Australian Earth Observation Data Quality: Exploring the need for a coordinated calval network to support the commercial EO Sector

Authors
Dr Jia-Urnn Lee, Dr Fang Yuan,
Dr Claire Fisk

Acknowledgements

We respectfully acknowledge the Aboriginal and Torres Strait Islander people of Australia, first custodians of the lands, air and waters that sustain the places we live, work and play. These first peoples have had a vibrant, living culture that has remained in sustainable synergy with the natural environment for tens of thousands of years, and continues to do so.

Remotely sensed data about the natural world is a powerful tool to create shared community vision accessible to many groups about sustainable ways of living in harmony with Country. We believe that through capturing high-quality remotely sensed data of the real world, we are doing our best to respect Country by accurately representing this real world in a way that ensures the integrity, trust and quality of data used for decision for the good of people and planet into the future.

We recognise that the lands of the Aboriginal and Torres Strait Islander people of Australia were never ceded, and coexist with the Commonwealth of Australia.

Contents

Summary

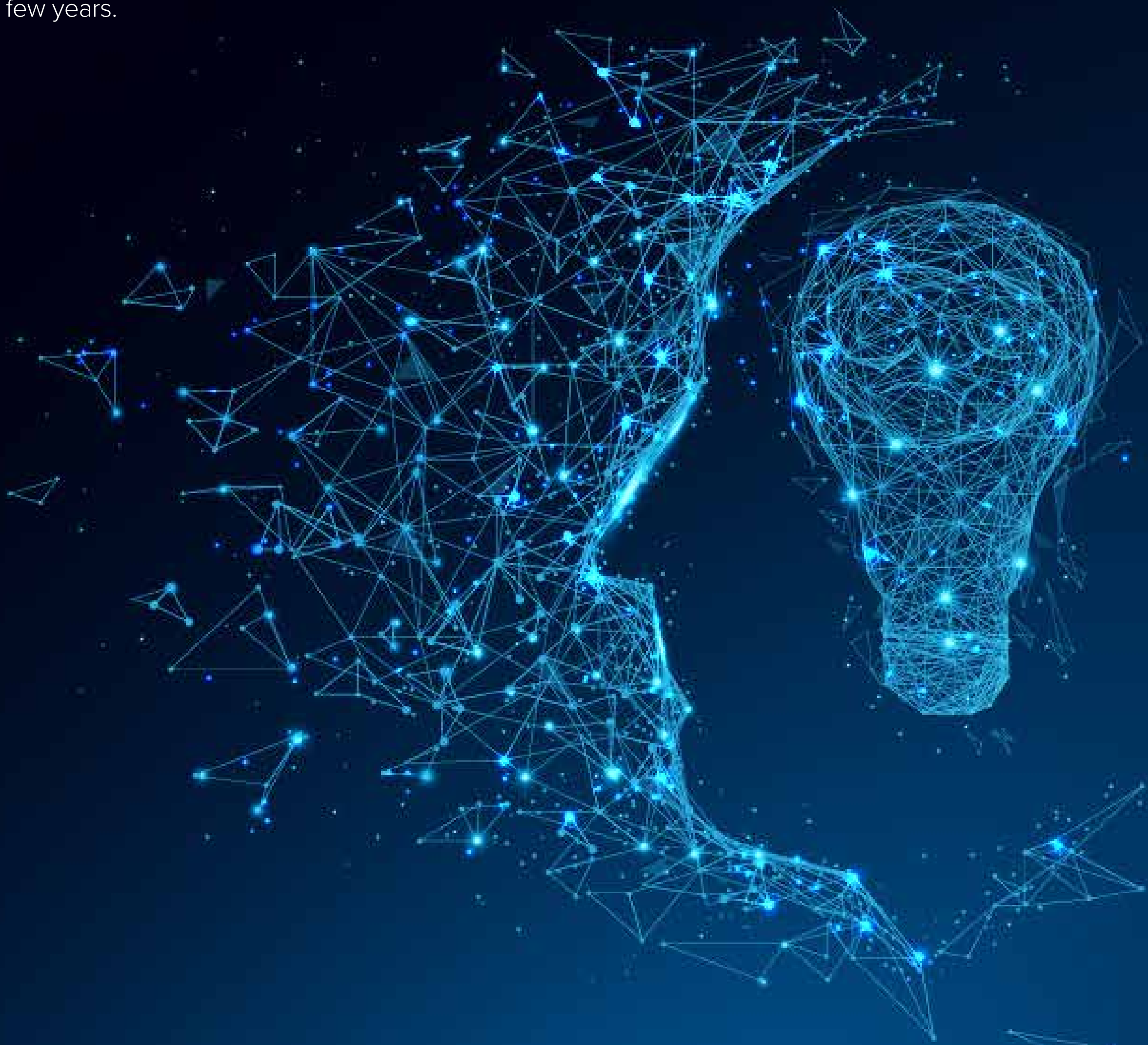
Calibration and validation (cal/val) of Earth Observation data is an essential part of the Earth Observation (EO) data management pipeline, and ensures that data is usable once it has been acquired by satellites.

Downstream cal/val is done well by the Australian EO community on data from either high-fidelity commercial satellites or gold-standard reference satellites such as Sentinel and Landsat. With the increase of smallsat development (including cubesats and microsats for the purposes of this paper) in Australia, cal/val throughout the lifecycle of mission development has become even more essential for enabling high-quality data.

Australia has a number of existing cal/val sites and supersites, providing a rich resource of world-class infrastructure and cal/val datasets. However, the use of and information from these sites remain fragmented and some remain inaccessible outside of the scientific community. A federated cal/val network would enhance the value of existing infrastructure, and significantly benefit Australia's smallsat missions, which account for 100% of all launched or planned Australian EO missions within the next few years.

This study engaged with selected Australian organisations and businesses to understand what cal/val activities and resources are required to support their needs. The following recommendations have been developed:

- Coordinate an Australian cal/val network to support the growing commercial EO community.
- Develop a one-stop-shop repository of cal/val methods, datasets and protocols to guide mission design, operations planning, and ensure the quality of EO products in supporting client / user trust.
- Leverage Australia's enterprising commercial space sector to drive up the value of existing cal/val infrastructure.
- Pilot an end-to-end cal/val case study of an Australian government-sponsored cubesat EO mission to provide valuable EO data for the science and research community, but also provide reference data for Australian commercial operations.



