FRONTIER SI

GROUNDWATER DEPENDENT VEGETATION PROJECT STAGE II

REQUIREMENTS DOCUMENT & OPTIONS REVIEW

17 November 2020

Background

The Problem

Mining companies must monitor the potential impact of mine dewatering on groundwater-dependent vegetation (GDV) as a compliance requirement. Mine dewatering is the removal of unwanted groundwater from a mine to allow rock and mineral extraction from beneath the water table. In some circumstances, this can affect the health of GDV in the vicinity, as many plant species rely upon a stable water-table for seasonal water requirements. Mapping the spatial distribution of GDV species at the tenement scale, and appropriately selecting monitoring sites to facilitate assessment of dewatering impacts on GDV, requires comprehensive knowledge and extensive time to set up. While earth observation data has been trialled to monitor the effects of dewatering on GDV health, its potential is still emerging, and no routine or user-friendly tool yet exists.

The Research

An initial project (Stage 1) conducted by Curtin University researchers, in consultation with mining companies (Roy Hill Iron Ore and BHP), used a completely knowledge-based approach to reduce the "search space" for GDV. The research team successfully trialled the model at three discrete locations in the southeast Pilbara region. Initial validation statistics indicated that the model predicts GDV at an accuracy exceeding 86%, and the team believe with further improvements accuracies of >95% could be attained.

The Stage 1 project enabled the rapid collection of a significant amount of sampling, which is being used in a Stage 2 project currently underway to assist data-driven modelling for prioritisation of ground-based response by folding together multiple remotely sensed data sources. The Stage 2 project will also include near real time detection of GDV health decline by utilising 40 years of image data. These outputs will assist prioritisation of ground-based response and better inform dewatering strategies.

The Stage 1 project was delivered as a Jupyter notebook, designed to run on the Digital Earth Australia Sandbox, hosted by Geoscience Australia. This notebook can be found at https://github.com/lewistrotter/GDV_TOOL/. This can be leveraged by developers of an operational tool (this project).

As part of the second Stage of research, the project team are seeking to develop and deploy a user-friendly tool to operationalise the Curtin University research. The development of the GDV Tool is considered as part of one of the work packages within a FrontierSI Collaborative Research Project and will therefore need to meet project timelines. The timelines for development and recommended approach are outlined further in this document.

End-user Consultation Process

The project team conducted a two-part stakeholder consultation process as part of the Stage 2 project. Stakeholder Consultation 1 sought to understand the workflows in which project participants use Earth Observation data for GDV detection and monitoring. This information has helped inform the technical requirements the software needs and options for development of a suitable solution.

Stakeholder Consultation 2 was then undertaken where development options were reviewed, and feedback sought from project parties to help inform the future development.

Document Purpose

The purpose of this document is to provide information to prospective solution providers to help inform the requirements for the GDV tool development. This document is presented in two parts:

1. The requirements for tool development
2. An overview of the tool development options and project stakeholder feedback

PART 1: REQUIREMENTS

1. Overall Objective of the GDV Tool

The main objective of the GDV Tool to be developed is to identify groundwater dependent vegetation and monitor it over time in a cost and time effective manner. The tool will assist prioritisation of ground-based response and better inform dewatering strategies. In addition, it is expected that future functionality as an outcome of additional research and refinement can be incorporated into the tool over time. Therefore, ensuring the technical architecture is scalable will be an important consideration. In more detail, this tool will:

   a) Identify GDV Likelihood;

   b) Detect trends and change over time; and

   c) Provide triggers when early changes are detected.

2. Prioritisation of Requirements

To ensure proponents are clear on the components that are priorities and those that could be flexible as part of the delivered solution we denote each component with either Must, Should, Could or Would, which are defined as:

MUST - Must have this requirement to meet the business needs

SHOULD -Should have this requirement if possible, but project success does not rely on it

COULD - Could have this requirement if it does not affect anything else on the project

WOULD - Would like to have this requirement later, but delivery won't be this time

3. Development Technology Readiness Level 8-9

Code from the Stage 1 project was delivered by Curtin University researchers as a Jupyter Notebook, designed to run on the Digital Earth Australia Sandbox, hosted by Geoscience Australia. This notebook can be found at https://github.com/lewistrotter/GDV_TOOL/. This can be leveraged by developers of an operational tool (this project). The research being developed in the Stage 2 project may also be developed in Jupyter Notebook; however, this is yet to be determined.

