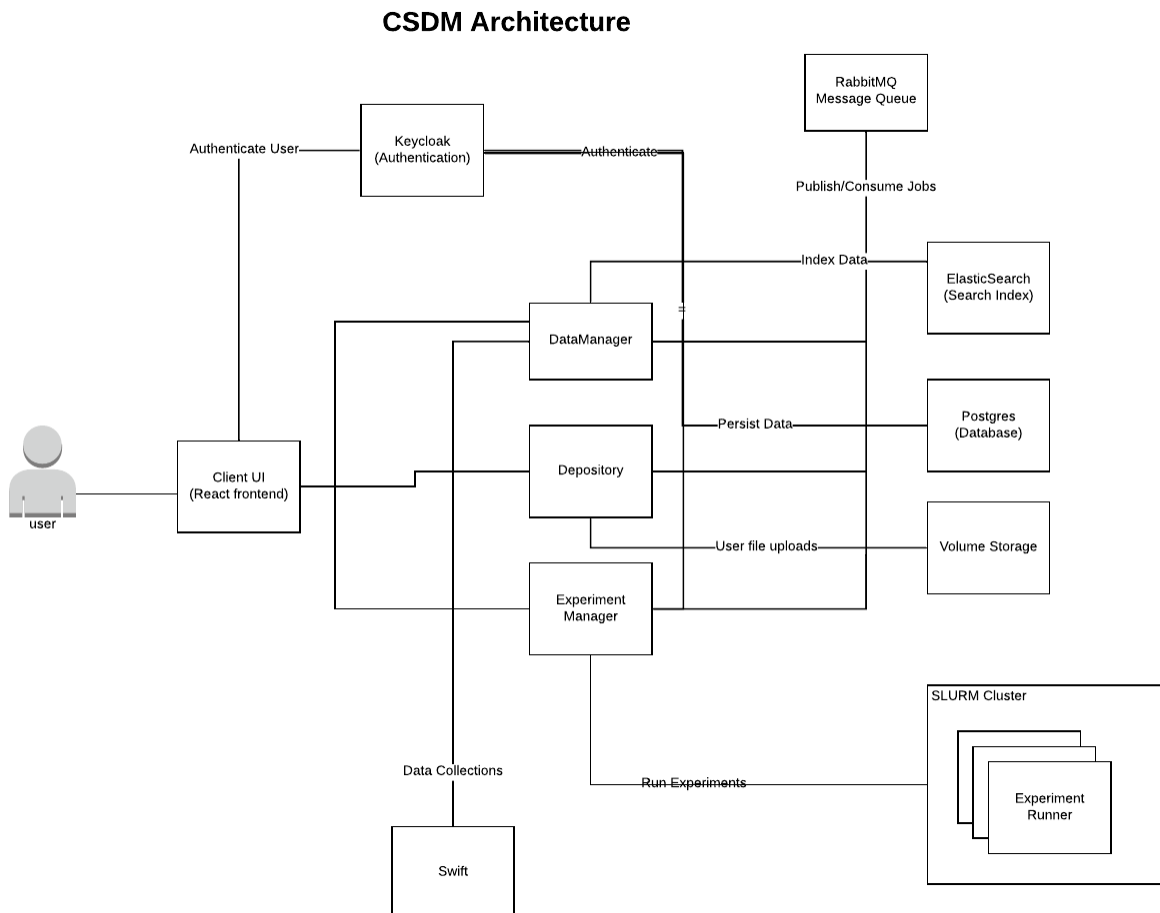
The functionality of EcoCloud will be made available in the EcoCommons platform through the Analysis Sandbox component of the platform. This component will contain the current functionality of EcoCloud, with some changes to the underlying architecture. The architectural changes will be done to increase stability and maintainability of the component, and to make it integrate with the structure of the new platform and take advantage of technological advances since the creation of EcoCloud.

2.3.3 Collaborative Species Distribution Model (CSDM) Portal

Description

The Portal focuses on the production of scientifically robust species distribution models (SDMs) supporting transparent and coordinated decision making by government at state and federal levels. The modelling workflows available through this Platform have been guided by an Expert Committee.

CSDM Current Architecture



CSDM Integration Plan

The integration of CSDM in the EcoCommons platform is two-fold.

Firstly the new platform will be utilising a similar (but not identical) platform architecture as the one used in the CSDM prototype. The implementation of Kubernetes configuration, and services/components like the containerised visualiser and experiment processing manager utilising a Slurm cluster for workload processing in CSDM - will be used for inspiration and a starting point when developing related components in the new platform. CSDM could indeed be viewed as a small prototype of a subset of the new EcoCommons platform, and ideas and findings from creating CSDM will be used in the creation of the new platform.