1 Introduction

Within the global Earth Observation (EO) community, Australia is considered to have an expert scientific and research user base of EO data.

Australia is also well-known internationally for hosting world-class infrastructure for calibration and validation of EO data. For the purposes of this report, calibration and validation (cal/val) can be explained more simply as checking and refining the accuracy and quality of EO data collected from satellites.

This work explores the need and maturity of Australia's EO sector on ensuring data quality through cal/val activities, with a focus on commercial space. It builds on previous satellite cal/val studies and the studies referenced within them:

- 2021 AusCalVal report[1] which presents the importance of quality assurance for EO data, and proposes a federated cal/val network to build broad consumer trust in EO data.
- 2022 UK-Australia Cal-Val Space Bridge[2] report which presented opportunities for strategic collaboration on cal/val services, infrastructure, and research with the UK.

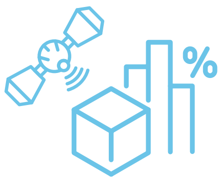
This current report has been prepared by FrontierSI for CSIRO to understand domestic commercial EO cal/val activities, with the following objectives:

- To gauge the status, maturity, and evolution of domestic commercial EO calibration and validation (cal/val) activities;
- To explore if there is a commercial need for a coordinated cal/val network; and
- To identify needs that a cal/val network can address.

Domestic stakeholders engaged for this study via interviews are classified across:

Mission developers, who plan on or are actively developing satellite missions (e.g. Spiral Blue, LatConnect60, Esper Satellite Imagery, ANU OzFuel, Smartsat Kanyini, CSIRO CyanoSense[3]). Some stakeholders in this group have long-term plans of being vertically integrated through the EO supply chain, including in developing hardware, tasking satellites, selling data, and providing insights to end user clients.

Satellite data resellers, who sell or on-sell satellite imagery, and who often also provide value-add services or insights to clients (e.g. Geospatial Intelligence, Geoimage, Arlula). Stakeholders in this group may also have significant experience benchmarking and verifying the data quality of new satellite missions, and/or providing agile data management infrastructure to meet a wide range of client needs.



Value-adders, who use satellite data or Analysis Ready Data to undertake further analysis and provide insights to end user clients (e.g. Ozius, Maitec, Regrow, Haizea Analytics). Some stakeholders in this group have a multi-faceted role in also collecting and providing validation data for mission operators and other value-adders.

[1] https://frontiersi.com.au/wp-content/uploads/2021/05/FrontierSI_AusCalVal_27052021_Final.pdf.pdf
[2] https://symbios.space/wp-content/uploads/2022/07/CalVal-Spacebridge-Final-Report-Jun-2022.pdf
[3] Formerly CyanoSat.

Snapshot: Awareness of Australian cal/val infrastructure and datasets

A snapshot of the organisations engaged as part of this current study indicates the following:

Regardless of where they sit within the EO value chain, >90% of interviewees undertake or intend to undertake some form of EO cal/val before providing the data or derived insights to their clients. The exceptions are companies whose model is to pass on satellite data un-interfered to their clients.

All interviewees who are developing missions are aware of various cal/val processes. Amongst these mission developers, 3 organisations have a general cal/val plan for their missions, and the same 3 organisations have clear intentions to include data capture over dedicated global calibration sites as part of their cal/val activities.

Almost two thirds of all organisations interviewed were unaware of the existing cal/val infrastructure and datasets in Australia. When referred to the 2021 AusCalVal report, some interviewees engaged for this study have no knowledge of the types of data provided by the sites, nor how to access cal/val site datasets.

The snapshot indicates that the existence and value of Australian cal/val infrastructure is not well-known within the domestic commercial EO sector. Meanwhile, the responses gathered from the interviews demonstrate that more and better support cal/val is needed.

Interview set-up

Interviews were undertaken with the above-mentioned organisations between September 2023 and January 2024. De-identified responses[4] have been aggregated into the following questions (A-M) and are detailed in the Appendix:

A. What type of sensor data do/will you provide, sell, or use?

B. What imagery / data level do you provide to your clients (i.e. Level 1, 2, 3, Analysis Read Data (ARD))?

C. What cal/val activities do you do with the data you receive (from satellite providers) or your own mission? Or does the provider undertake all of the cal/val and you trust in the data?

D. Will you capture over Australian calibration or validation sites? Do you know of any?

E. What data quality attributes do you want or request from your satellite providers? Have you encountered issues with data from your satellite providers?

F. What are your criteria to sign up satellite data providers or be an early adopter of their services?

G. What data quality or cal/val attributes do you aim to achieve? What data quality or cal/val attributes do your clients request?

H. Are your clients aware the data could come from different operators?

I. What data quality issues have you encountered from providers, and any impacts on your business?

J. Do / will you validate your products or models? Do you undertake validation on derived products you use as input to your own products?

K. Would a federated cal/val infrastructure / service be of value to your business and how?

L. Would you be willing to pay for outsourced or commercial cal/val services?

M. Accessing Australian cal/val infrastructure and datasets.

In the sections below, we link our analysis back to the relevant question ID and response numbers in parentheses.

[4] Some of the responses to questions are referenced as evidence throughout this report. Evidence (A2) for example refers to Question A in the Appendix, response 2.