The software solution must be delivered to a Technical Readiness Level (TRL) 8-9 as outlined in the graphic below. The tool should be able to be operational and ready for deployment by a preferred service provider. (MUST)

It should be noted that ongoing service and maintenance is not expected as part of this project, however the cost to provide the solution as a service will be a factor considered as part of the assessment process.
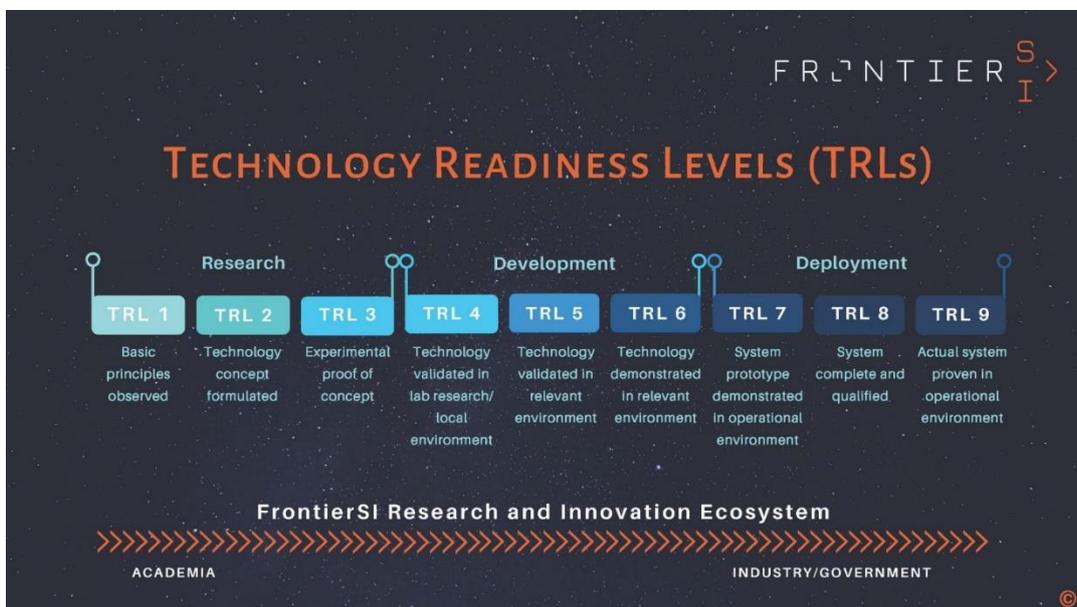
4. Key Functional Requirements

4.1 Deriving a GDV Likelihood Layer

The GDV likelihood layer is the core component of Stage 1 of the project and stands alone as having the advantage of not requiring any upfront information. Stage 2 also creates a GDV Likelihood layer, but it is a data driven (e.g. known presence/absence locations, species) ensemble of models. Each of the models in the ensemble may also stand alone and have parameters that can be modified by the users to achieve a more accurate result on their tenement.

A relatively small subset of high-resolution mapping of known GDV species can also be used to map their fractional presence in lower resolution, but larger footprint imagery (e.g. freely available Sentinel-2 data). The subset layer could come from a range of sources, such as commercial satellite data, aerial imagery or field surveys. Originally the system was envisaged to directly use hi-resolution imagery, but due to the set of constraints around licensing of the data, the creation of the high-resolution areas of GDV is proposed to be done locally by the user, outside of the system.

4.1.1   The architecture must support the components identified in Figure 1 (see exceptions in 4.1.4). (MUST)

4.1.2   Users must be able to view the layer as a web map or layer in a suitable desktop environment such as ArcGIS. (MUST)

4.1.3   Users must be able to download this data for their own purposes such as offline field use, or ingestion to an internal library. (MUST)

4.1.4   The architecture must be able to support users to upload their own high-resolution data (subject to licensing requirements) to upscale lower resolution imagery over their own area of interest. (MUST).