Secondly, once the EcoCommons platform has been developed, the current CSDM prototype/proof of concept can be migrated to use the EcoCommons platform components such as microservice functions and workload processing instead of being deployed completely separately.

2.3.4 SilverEye

Description

"Daily Climate Grids Microservice"

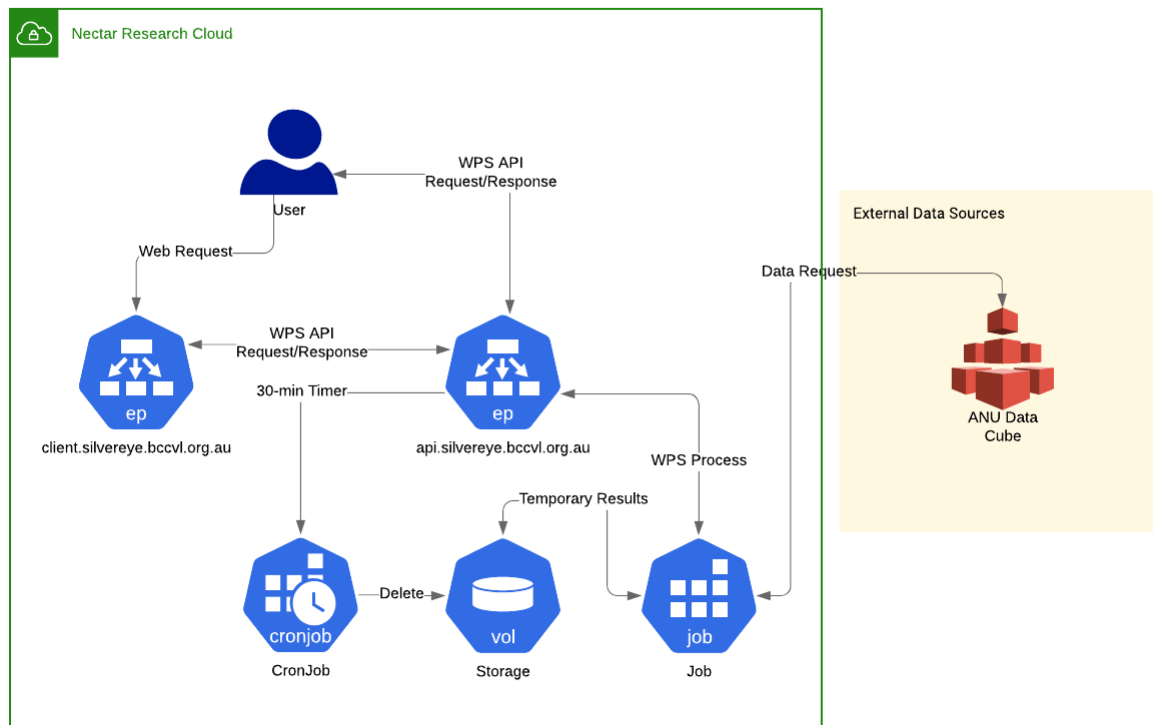
SilverEye is a cloud-based micro-service that allows spatio-temporal analytics. At present, the service provides access to ANU Climate Data Grids for Australia. It has a front-end and a back-end component.

The front-end is for easy point-and-click access. The back-end has a standard OGC:WPS API through which the user specifies a selection of latitude, longitude, and time coordinate regions for extraction of the data variables: maximum temperature, minimum temperature, rainfall, solar radiation, and vapour pressure. It offers a variety of time-series selection patterns.

A statistical aggregation function is applied to the extracted data, and the results are returned to the user in multiple formats, including: netCDF, CSV, geoTiff, PNG 2D-line plots, and Zarr files.

SilverEye Current Architecture

SilverEye Overview



SilverEye Integration Plan

The functionality of the SilverEye tool will be implemented as part of the microservice toolbox and will be available to users of EcoCommons platform that way. It has already been developed with the new platform structure in mind, so integration should be relatively straightforward.

3. BASELINE TARGET SOLUTION ARCHITECTURE

3.1 Solution Architecture for the Proposed Solution: EcoCommons

The new platform will have 4 main areas of functionality:

1. Microservice Toolbox
2. Workflow and result management API service
3. Data Explorer
4. Analysis Sandbox

It will be architected to support and connect these 4 areas of functionality, whilst allowing for adding additional features and functionality in the future.