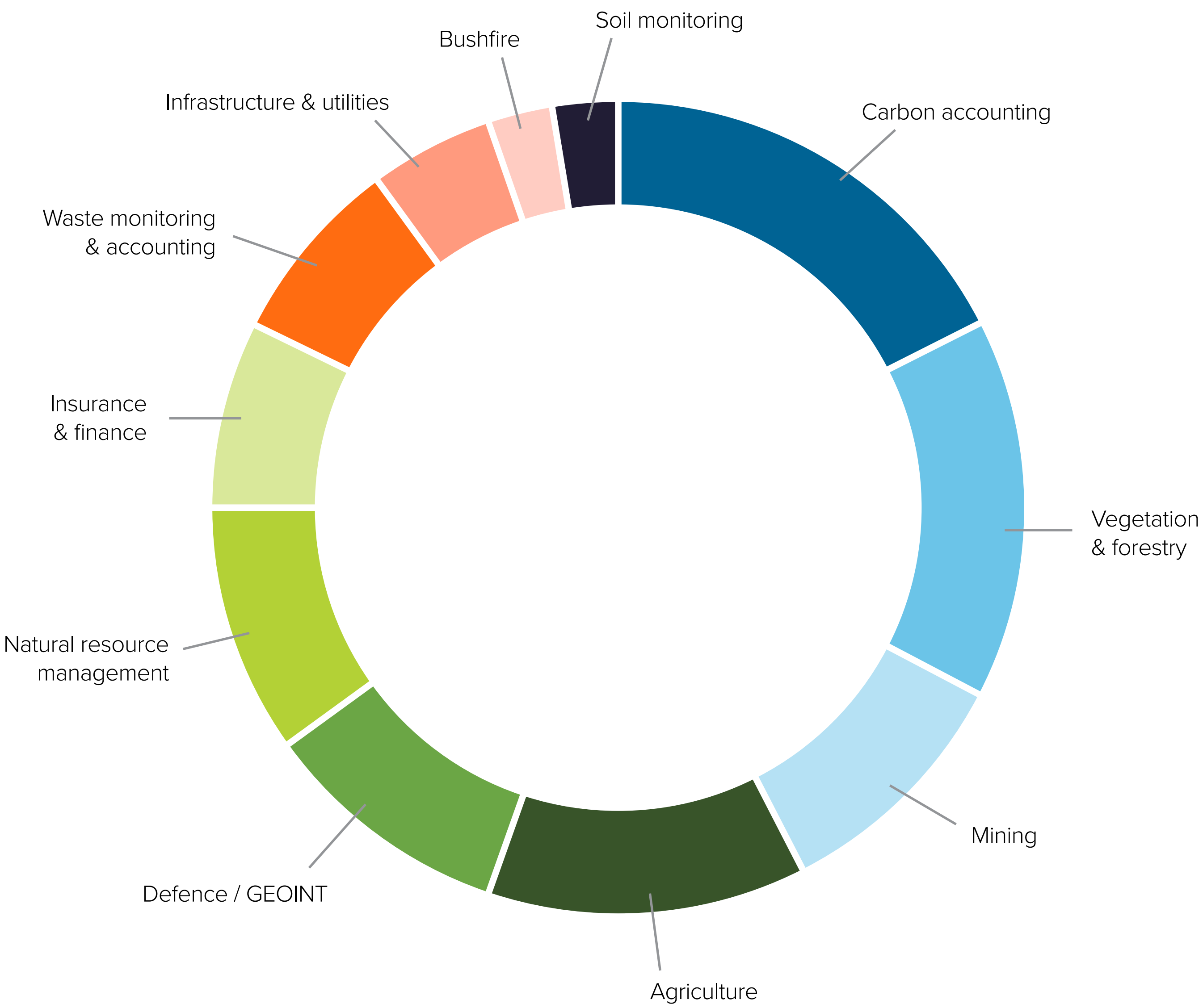
2 Evolution of the Australian EO ecosystem

Applications and sectors

For organisations interviewed for this study, the most common target markets or applications are carbon accounting (including greenhouse gas emissions), vegetation and forestry, mining, agriculture, geospatial intelligence for Defence (GEOINT), and natural resource management (Figure 1). While insurance and finance are categorised here as a client sector, realistically they require environmental intelligence auditing, compliance and governance uses for their own clients in many of the other applications.

Responses from the interviews allowed categorisation of EO data users in Australia into two broad groups. First is a small but experienced scientific user community mainly within government and research institutions (B2, G17). Second is an increasing number of industrial sector clients seeking insights derived from satellite imagery, with some end users not necessarily having or needing EO technical expertise (G16, G17, G19).

Figure 1. Common market sectors and applications of Australian EO businesses.



Vertically-integrated companies entering the market

Since around 2018/2019, coinciding with the establishment of the Australian Space Agency, Australia has seen multiple new startups enter the EO market. The startups interviewed for this report specialise in developing space subsystems such as hyperspectral sensors and edge processors. To bolster their business models, several of the startups have diversified from purely developing hardware, to also re-selling satellite imagery and providing insights for bespoke applications (A5), with initial lower barriers of entry. Some of the startups have longer-term plans to launch and operate their own cubesat or microsat constellations (generally sub-50 kg), to collect their own data to better serve their clients.

Diversification from exquisite satellites to cubesat constellation data

Data resellers, particularly those with established government and Defence clients, have a target market demanding high resolution imagery from "exquisite satellites"[5] such as Pleiades, WorldView, and SPOT. In terms of data quality, calibration and validation, interviewees found that high resolution data at the sub-metre scale requires minimal additional calibration after the data has been provided by the satellite operators (C1, C2), and is generally "easy to co-register" to and against other data sources (I1).

However, there is an increase of data from smaller satellites entering the EO market. Combined with evolving client needs for aesthetically-pleasing satellite data, that can be collected at high frequency and in customised areas of interest, data providers expect that they are approaching the cusp of needing to diversify to offer a mix of exquisite and cubesat / microsat constellation data (A1, G20).

> ***"We are always on the lookout for new satellite providers... and diversifying from big primes. There would be a large impact if anything happened to prime satellites."***

Some interviewees articulated that the downside of cubesat / microsat acquired data is clearly in poor data quality (K1, K2, K7). Data quality in this case encompasses spatial resolution, geolocation accuracy, image artifacts, correctly calibrated sensors and sensor data, and consistent data quality throughout a constellation with high turnover of satellites. The space, power and weight limitations of cubesats and microsats generally translate to acquisition of lower resolution data (5-10 m increasing to 20-30+ m GSD) as well as lack of onboard calibration (C11, C14). Additionally, in Australia, cubesats and microsats are likely to be launched by startups and smaller commercial companies, who do not necessarily have the established capability, expertise and resources to undertake robust pre-launch, onboard, or vicarious calibration to the same extent that multinational primes and international space agencies can.