Figure 1: Components of the GDV Likelihood layers
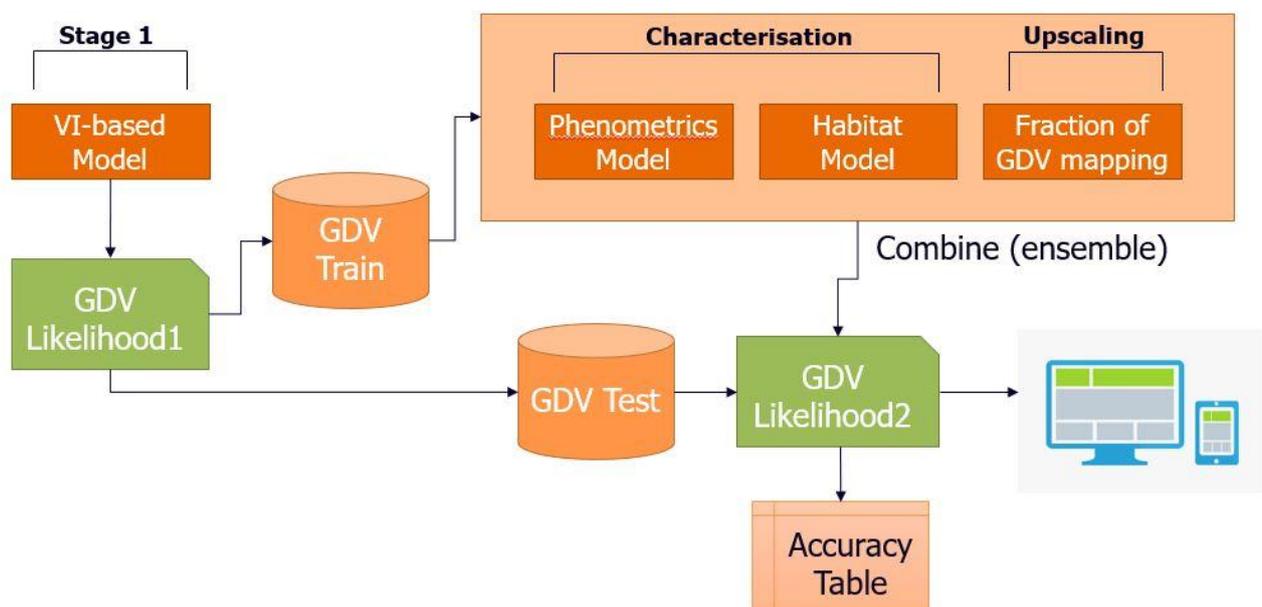
4.2 Trends and Change Detection Analysis

The trends and change detection analysis component allows users to explore trends and changes in vegetation over time, using the full Sentinel-2 and Landsat archives. For planning purposes and historical investigation, the ability to view the timeseries of a vegetation index for a given site allows a baseline to be established. For on-going monitoring, the ability to compare sites with the historical timeseries, or with control sites situated outside areas of impact allows users to asses any changes.

4.2.1   The architecture must enable users to define sites as both single pixels/points, or as polygons representing riparian zones, stands of trees etc, either by uploading exiting shapes or drawing directly on a map. (MUST)

4.2.2   The architecture must enable the timeseries results to be viewed as a graph as well as downloaded in a common format such as CSV for local analysis with other data sources. (MUST)

4.3 Real Time Monitoring

This component uses near-real time data, most likely from the Sentinel-2 sensors, to provide the current state of GDV areas. These areas are user-configurable, as with the Timeseries Analysis component above. Warning triggers may be set by configuring a threshold change in comparison to either a historical seasonal average or a control site.

4.3.1   The architecture must have the ability for users to set threshold and be notified when those thresholds are exceeded by email and an API call suitable for integration into external systems. (MUST)

4.3.1   **The architecture must deliver a** dashboard to easily show the status of each defined area, such as with a traffic-light system, as well as being able to show the status of a group of areas or entire mine site. (MUST)

5 Non-Functional Requirements

5.1 Integration with other tools

Stakeholders expressed a desire for integration with their existing tools where possible. All users made use of ESRI tools with some also using Google Earth Engine or eMapper.

Users must be able to integrate their existing tools (as noted above) where possible with the developed solution. (MUST)

5.2 IT cybersecurity (software)

Stakeholders expressed a range of issues around organisational cybersecurity. This involved:

- Extensive procedures for installation of software within the organisation.
- Limitations on accessing websites and services hosted externally to the organisation's network.
- Controls and policies of transferring data to external systems.

In some cases, approval by the IT security team would be provided once risk factors of the system had been addressed. Stakeholders indicated this could often be a lengthy process.

The architecture design should ensure cybersecurity (i.e. points noted above) is considered. (SHOULD)

5.3 Data security

5.3.1 The architecture design must ensure the privacy of use of the data for each user. (MUST)

5.3.2 The architecture design must consider the licensing terms of the data used in the system. (MUST)

5.4 Ease of access to satellite data

The solution must be able to access large quantities of satellite data, whether co-located with compute resources or accessed remotely, to be considered. (MUST)

5.5 Performance

The tool must be able to provide mapped results to users in an acceptable timeframe for the task being performed. (MUST)

5.6 Suitable for a range of user expertise

Users of the GDV tool are likely to range from environmental management personnel perhaps with limited GIS experience to users with a high level of GIS expertise. The system developed should aim to enable all levels of users to interact regardless of expertise. (SHOULD)

6 Preferred Approach

The development of the GDV Tool is a discrete Work Package within a FrontierSI collaborative research project. A project agreement has been executed with an outline of the project timeline and deliverables. The timeline for development (this project) and preferred approach is provided below.