In addition to the 4 main areas of functionality, the platform will also have the following core functionality to support running the platform.

- User Management/Authentication
 - This will enable the ability to add/authenticate users and roles/permissions for the platform. Includes management of both access to the platform via UI and via API.
- Message Queue
 - The platform will be designed around a central message queue that will help connect the various services of the platform. Individual platform services will be able to post messages to the queue, and also to pull messages from the queue for processing.
- Database
 - A central database on the platform will enable the persistence of configuration options, certain metadata and statuses.
- Search Index
 - The platform will use a search index to enable better and faster search of metadata collections in the user interface. Various search indexes will be created to be used by individual components such as data explorer, microservice toolbox, workflow & result management service etc.
- Monitoring
 - A suite of services to enable monitoring and logging on the platform will be installed and configured.
- Reporting
- Job/Workload processing:
 - This is the part of the platform/EcoCommons that will run the compute intensive experiments/modelling/calculations. We choose to separate the architecture of the workload processing from the main platform architecture to allow better flexibility to scale resources, and to potentially use external clusters/compute resources to do processing. The job handling for the workload processing will initially be developed to run jobs from the microservice toolbox, but additional work will go in to attempt to generalise the access to

the workload processing so that it could be used by additional components of the platform such as the Analysis Sandbox.

- Works by setting up one or many schedulers that would be available to pull new jobs from the queue and process them with the available workers.

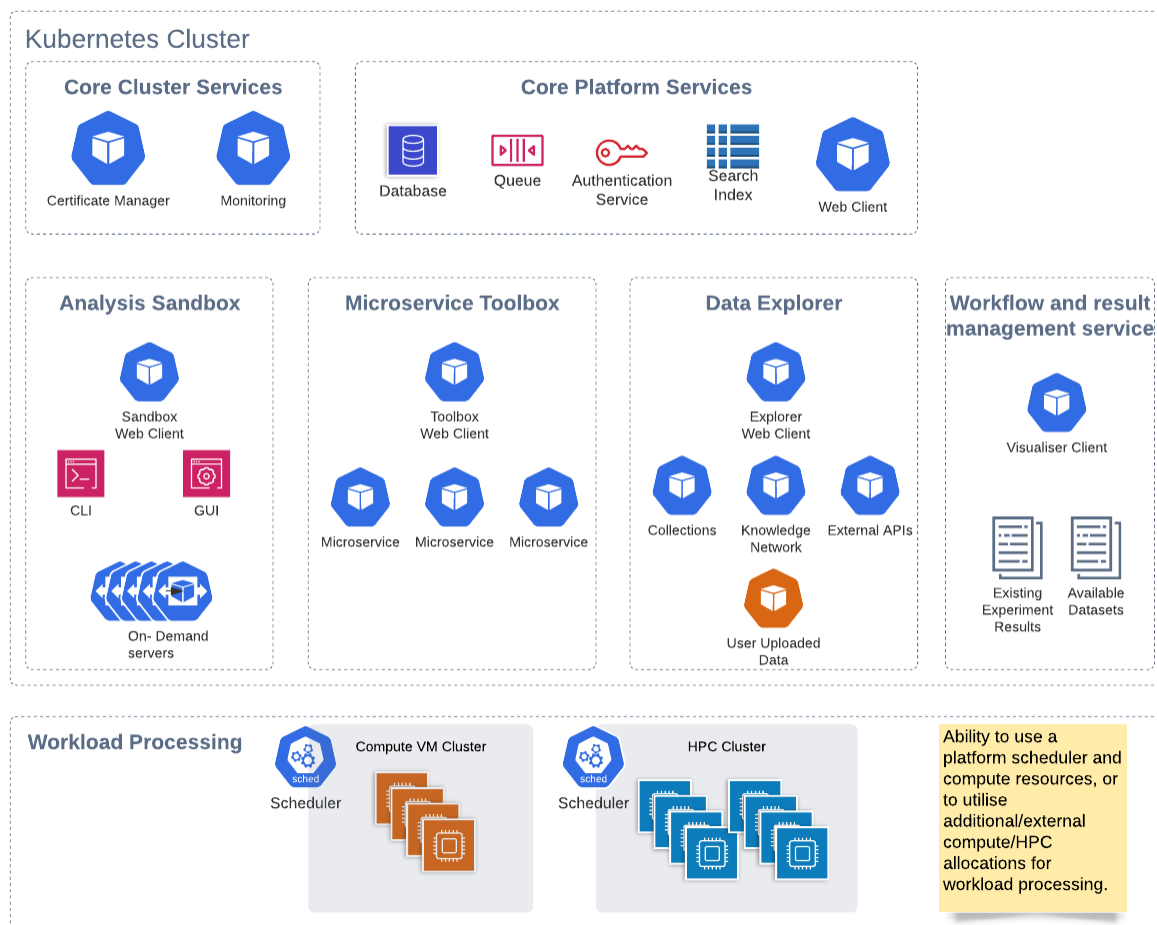
The platform will be a set of containerised services deployable on a Kubernetes cluster. This will allow for both easily scaling the cluster by adding additional resources, and also the ability to deploy the platform, or part of the platform on private/separate clusters if needed.