[5] "Exquisite" refers to larger, more sophisticated, expensive, generally one-off satellites rather a constellation, often developed by well-resourced primes and/or international space agencies.

3 Cal/Val practices in the Australian EO sector

Cal/val can occur both upstream or downstream. Upstream calibration is undertaken during development of the satellite payloads (pre-launch calibration), and during operation of the mission (onboard and vicarious calibration). Downstream calibration and validation occur as part of processing of the acquired EO data. Interviewees were asked about the cal/val activities they undertake when they receive data, whether from satellite providers or for their own missions (Question C in Appendix).

Depending on where they sit in the EO value chain, businesses undertake varying levels of data processing and provide different products to clients. Some companies act essentially as conduits for commercial data and do not have visibility into the satellite operators' processing pipeline. The companies obtain post-processed and calibrated commercial data from satellite operators, which is passed on un-interfered to customers. These companies may not be able to correct or undertake additional processing on the satellite data, as the data is essentially owned by the customer they are selling to (C20, G10).

In contrast, other data providers or solutions providers undertake common checks to provide native data, post-processed data, specific spectral bands, cloud-free data, or synthetic data, depending on client requests. Mission operators typically play a large role in cal/val both pre- and post-launch. Typical cal/val practices conducted are outlined below.

3.1 Upstream

Pre-launch calibration. Mission developers consider pre-launch calibration of sensors as vital for minimising post-launch sensor calibration issues, as well as informing the design of in-orbit calibration activities (C11, C15, C22). The extent to which they can do pre-launch calibration themselves depends on infrastructure and expertise available. On the fully self-sufficient end, a couple of research organisations can undertake detector characterisation, testing and calibration in-house, using state-of-the-art qualification infrastructure developed for astronomy research (C15). At the other end of the spectrum, other mission developers procuring commercial-off-the-shelf sensors from overseas may leave it to the manufacturers to calibrate the sensors (C4, C5).

Onboard calibration. Onboard calibration is generally considered a stretch goal for mission developers. Restrictions on spacecraft space, weight and power would likely preclude onboard calibration systems for their satellites (C11). One organisation mentioned configurable onboard sensor calibration as an ambition, where spectral and radiometric calibration could be updated on the fly as the satellite passed over a calibration site (C21). Onboard data processing is different to onboard sensor calibration, but is included in this description. At least two of the organisations undertake onboard data processing and can downlink relevant analytics and raw data within constraints of data downlink bandwidths (B11, C10).

Capturing over Australian cal/val sites (vicarious calibration). Interviewees were asked if their missions would capture images for calibration as the satellites passed over Australian cal/val sites, or request their satellite operators to capture over Australian cal/val sites (Question D). Research (scientific) missions were able to pinpoint specific Australian sites as part of their cal/val plan (D1, D5), including the Pinnacles (Western Australia), TERN sites including the Cumberland Plain and Tumbarumba supersites in Queensland, and the National Arboretum in Canberra. One organisation explained that their off-the-shelf sensor may likely capture over calibration sites overseas (for example the Libya-4 pseudo invariant calibration site), to assess the performance of the sensors against previous missions that have captured over the same calibration site using a similar sensor build and specifications (C6).

Some companies encourage their partner satellite operators to acquire imagery over calibration sites regularly so that they can assess errors to enable appropriate corrections (during processing, calibration and validation). Experiences vary between the companies and their partner operators. One company may find that it is the operator's decision to capture over cal/val sites (some operators may not) (C17). Another company may have a long-term partnership with a satellite operator who understands their needs and requirements (C3).

While Australia has numerous "supersites" for vicarious calibration and validation[6], this information is not readily known or accessible by companies outside the scientific community or even within those with close ties to the scientific community (C22, D2, D3, D4, L9). Overseas launch providers play a role here in recommending calibration sites to their Australian rideshare customers, or refer them to a USGS Test Sites Catalog[7] (D7). Unfortunately, Australian sites may not be well-represented in such overseas resources. Some respondents are clear on the need for Australian datasets calibrated and validated to Australian conditions (L10), however this quality parameter may not necessarily be appreciated by other companies or those with international markets.

3.2 Downstream

Geolocation and geometric accuracy. Interviewees noted they spend significant effort in ensuring geometric accuracy and consistency of satellite EO data, EO products and EO derived products (C7, C8, J2, J3, J7). Effort is spent on band alignment, distortion correction and correction of other geometric-related issues, as well as registering timeseries data to ensure locational consistency of imagery. Some value-adders have quoted a general requirement of 5-10 m, though this is less of a concern when target features are larger. Often, orthorectified images are geolocated against a reference standard.

Comparison against a known/trusted reference. Most organisations cited calibration and validation of images, datasets, band aggregation, and other baselines against Sentinel and Landsat as a common process (A6, C7, C9, C17, G14, K21) to assess radiometric and geometric accuracies. This includes: i) assessing for drift against a known standard; ii) referencing an orthorectified image to identify any offsets, warping, or orthorectification issues. One organisation may also look at self-cross calibration where they intend on checking for synchroneity of images from their in-orbit sensor systems, as an element of cross-calibration within their future satellite constellation (C14).

Validation of products and/or models. A number of interviewees, especially those delivering ARD and EO insights, perform validation of datasets and derived products as part of their own product or algorithm development. Examples include validation of surface reflectance products (J2, J10) and derived biophysical parameters (J9, K17, K18). Sometimes, additional corrections such as Bi-Directional Reflectance Distribution Function is applied to improve data consistency (J4). This type of validation is critical to ensure quality of derived insights to be provided to their clients.

For these organisations, applying their in-development algorithms and models on datasets is a way to validate their own models, while applying established models on new datasets is a way to assess the output accuracy and quality of input data (I3, I12).

3.3 Customer expectations and quality targets

Consistency over time. Interviewees indicated that consistency over time for geometric, geolocational and radiometric accuracy, and comparability to global datasets such as Sentinel-2 is often desired. Geometric accuracy is the most often mentioned quality requirement. There is a universal need by all types of end user segments and applications to have accurately-located objects or features on the ground.