Timing	Activity/Deliverable
March 2021	Contracting completed
April- September 2021	Develop application involving a series of sprint cycles to develop the interface iteratively. This will overlap with the Curtin University research workplan.
30 September 2021	Deliver product demonstrating GDV likelihood layer
October 2021 – March 2022	Continue to develop application involving a series of sprint cycles to implement complete functionality.
31 March 2022	Deliver fully developed GDV tool.

It is important to note that development will need to be done in close collaboration with Curtin University researchers because timing of the development of the models will be in parallel with the GDV tool development.

7 Open source

There is a requirement within the project agreement from the project partners that all deliverables be open source and licensed under an Apache 2.0 licence. However, alternative models could be explored as long as the models and algorithms developed by the research team and a DEA Sandbox version or openly available option of the tool be made available to interested parties.

8 Constraints

- Timing constraints: As outlined above with a final deliverable due in March 2022.
- Budget: The allocated budget for this work is $80,000

PART 2: OPTIONS REVIEW AND STAKEHOLDER FEEDBACK

1. Stakeholder Consultation Process

The stakeholder consultation process was conducted in two parts.

Stakeholder Consultation 1

The focus of Stakeholder Consultation 1 (SC1) was to understand the workflows in which project participants use the GDV data, and how they use it within those workflows. Stakeholder consultation 1 comprised design-led collaboration workshops with the mining- focused project partners to understand the use of the GDV data. The online collaboration software 'Mural' was used for these workshops (a sample Mural board is included in Appendix 1). This enabled project partners to describe the workflows in which GDV data are used, as well as the particular applications. The workflows and applications utilising GDV data identified through SC1 are summarised in Figure 2. Brief descriptions of the GDV applications from the figure are also provided in Table 1 below. The information collected was sense checked with the Department of Biodiversity, Conservation and Attractions (DBCA) and the Western Australian Biodiversity Science Institute (WABSI). The information was also used to update NGIS and Astron as parties with interests in use of the data from the perspective of possible provision of the software solution at a later Stage of the project.
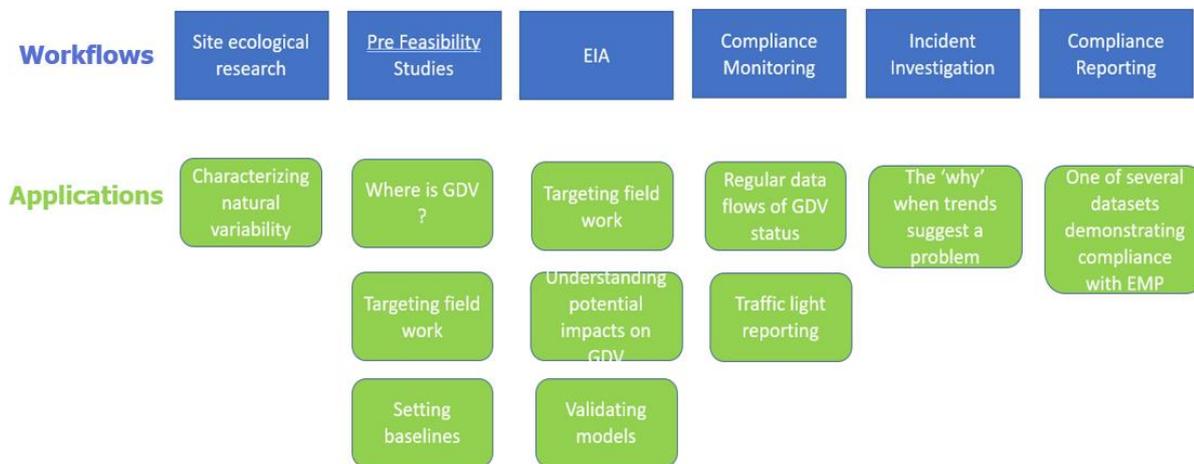
Summary of GDV Data Use

Figure 2. Summary of workflows and applications utilising GDV from Stakeholder Consultation 1