Kubernetes is a mature, well documented container orchestration software. It is currently being used worldwide, and is seeing a lot of uptake in Australian higher education/research sectors.

Services within the *microservice toolbox* will be accessible via the WPS interface standard from approved external clients where possible. The platform will track usage of its resources both via UI or API to ensure correct usage metrics can be extracted and used for KPI measurement and improvement.

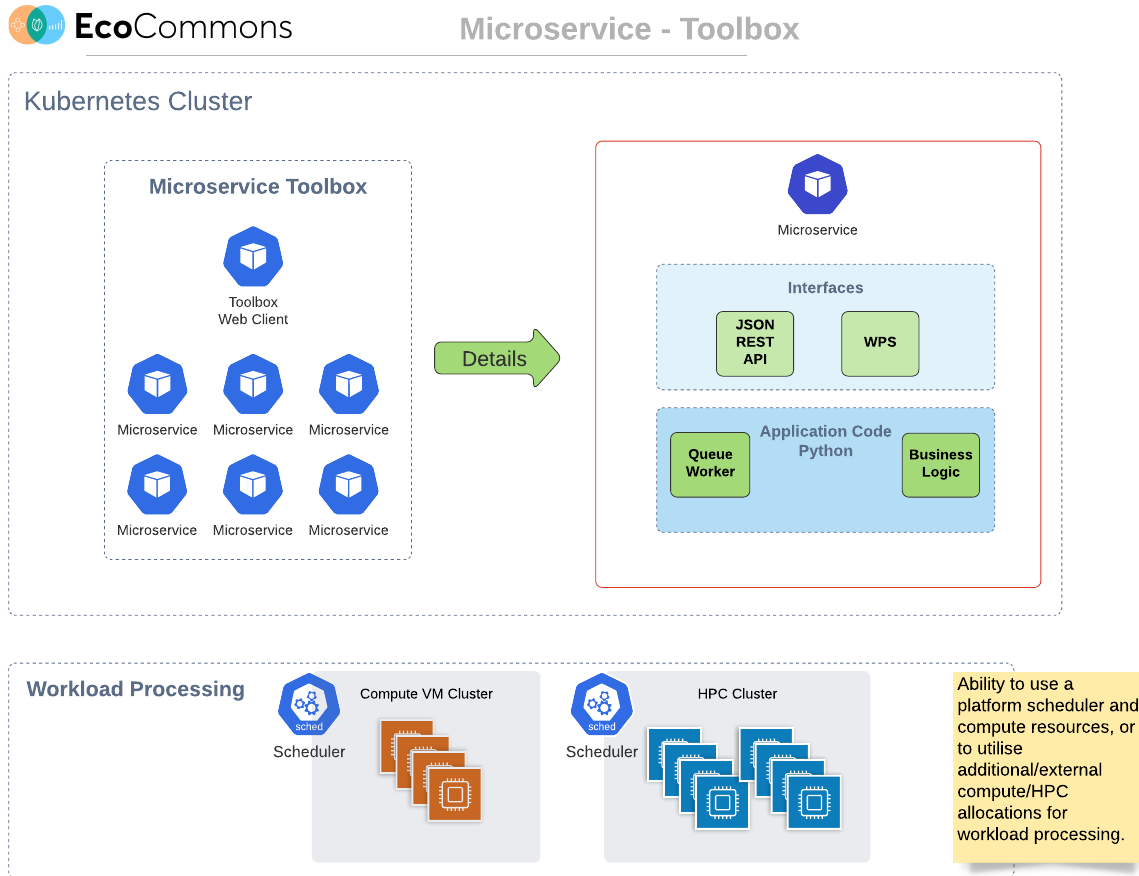
3.2 Proposed System Architecture Baseline