Geolocational accuracy requirements can vary depending on application, and often depend on the size of the target of interest. Clients in the geospatial intelligence sector may want better than 5-10 m accuracy (G7, G9). For broader or landscape-scale environmental assessments, these requirements can be relaxed to 25-50 m, where some clients prefer a dataset to "look good" rather than require absolute geolocational accuracy (G3). Absolute geolocational accuracy (up to 25-50 m) may not be an issue, so long as all images are out to the same amount of error, and the data can be stacked on top of one another to derive time series information for change detection (G1, G7, J3). This requirement supports monitoring applications and machine learning-based algorithms to be applied across sensors.

Many groups have established methods that they apply to co-register images to ground control points or baseline reference images (for monitoring over time). This does add to the processing and quality control efforts required and tends to be a bigger burden for smaller businesses.

How "good" satellite data is depends on what it is used for. One organisation who provides Defence intelligence uses very high resolution commercial data, remarking that Sentinel data for one-off projects is the "poorest resolution" data they use (I1). Other companies who work more in environmental monitoring deem Sentinel as fit for their purpose. Several interviewees explained that open data is suitable for landscape-scale analysis and and product upscaling (I8, I9), with gaps filled either with commercial satellite imagery or drone data.

[6] TERN Ecosystem Processes, www.tern.org.au/tern-observatory/tern-ecosystem-processes
[7] USGS Test Sites Catalog, https://calval.cr.usgs.gov/apps/test_sites_catalog

Sector-imposed quality criteria. Some clients have strict quality criteria imposed on them as a function of their application/sector, and these quality criteria are passed onto geospatial companies / data providers. For example, within the carbon credits sector, client outputs or algorithms are certified by energy regulators, and hence supplied satellite imagery need to meet certain regulator requirements for their clients (G2). There is thus an untapped and unexplored opportunity to coordinate Australian cal/val efforts to help domestic EO data providers meet Australian industrial or regulator standards, where EO data is a critical input.

3.4 Impacts of poorly calibrated or unvalidated data

Interviewees were asked whether they have encountered poor data quality issues from satellite operators, or whether they themselves had encountered issues from their own cal/val processes. (Question I in Appendix). For data from satellite operators, experiences appeared to vary depending on whether the companies had access to larger, more sophisticated "exquisite satellites" where robust cal/val is provided. One such company indicated high resolution (spatial) data at the sub-metre scale makes co-registering easy (I1) and requires minimal additional calibration after the data has been provided by the satellite operators (C1, C2).

In contrast, some R&D teams in both research organisations and businesses expend significant effort to incorporate datasets from new sensors into existing products, but often find that commercial data is not suitably calibrated to their needs. One company cited an example of being asked to provide insights from high resolution commercial smallsat data, with the processed data and surface reflectance data showing more registration, jitter and signal-to-noise issues compared to Sentinel-2 data (I4).

That said, Sentinel data has also had its critics amongst the environmental intelligence community. One company which provides a tree canopy product said: "There are geometric issues in the earlier Sentinel era, meaning we cannot use time series information for tree canopy because the registration wasn't accurate enough. The lack of previous geometric accuracy limits use of the data, limiting a number of projects for us and across the community."

For those developing their own cal/val pipeline, especially for startups, the learning curve has been steep. Experience has informed major improvements to data processing and data management over the last 2-3 years (I2, I10).

Several interviewees disclosed that they are frequently approached by both global and domestic satellite operators to be early adopters of data from new sensors and new technology. The interviewees play a significant role in quality checking and assessing if the satellite data is fit for market (F6, F7). Data quality can vary, with poor quality data being spatially inaccurate and needing significant calibration to ground controls, poor band alignment, and a lot of noise due to compression processes (F1, F2, F3, F4, F8). The interviewees commented that on-selling poor quality data can lead to severe consequences in terms of environmental intelligence and decision-making, but also to company reputation (K2).

The other issue raised by a few interviewees is a lack of transparency from satellite operators on their processing pipelines and the resulting data quality, (I11, I12), hindering further validation and use of such data in applications.

3.5 Trend towards insights / Analysis Ready Data only

Interviewees were asked about whether their clients were aware that EO products, insights and results could be based on different satellite data sources. Responses indicate that more recently, there is an increasing trend of clients requesting final insights or analysis ready data, rather than including the raw imagery, processed data, or metadata (B1, H2, H4). Clients essentially "outsource" assurance of data quality to the data providers, either because a) they place increasing confidence in the solutions provided by trusted geospatial companies; or b) the clients are indifferent.

For the latter, some clients *"don't care about calibration*" (G16). This may not be an issue if the data being resold is from exquisite satellites, where robust cal/val is undertaken as part of the satellite data acquisition and tasking service. Client indifference should be of concern where:

- Data comes from cubesats or smallsats where rigorous cal/val is not a priority;
- Clients do not necessarily understand the limitations of EO data (G17);
- Clients have limited scientific knowledge about biophysical / environmental processes.

A company cited an example of a customer providing machine-learning based solutions for agriculture. In this case the non-functionality of the customer's algorithm was due to the customer's lack of knowledge about scientific processes (e.g. lack of knowledge of multispectral vegetation indices) rather than data quality itself as an issue (G19).

4 Value of a federated cal/val network

Interviewees were asked whether federated and coordinated cal/val infrastructure, datasets or services would be of value to their business or missions, and what this might entail. The following themes emerged:

- While larger gold-standard satellites would benefit from a more coordinated Australian cal/val effort, federated cal/val infrastructure would significantly benefit Australian smallsat (including microsat and cubesat) operators more (F6, K1-K7).
- Additional cal/val datasets would provide mission developers with confidence for upstream sensor cal/val (D7), as well as provide a measure of quality for clients to to verify their downstream data products against (K8, K11).
- Cal/val datasets need to be accessible and supported in a manner that is scalable and interoperable (K9, K10), otherwise the benefits of more cal/val data cannot be realised.
- Access to well-collected, reliable field data for validation of sensors, data algorithms and new techniques (K16, K17).
- Frequently published data and updates of biophysical data from cal/val sites, in a coordinated and consistent manner across the Australian network of data, and building upon already invested-upon infrastructure would be valuable (K18).
- Australian cal/val datasets should be able to integrate and be consistent with global datasets. Interoperability between datasets and protocols, such that data can be normalised between datasets would also be useful (I5, I6, K19, K20).
- Cal/val protocols need to be updated or developed. Field validation protocols need to be updated for ground validation through a national effort. A set of protocols also need to be developed or aggregated for pre-launch sensor calibration and spaceborne data validation. It is unclear to industry whether these exist or how to gain access to these protocols (K22).