GDV Applications

WorkFlow	Application	Description
Site ecological research	Characterizing natural variability	Natural variability in vegetation condition is important to know so that appropriate thresholds can be set for compliance assessment
Pre-feasibility studies	Where is GDV?	Identification of GDV is important to planning of future operations
	Targeting field work	Enables efficiencies through concentrating field efforts on known areas rather than searching for areas of importance
	Setting baselines	Understanding variability in vegetation condition prior to any operations so change can be monitored
EIA	Targeting field work	As above
	Understanding potential impacts on GDV	One input into models for understanding what environmental impacts future operations may have
	Validating models	Used as data for validating models predicting environmental impact of future operations
Compliance Monitoring	Regular data flows of GDV status	Used as an index of vegetation health which can be regularly viewed
	Traffic light reporting	Regular monitoring can flow into operational decision making
Incident investigation	The 'why' when trends suggest a problem	If a vegetation problem is identified GDV can be used as one dataset in a forensic analysis to discover causes
Compliance reporting	One of several datasets demonstrating compliance with EMP	Useful measure of one element of environmental impact since it can be regularly acquired

Table 1. Brief descriptions of GDV applications

Stakeholder Consultation 2

Once the information from all the project partners had been collected, FrontierSI synthesised the outputs to identify several possible software solutions that would meet the technical requirements of users. The possible software solutions and key issues relating to different options were presented to the project participants in an online workshop through a second stakeholder consultation session, Stakeholder Consultation 2 (SC2).

Following this workshop, project partners were provided a feedback form (See Appendix 2) to record the preferred solution for their organisation. The feedback provided was reviewed and the preferred solution was identified.

The following section outlines feedback on the key issues identified which may influence future development solutions.
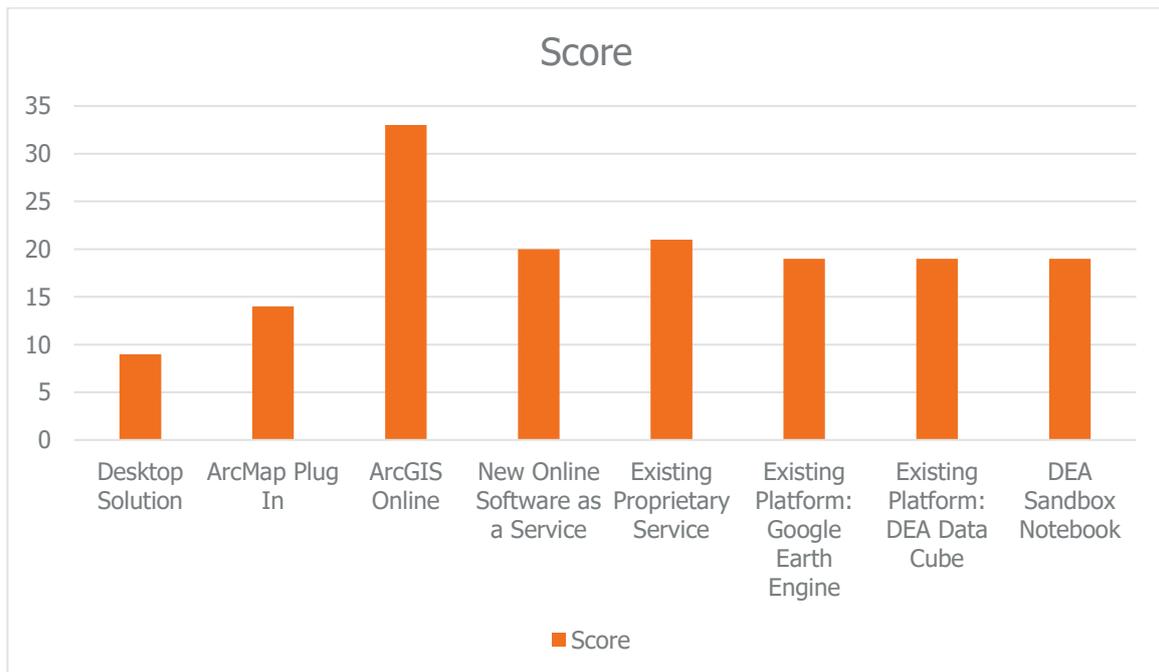
2. Feedback on Options

The stakeholders completed a feedback form (Appendix 3) responding to the options. The results are summarised below.

Q1. Online Solution

All users agreed that an online solution was preferable.

Q2. Option Preference

NOTE: Based on feedback we believe the preference for the ArcGIS Online option reflects a strong desire for integration with their existing toolsets, rather than a specific usage of the ArcGIS Online suite, and that any online solution that integrates with ArcPro would be acceptable.