EcoCommons High Level Platform Architecture



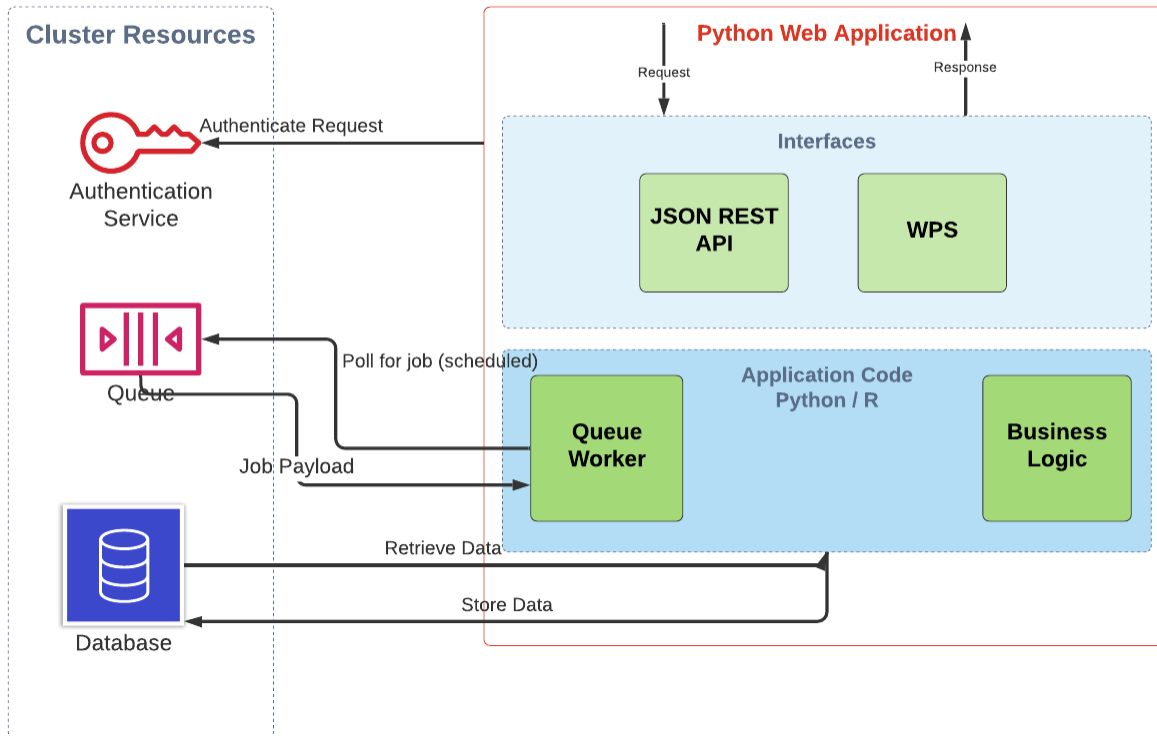
3.3 Proposed Architecture for Microservice Toolbox

Below is a high-level diagram showing Microservice Toolbox inside the platform.



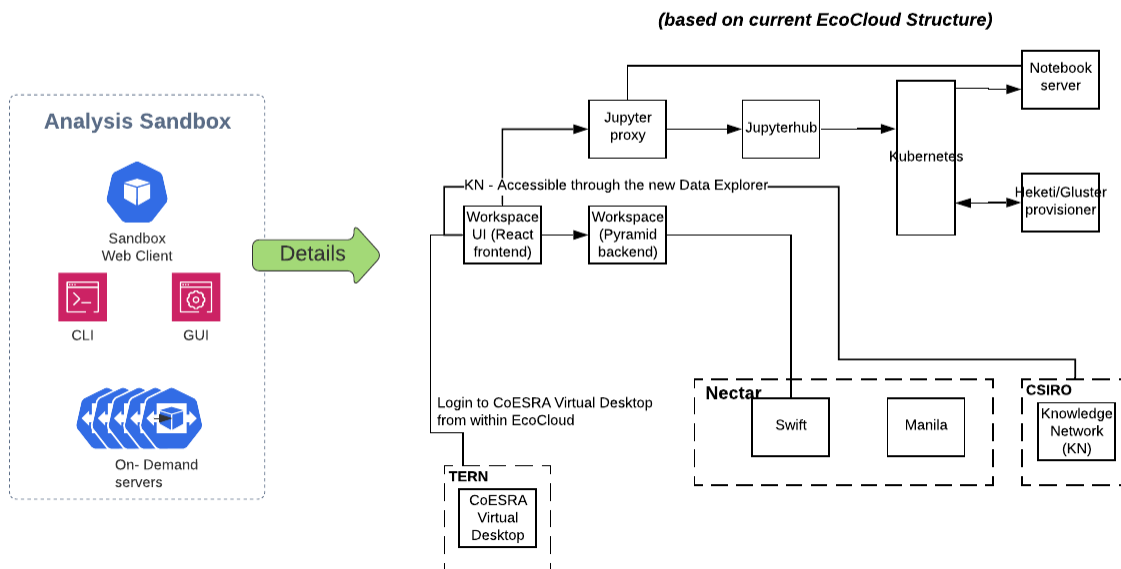
3.4 Proposed Architecture for platform Microservices

Below is a high-level diagram showing individual Microservice structure inside the toolbox.