5 Discussion and Recommendations

5.1 Upcoming missions and strategic international partnerships

It is worth describing here: i) the imminent launch of a couple of flagship Australian EO missions; and ii) Australia's involvement in the USGS Landsat Next Mission, and how these programs could help elevate an Australian cal/val network to benefit smaller, commercial EO missions.

Government-funded demonstrators

In mid 2023, CSIRO's CyanoSense-1 was launched as a technology demonstrator for the AquaWatch mission. The primary mission focus was to produce engineering data to confirm performance of all payload sub-systems. The next demonstrator, CyanoSense-2, is currently in development and is anticipated to generate EO data for science and testing purposes. At the time of this survey, CSIRO had expressed intention to undertake cal/val, but at the time of the interviews (late 2023-early 2024) have no firm plans for post-launch cal/val.

The South Australian Kanyini mission is due to be launched in July 2024. It is also not intended as an operational satellite but to serve as a research platform. Cal/val plans have progressed since the time of the interviews, though updates have not been captured in this report or the Appendix.

Demonstrators help to de-risk technical and satellite operation issues. However, there is an opportunity to use them to de-risk data issues as well. Considering that the value of EO is in downstream applications (with over 90% concentrated in agriculture, utilities, public services, mining, transport and others[8]), investment and effort should be proportionately allocated to the data value chain, which is enabled by cal/val.

CyanoSense and Kanyini, as government funded missions, should lead by example and demonstrate the value of rigorous cal/val within their missions, both pre- and post-launch. This would help set the precedence of Australian missions utilising Australian cal/val infrastructure. It would also provide valuable EO data for the science and research community, and reference data for Australian commercial operations.

International partnerships

The Australian partnership on the NASA and United State Geological Survey Landsat Next[9] satellite program was announced after the period for this survey. Australia's involvement in the program will be led by Geoscience Australia and will include enhancing ground station capabilities, but also in Australia's areas of competitive advantage, i.e. downstream EO data processing and analytics. While not explicitly highlighted, it is assumed the latter would include rigorous cal/val throughout the sensor and data value chain. Landsat Next will undoubtedly become a key data reference mission for Australian space and geospatial companies in undertaking cal/val. Critically, the program will help strengthen international collaboration and continue to secure Australian access to data from international missions.

There is thus an opportunity to anchor a cal/val network to national uplift programs (CyanoSense, Kanyini) to further demonstrate the importance of accurate and quality data in Australian downstream EO excellence. Signalling certainty through a federated program of work will have a multiplier effect by generating awareness and confidence of cal/val datasets within the EO community.

[8] Amplifying the Global Value of Earth Observation, www3.weforum.org/docs/WEF_Amplifying_the_Global_Value_of_Earth_Observation_2024.pdf
[9] Australia to join US satellite program in Landsat 2030 International Partnerships Initiative, www.minister.industry.gov.au/ministers/king/media-releases/australia-join-us-satellite-program-landsat-2030-international-partnerships-initiative

5.2 Opportunities and recommendations

A high-level SWOT analysis in Table 1 presents the strengths, weaknesses, opportunities and threats for an Australian cal/val network and the current cal/val ecosystem. This report concludes with the following opportunities and recommendations to address the factors highlighted in the SWOT:

Coordinate an Australian cal/val network. Outside of the research and government community, a nationally-coordinated front door for cal/val would make the infrastructure and data more easily accessible to Australian space and geospatial companies. There needs to be urgency in coordinating and communicating the existence and value of cal/val infrastructure to Australian industry, who are designing cal/val plans for their missions. Lack of promotion and accessibility to the Australian cal/val network is not only wasted value of Australian research infrastructure, but a lost opportunity to support commercial EO.

Develop a repository of cal/val methods, datasets and protocols. As an increasing number of space companies become vertically integrated, a coordinated cal/val data hub could essentially act as a one-stop-shop for information. A compendium of standardised methods, protocols and tools could guide mission design, operations planning, and quality assurance of EO products in supporting client / user trust.

Leverage Australia's enterprising commercial space sector to drive up the value of existing cal/val infrastructure. Federating and centralising cal/val assets and resources may open up industry-led opportunities such as data marketplace services to address industry cal/val problems and capability needs.

Pilot an end-to-end cal/val case study of an Australian cubesat EO mission, and evaluate potential policy, technology, and interoperability impacts and benefits of a federated cal/val network. The cal/val value chain should capture both upstream and downstream cal/val, including cal/val needs of data providers and value-adders. CyanoSense and Kanyini, as government funded missions, should lead by example and demonstrate the value of rigorous cal/val within their missions, both pre- and post-launch. This would help set the precedence of Australian missions utilising Australian cal/val research infrastructure. It would also provide valuable EO data for the science and research community, and reference data for Australian commercial operations.

Table 1. Identified Strengths, Weaknesses, Opportunities & Threats.