Q3. Vendor-Lock In:

CONCERN ABOUT VENDOR LOCK-IN

High 29%
Low 14%
Moderate 57%
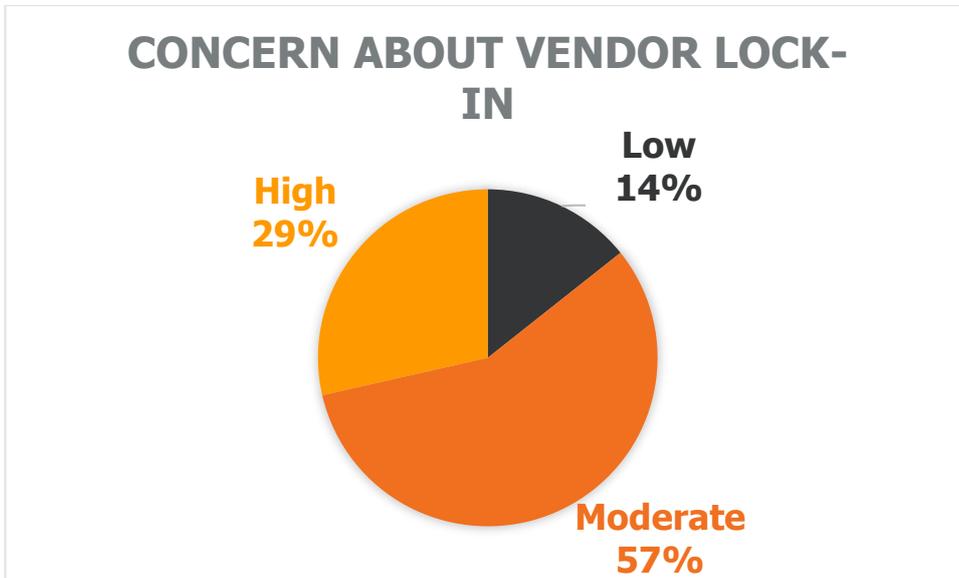
Comments:

- Need to understand licensing, business and financial model of this solution.
- The vendor's long-term interest in the project would need to be assured.
- The costs and benefits would need to be compared.
- Many respondents preferred a tool that any provider could run as a service, allowing a competitive tender process for providing the service and any further development work.
- The "Low" concern response was caveated around the solution being provided by the respondent's existing vendor.

Q4. Sandbox Version

IMPORTANCE OF A SANDBOX VERSION

High 43%
Low 14%
Moderate 43%
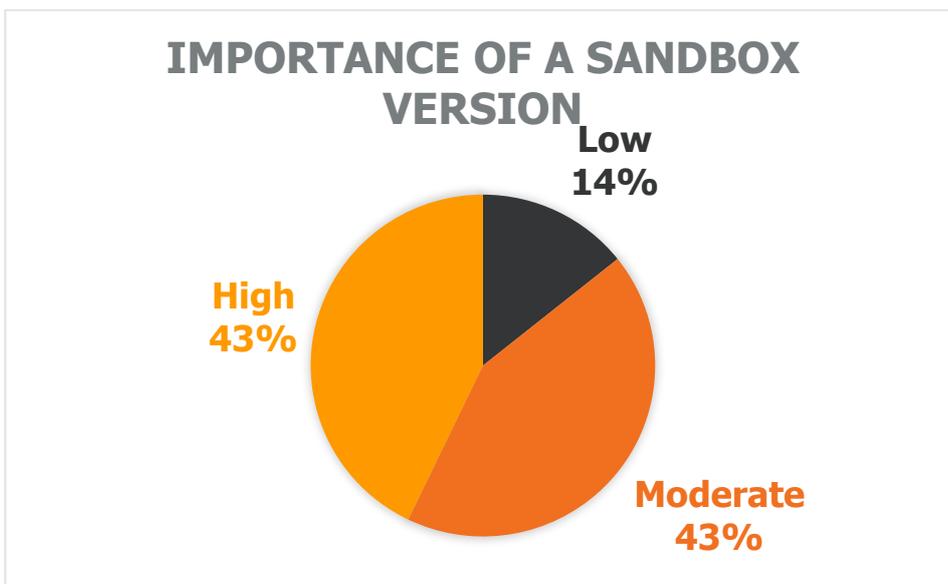
Comments from industry partners:

- Seen as "Nice to have" and "not a high priority".
- Useful to test the model.