3.5 Proposed Architecture for Analysis Sandbox

The below diagram shows how the Analysis Sandbox will sit inside the platform. Details on the right are showing the current internal structure of EcoCloud. Most of this structure will remain the same, but some re-engineering and updates to provide better stability and to integrate with platform infrastructure will be done.



3.6 Implementation Plan

This is a 3-year project which was initiated in March 2020. The EcoCommons Australia core development team was assembled in May 2020.

The first year of the project will be dedicated to building the new platform, and integrating the existing functionality of BCCVL, EcoCloud, and CSDM to enable the migration of users onto the new EcoCommons Platform. This Platform will consolidate all existing services and provide space for expanding under addition of new functionality.

The last 2 years of the project will be focussed on extending the capabilities of the platform to new domains such as Biosecurity and Agriculture, collaboration, interoperability with existing platforms and training.

3.7 Migration Plan

Overall, all current services will be migrated in some form into the new platform, and after verification of successful migration, the existing services will be turned off.

Migration plan breakdown

Data

- User generated data such as experiments and results **will not be migrated** from the existing systems (BCCVL, EcoCloud, CSDM) into the new architecture.
- User account data **will be migrated** to allow for continued access for current users.

Functionality

- **BCCVL** - Current functionality of BCCVL will be available through the Microservice Toolbox in the new platform
- **EcoCloud** - Current functionality of EcoCloud will be available through the Analysis Sandbox in the new platform.
- **CSDM** - Current functionality of CSDM will be available through the Microservice Toolbox in the new platform
- **SilverEye** - Will be available as a microservice in the microservice toolbox.
- **ecoEd** - Current content and functionality will be available through the new platform.

Integrations

- Analysis of existing services and integrations that use the current services, and design a migration plan to start integrating with the new platform when ready.

Users of the current services will be notified in due course to allow them to export/download their data before the existing systems are turned off. Please see the integration/migration overview from section above.

User acceptance testing will be done to ensure required functionality exists and works in the new platform before existing systems are turned off.

A comprehensive set of functionality tests are being developed for the existing systems, and these tests will be used to create a new set of tests that will measure how the new platform is able to provide similar functionality as the existing systems.

To start with, these tests will be done manually by team members, but automated tests will be gradually developed and included as part of the deployment pipeline for the platform.

3.8 Implementation and Migration controls

Control	Yes, No, N/A	Notes (if applicable)
Functional and integration testing	Yes	
User acceptance testing	Yes	
Load testing	Some	
Version control	Yes	Git
Migration rollback / reversion plan	Yes	