<table>
<tr><th>Opportunities</th><th>Threats</th><th>Strengths</th><th>Weaknesses</th></tr>
<tr><td>Australian investment in domestic missions such as AquaWatch and Kanyini.</td><td>There is significant investment in EO satellite technology development, but comparatively little in supporting cal/val and deriving quality data from these missions.</td><td>Australian scientists and researchers already undertake cal/val R&D, well regarded by the international research community.</td><td>Lack of coordinated cal/val effort across government and industry results in inefficiency in leveraging cal/val investment and expertise.</td></tr>
<tr><td>Australian EO is regaining momentum through strategic partnership in the USGS LandsatNext program.</td><td>Political and policy instability (change in government and their priorities) impacts investment in infrastructure that supports Australia’s EO sector.</td><td>There are established cal/val sites across Australia.</td><td>While Australia has numerous sites for vicarious calibration, this information is not readily known or accessible by companies outside the scientific community. Companies are unaware of Australian cal/val sites and not including these sites in their cal/val plans.</td></tr>
<tr><td>Broad international support for an Australian nationally coordinated cal/val network to support global initiatives, ensuring greater coverage of cal/val data.</td><td>Lack of standardised cal/val processes. Mission developers are feeling their way through based on an undefined idea of what constitutes "good enough" cal/val.</td><td>Customers of established geospatial companies trust them to provide insights derived from quality data.</td><td rowspan="3">Startups are unable to undertake the scale of cal/val that large companies and international agencies can do due to their scale, capacity, access to infrastructure, and in some cases lack of calval expertise.</td></tr>
<tr><td rowspan="2">Other industrial applications are approaching the cusp of defining ESG standards and credentials. There are sector-imposed quality criteria, such as energy regulations or mine site operation standards, that EO data products support.</td><td>Poor quality data may negatively impact on the analysis and decisions based on its use, limits wider adoption.</td><td rowspan="2">There is growing awareness within the commercial EO sector in Australia of the need for cal/val.</td></tr>
<tr><td>Instead of leveraging domestic cal/val expertise / capability, Australian commercial operators rely on overseas providers for solutions and advice.</td></tr>
</table>

Appendix: Interviewee Responses

A: What type of sensor data do/will you provide, sell, or use?

1. Using / selling optical (VIS to NIR), SWIR, thermal infrared, mostly commercial. Also hyperspectral, which are new satellites from VNIR to SWIR, mostly commercial. We are always looking for new satellite providers. There has been an explosion in SAR and hyperspectral, and have signed up with some start ups.
2. Selling optical, SAR, and signal. Commercial.
3. Using optical (Sentinel, Landsat, MODIS, Himawari), sometimes clients provide commercial datasets.
4. Hyperspectral VIS-NIR, but it is a research / educational platform and not considered an operational capability. Not selling. Will be delivered to the end user [sponsor of mission], and then a data tasking form will be provided for those who to access the data.
5. Developing SWIR from our own satellites in the longer-term. We wanted to build a fleet of satellites but raising capital was impacted by COVID. So we're developing focused solutions with a capacity lease from our partner satellite operator that provides multispectral VIS-NIR.
6. Using multispectral, SAR, lidar. Baseline is Sentinel 1 and 2, Landsat for timeseries, MODIS, GEDI.
7. Using / selling multispectral VIS-NIR. One of our satellite providers does capture SWIR data but we don't sell it because it's hard to calibrate. Future lidar.
8. Using optical (MODIS, VIIRS, Sentinel-2/3). Research / experimental hyperspectral.
9. Using optical and SAR. SAR is free and from commercial.
10. Developing hyperspectral VIS-NIR pathfinder and then SWIR constellation.
11. Using multispectral SWIR. Developing hyperspectral SWIR.
12. Selling multispectral VIS-NIR, possible hyperspectral and Lidar in future.
13. Hyperspectral VIS NIR.

B: What imagery / data level do you provide to clients (i.e. Level 1, 2, 3, ARD)?

1. Can be all of the above. Depending on client needs. Some clients don't want imagery; we may just provide results as a kml file with analytics and results.
2. Can deliver raw data from vendor with rough geometric accuracy. Our clients are usually researchers who want the raw data; they will calibrate using their own data.
3. The instrument will be able to download Level 0 data and can process onboard from Level 0 to Level 1C, which is Top-of-Atmosphere georeferenced but not orthorectified. Some processing will be done onboard and some on the ground. The manufacturer of the sensor provides utilities (processing code) to deliver up to Level 1C. Ideally we can get to Level 2 Top or Atmosphere (georeferenced and orthorectified) before we supply it to others for research but unclear yet on process.
4. We can task the satellite directly out of the office, get raw data and process it into a form our customers can use. The processing chain has been developed in-house; been a learning curve. Depends on client need. We start with a baseline product, then build out bespoke processing chains if required for clients. Baseline is Level 1B, top of atmosphere, georeferenced, radiometrically-corrected, stopping at orthorectification. Our clients can do their own orthorectification. We also produce Level 3 data, such as vegetation indices, NDVI, and provide analysis as a fully-formed product for some clients.
5. Level 1 (direct from the satellite provider) and Level 3 (corrected and adjusted to make it look nice)
6. Process raw data to application data.
7. Level 2A. Data will have undergone rectification and processing. We will not provide analysis or insights.
8. Level 0 to Level 2 for the demonstrator product. For a future constellation, we expect to provide Analysis Ready Data to users as well as the flammability product.

9. Typically, Level 2 product (L2-B specifically bottom of atmosphere) – can provide Top of Atmosphere and other specifications - user would specify at purchase. As a baseline, we try to get all data to TOA L2 product in their system and then can do BOA as needed. Resellers or customers can access our system and do the processing. Cloud masking is challenging to create with only RGB+NIR, don't have the short-wave info that would make this simpler. Analysis Ready Data is a target, not where we are at but working towards it.

10. Conduct prelaunch calibration - spectral and radiometric calibration, then aim to produce Top of Atmosphere radiance data. We want to correct for the instrument response and ultimately pass that data on to those who can work with it and have the knowledge to do top of atmosphere corrections.

11. Onboard processing is designed to be modular because we have a limited data budget, i.e megabytes a day but we're capturing gigabytes a minute, so had to think about how to squeeze the most data out of that. We have algorithms that take data, compresses it as much as possible and then we downlink it to make sure the image is in focus. Then build upon that data stream. At the highest level of product, it's that TOA Radiance spectral data cube – whether they will do that on on-board or on the ground is still something they are considering. We won't go past that processing level on board.

C: What cal/val activities do you do with the data you receive (from satellite providers) or your own mission? Or does the provider undertake all of the cal/val and you trust in the data?