Comments from government partners:

- Important for future development and modification of the model.
- Concerns over maintenance in the long-term.

Q5. Fractional Coverage

There was unanimous support of the local component option.

Issues were raised around licensing of commercial satellite data, and the limitations of analysis.

IMPORTANCE OF FRACTIONAL COVERAGE FEATURE

High 29%
Low 28%
Moderate 43%
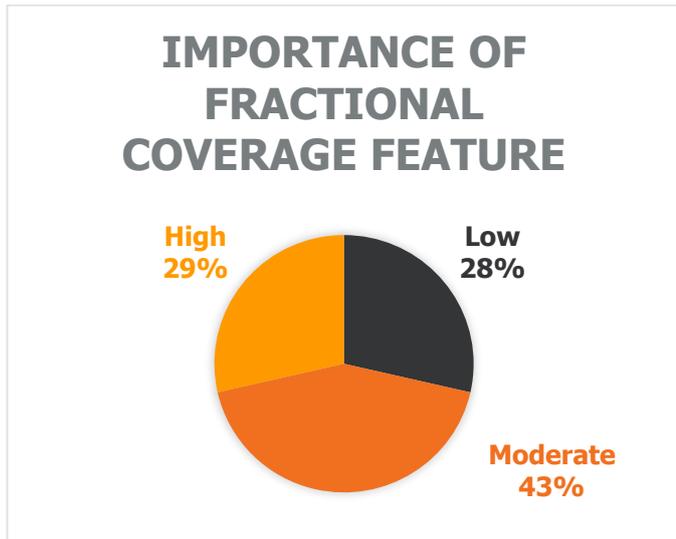
Other Comments:

All stakeholders indicated a preference for a solution that integrated with their existing toolsets. This was primarily the ArcGIS suite, with some additionally using Google Earth Engine or eMapper.

There was concern of the source of the Sentinel-2 data being used in the system with regards to the atmospheric correction applied. If the data used is different to the Geoscience Australia-corrected data that was used to train the model, this impact would need to be investigated.

Appendix 1 – Sample Mural

Sample of the Murals which were used to map workflows in which GDV are used. This methodology was used with each of the mining partners.

Process Map - what is the workflow in which GDV is used ?
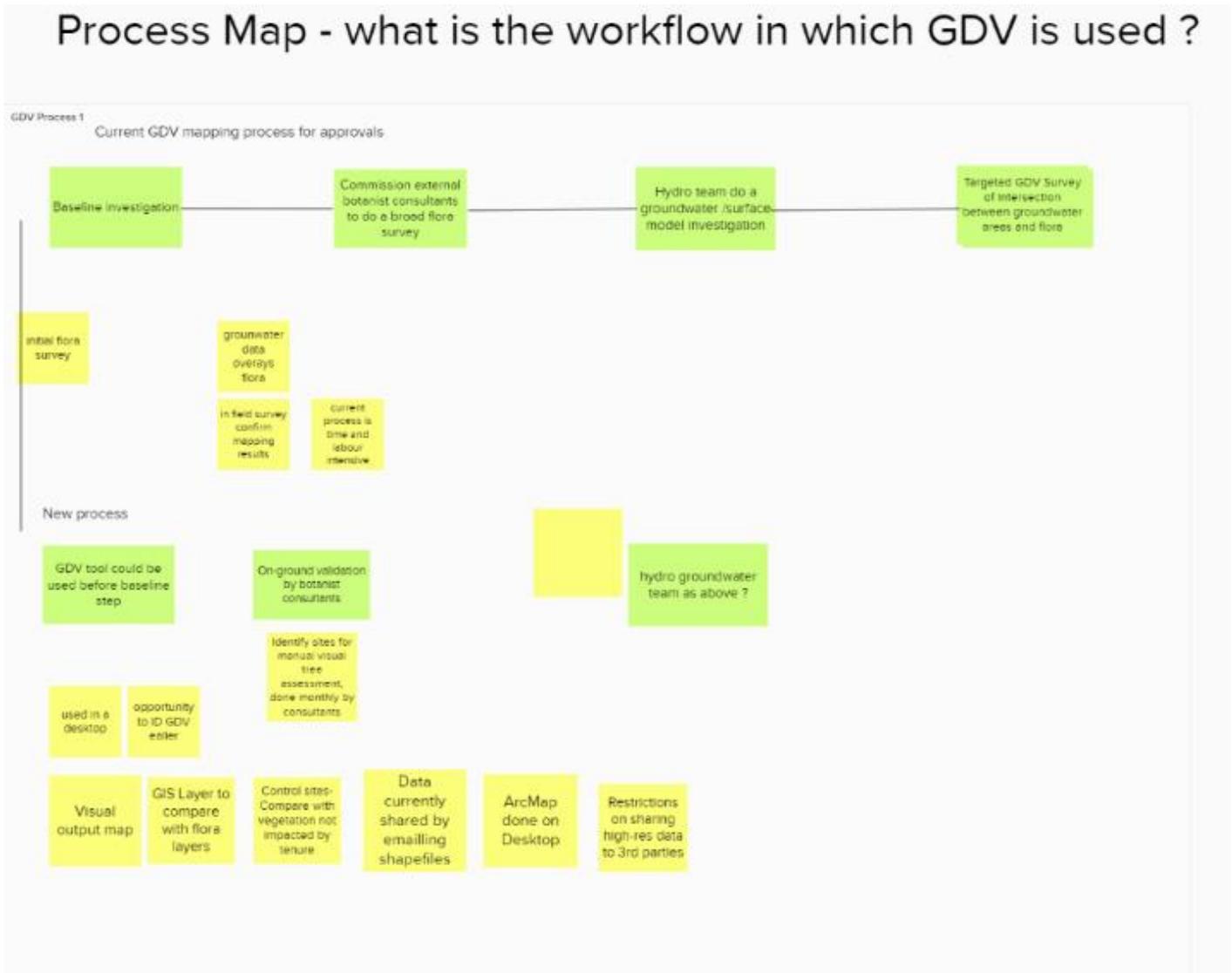
Appendix 2 - Stakeholder Feedback Form on options presented for development of the GDV Tool