4. INTERFACES

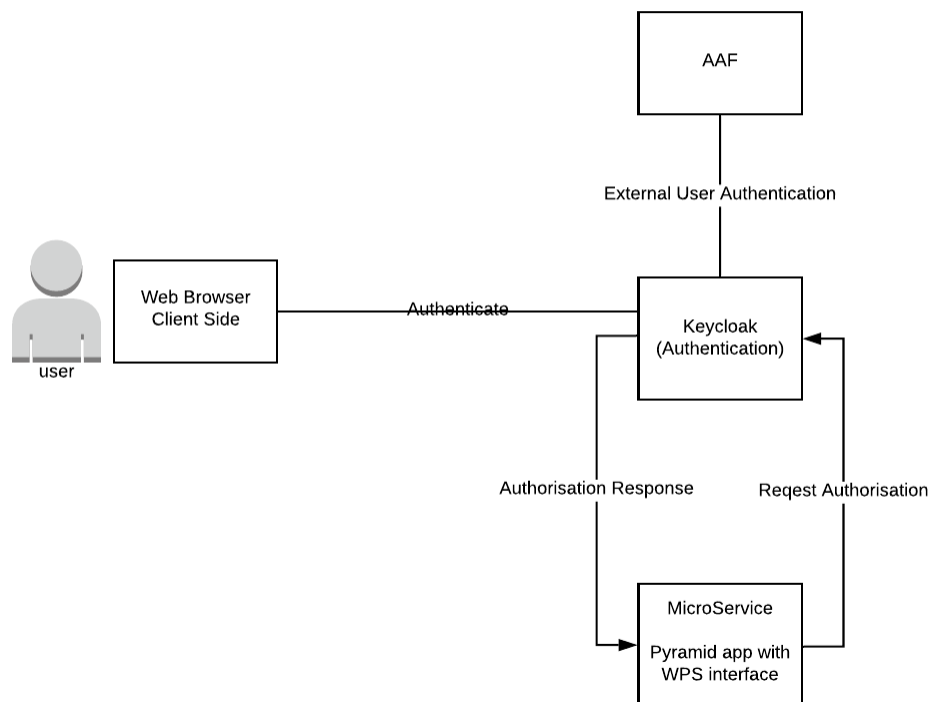
4.1 User Interfaces

The platform will be accessed through the following user interfaces:

- Web Client - <https://app.ecocommons.org.au>
- Direct MicroService Access via API (JSON or WPS) via HTTPS (approved clients with valid API token only)
- Custom portals/virtual laboratories can be created 'on top' of the platform that provides the user with access to a customised front end client with additional context and clarifications for specific use cases. An example of this would be the creation of a custom front end to provide a BCCVL like virtual laboratory.

4.2 API Interface Example

EcoCommons - API Interfaces



Elements of note in the diagram include:

- AAF: Australian Access Federation
- Tokens used on both client and server side
- Tokens are only stored on the client side for the current session
- Tokens are processed and utilised using standard JWT (JSON Web Token) standards

5. INFORMATION SECURITY

5.1 Information asset register

- User uploaded datasets.

Requirements for this will be developed with regards to privacy, security, licensing and storage requirements when developing the data explorer component of the project.

5.2 Data Entities

Entity	Description
User Experiment data container	
Curated Set of data models	
User uploaded data	
System User accounts	

5.3 Information Security Controls

Abbreviation(s): Australian Access Federation (AAF).

Control	Description
Information Confidentiality Controls	
Database Column Encryption	When any reads or writes are undertaken between the application and the database, it is encrypted using the SQLAlchemy-Utils class. This uses the AES Engine (AES 256) with pkcs5 cryptography. This key is stored in Kubernetes as a secret in accordance with its solution architecture. A copy is also kept in the project groups password manager tool.
Information Availability Controls	
Availability Target	99.9%
User Authentication	Users will authenticate via either AAF or via local user record
Recovery Time Objective (RTO)	48 hours
Recovery Point Objective (RPO)	48 hours (nightly backups)
Performance Effect on Other Systems	None
Growth and Scalability of System	Very Scalable
High Availability or Disaster Recovery	Disaster Recovery and high availability required
Availability Monitoring	Provided by the project team by using cluster monitoring tools such as Prometheus, Grafana and alert manager.
Load Testing	Provided by the project team by running scheduled load tests on the system using tools such as Locust.
Encryption-in-Transit	HTTPS
Encryption-at-Rest	AES encryption of Database

6. SUPPORT

6.1 Model for support

The model for support is documented below.

Support Tier	Details	Responsible Actor
T0	Online Documentation/FAQ	Project Team
T1	Application Assistance	Project Team
T2	All other Help	Documentation
T3	Kubernetes support	Project Team
T3	Infrastructure	Project Team/Project Partners

6.2 Solution Lifecycle

The initial project length is 3 years, which should result in a mature and fully working platform in a 'business-as-usual' state that should be operational for at least 7 years in total.

The sustainability and support model for post project operation is currently being discussed and formulated and will be made available as soon as it is ready.

7. COMPLIANCE

7.1 Policy and legislation compliance

At a minimum, the platform will comply with the following policies and legislation:

- FAIR Data Guidelines for Project Data Outputs
(https://ardc.edu.au/about_us/policies-and-guidelines/fair-data-guidelines-for-project-data-outputs/)
- Information Security Policy
(<http://policies.griffith.edu.au/pdf/Information%20Security%20Policy.pdf>)
- Information Management Framework
(<http://policies.griffith.edu.au/pdf/Information%20Management%20Framework.pdf>)
- Information Privacy Act 2009
(<https://www.legislation.qld.gov.au/legisltn/current/i/infopriva09.pdf>)
- Information security - IS18 (<https://www.qgcio.qld.gov.au/products/qgea-documents/549-information-security/2704-information-security-is18>) Use of Standards