Organisation Name: _____   Contact: _____

Please provide feedback on the software development options for the GDV tool presented during the Stakeholder Consultation 2 Workshop.

Your feedback will be used to confirm the design options for the GDV software tool. The feedback from project partners will be collated to confirm the best solution for the group. We are seeking one feedback form from each project partner.

1. **Based on the consultation undertaken with project partners an Online Solution was identified as the best option for the GDV Tool. Please provide your feedback on the Online Solution below.**

The Online Solution is our preferred option.

☐   Agree

☐   Disagree

If you have selected 'Disagree' please provide your reasons below.

2. **Please score the development options for each Online Solution by providing a score out of 5 in the table below (1 least favoured to 5 most favoured).**

Option ID in Presentation	Option	Score Least …………………………….…Most				
		1	2	3	4	5
A	Desktop Solution					
B	ArcMap Plug In					
C	ArcGIS Online					
D	New Online Software as a Service					
E	Existing Proprietary Service					
F	Existing Platform: Google Earth Engine					
G	Existing Platform: DEA Data Cube					
H	DEA Sandbox Notebook					

3. **Preferences for Open vs Proprietary Software (Implication of Vendor Lock-in).**

This project is paying for the development of the software, but not an ongoing service.

An Open Source solution will be free to use, however there will be no provision for ongoing support or future upgrade of the tool.

There is the possibility we can tender for a proprietary system to deliver the results. The tool could be cheaper to build and be supported in the future, however users would be locked into using that provider and may need to pay a subscription (Vendor Lock-in).

What is your level of concern about Vendor Lock-in ?

☐ High

☐ Moderate

☐ Low

How willing are you to accept Vendor Lock-in to the options below? Please rate your willingness between 1 and 10 (1 least favoured to 10 most favoured).

Vendor	Score
Cloud Agnostic	
Cloud Provider	
Existing Platform	

Please provide any comments on Vendor Lock-in below.

4. **Option for a DEA Sandbox version.**

Stage 1 of the project delivered the projects via Jupyter Notebook within the DEA Sandbox.

How important is it for you to be able to use a version of the tool which runs using the DEA Sandbox in addition to the new online tool be to your organisation?

☐ High

☐ Moderate

☐ Low

☐

Please provide any comments on whether a DEA Sandbox version of the tool would be useful for you.

```

```

5. **Fractional Coverage Options.**

The Fractional Coverage component uses a single higher resolution image (e.g. WorldView3) to effectively "sharpen" the results for the lower resolution Sentinel-2 imagery over the course of the year.

How important is Fractional Coverage to the design of the tool?

☐ High

☐ Moderate

☐ Low

What is your preferred option for Fractional Coverage?

☐ Local Component

☐ Hosted Service

Please provide any comments of Fractional Coverage below.

```

```

If you have any other comments on the options for development of the GDV Tool please provide them below.

```

```