8. ADDITIONAL INFORMATION

8.1 Reusability

All components and services developed, will be developed according to the "FAIR Data Guidelines for Project Data Outputs" https://ardc.edu.au/about_us/policies-and-guidelines/fair-data-guidelines-for-project-data-outputs/

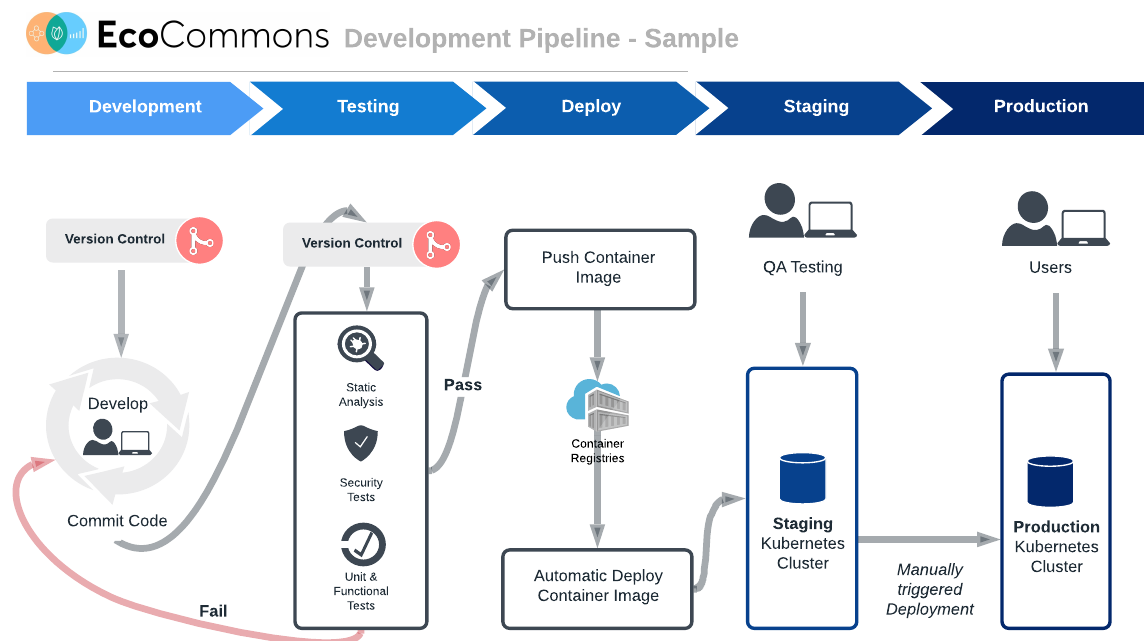
8.2 Risks

- Allocation of compute resources not sufficient or in time for project requirements
- As this is a cutting-edge project, there are unknowns when making initial decisions and assumptions for architecture and technology. These assumptions could prove to be incorrect.
- Data agreements are not in place or supported by policies
- NeCTAR NCI cloud infrastructure will be closing in September 2020 which requires the migration of our virtual laboratories to a new NeCTAR node (TPAC). This is uncharted territory.
- QrisCloud is not funded to support applications in the long term. Presently the EcoCloud and CSDM cluster is hosted on QrisCloud.
- External services and collaboration partners have different priorities; this may result in conflict or inability to meet obligations.

8.3 Development Process

The development team will use an agile development methodology, and heavily use automation of integration and deployment processes. This ensures that security tests, functional tests and integration tests are run before individual components and containers are deployed.

The diagram below illustrates the development process the team uses.



8.4 Output

All source code, tools and containers will be available through the projects' Gitlab group with appropriate licensing information attached and will adhere to *FAIR* practices.

8.5 Acknowledgements

Please include the following acknowledgement when mentioning EcoCommons:

"EcoCommons is a collaborative initiative, enabled by investment from nine partner institutions: Australian Research Data Commons (ARDC), Atlas of Living Australia, Centre for Biosecurity Risk Analysis, CSIRO, Griffith University, Macquarie University, QCIF, TERN, University of NSW. QCIF is the lead organisation. EcoCommons received primary investment (<https://doi.org/10.47486/PL108>) from the ARDC. The ARDC is funded by the National Collaborative Research Infrastructure Strategy (NCRIS)."

8.6 Referencing this document

Please use the following citation for this document:

EcoCommons Australia, 2020. *Ecocommons Solutions Architecture: Baseline Version*. 1.0. [online] Brisbane: EcoCommons Australia, pp.1-23. Available at: <(insert link)>