1. Depending on client needs, we do minimal preprocessing to get spatial accuracy, datum projections, atmospheric corrections. Geometrically – system orthorectification (terrain and Nadir angle are considered) or orthorectified. For quality assurance, we make sure the bands are aligned, though we rarely find issues with high resolution satellites by prime. We don't have to test the spectral accuracy too much. Main in-house correction is atmospheric correction derived from the calibration data from the provider. With smaller operators however that cal data isn't readily available. So a federated calval program would be a good opportunity to have small satellite operators get on-board.

2. The provider gives very good registration and calibration, and we also have tools we can use to do the cal/val. Sometimes the state government clients provide some of the satellite data they want us to use. We have access to good ground truthing data though we can't elaborate on these.

3. We have long term relationships with our satellite provider partners and they have a good understanding of our requirements.

4. The sensor manufacturer will provide an initial in-orbit calibration. Downstream of that we know we need a calval plan but we haven't explored that further.

5. We don't know what the sensor manufacturer's calibration processes are. These instruments come pre-calibrated and some are already in-orbit, so we assume they're taking pictures of sites they've previously flown over. We're unable to do any testing with the sensor due to lack of appropriate facilities.

6. Not sure what the post-launch calval will be. The sensor manufacturer will handle it, referencing data from their other missions (for example the Libya-4 pseudo invariant calibration site). We do intend to collect aerial hyperspectral imagery for calval.

7. We will compare our data to Landsat and Sentinel over time. Bands to be registered accurately and ortho-referenced. For band aggregation, we hope to come out to a similar accuracy to Landsat / Sentinel.

8. For the satellite we're tasking, the overall processing is done by the satellite manufacturer. In terms of calibration against specific sites, it has been a learning curve to get to the processed data. A lot of the calibration data in Australia is not just for spectral or radiometric correction, but for geometric referencing as well. We use sites in South Africa for the processing chain development to Level 1B.

9. We also leverage Sentinel dataset as a reference dataset to build out our georeferencing.

10. We do geolocation on board the satellites typically with a displacement of 300-500 m. No other corrections completed just the rough location within a target range. For example, return location of the object and showing its appearance. Customers have wanted to see data / algorithm performance, therefore raw images are retrieved as well during test phase.

11. For a future constellation, ideally we want the sensor to be calibrated pre-launch and hope that it stays calibrated. For us, it is most financially viable to undertake onboard processing or onboard calibration post-launch. We're looking to have a co-located camera with the sensor for some limited calibration. For the sensor itself, it would be well filtered, making sure we're just picking up the specific returns needed to reduce satellite size, weight, power.

12. We have calval datasets to train our models, and then validation datasets to validate. We access the freely available datasets through literature. For international work, we contract a local partner to get ground truth to train the model. We have a big science team with good connections to the universities.

13. We will not provide analysis or insights from our data, but will do internal analytics to improve the quality of the data, and better work with clients based on data from the demo launches to help understand the quality of our data. We will sell directly analytics companies for their own specific applications.

14. For the initial few launches, we have identified a couple of areas of interest and know the type of composite materials we will be looking at (know the composite materials of the area that we can extract based on spectral unmixing). Validate based on how closely we are able to replicate the calibration data we have. We will also look at how our detectors synchronise with each other when in orbit. We'll stitch images to together and want to make sure they are as synchronous as possible. Hardware restrictions and launch provider limitations in regards to power and thermal requirements. Want to analyse how best we can capture with the least amount of resources possible. We are building a data pipeline so we don't have to manually validate every time.

15. Pre-launch sensor calibration will be done at the Advanced Instrumentation Technology Centre. Once launched, we will coordinate data collection campaigns at the time of satellite overpass. For the first year, we expect to have 3-6 simultaneous data collections to ensure they are measuring on the ground the same thing that the sensors is capturing in orbit. For the technology demonstrator, we expect to only collect data around the ACT. This is due to downlink opportunities, orbit, ability to move and validate the data, data storage etc.

16. Depends on requirements and what the customer asks for. Geolocation system is based off a line-of-sight model with a number of corrections for elevation, atmospheric effects etc. It can estimate error based off error in ephemeris and the understand of the state the line is propagating through as well as know error for the modelling.

17. We encourage operators to capture over a calibration site regularly to reassess that error to enable corrections. However, it's the operators decision and not all operators are willing to do this. One of their operators has never recaptured over a calibration site. To help deal with long times between re-calibration captures, we have a system that takes the orthorectified image and checks geolocation against a reference standard like Sentinel 2 or Landsat, and use this to provide some error estimation. We use it for our own purposes but try not to quote as an authoritative error. Gives an idea when things are drifting and allows us to give a warning to operators to do a re-calibration pass if measurements go over a certain threshold; not exposing this to users but provide feedback to data providers if the location is off.

18. We have limited experience in radiometric calibration. Two of our providers lease the satellites and the ground station provides the radiometric processes. As a reseller, we don't provide sensor calibration for our downstream customers. Sometimes we have concerns about the accuracy of the satellite sensors, when the sensors don't seem to be drifting in their calibration over time as we would expect it to do. So not sure if we are really lucky with this sensor or is there is something off with the procedure.

19. For commercial data, we don't own it so can't do much with it. The data is owned by the customer we are selling to. We can do processing to mediate delivery but not much else.

20. We will still do pre-launch calibration as part of this, and aim for configurable onboard calibration post-launch. If we capture data over a cal/val site and want to update our spectral and radiometric calibration, then we can configure that onboard. But the logistics have not been thoroughly planned yet.

21. We defined their own prelaunch calibration plan based on what we thought the equipment was going to be and the testing equipment we had. For post launch calibration, we haven't looked at what standards are available. We don't know what the standard calibration process is, how to access the data, what format the data is in and how to process and relate it to our data.

D: Will you capture over Australian calibration or validation sites? Do you know of any?

1. We may suggest to the sensor team to use the Pinnacles calibration site (in Australia) as a minimum.
2. Don't know about the sites. We haven't considered capturing over these calibration sites. We're still a few years away from a pathfinder for our constellation.
3. We wouldn't even know how to access calval site data. We collect data in a way that's suitable for us. Don't know what existing calval sites have. Aware of TERN but data collected by TERN is not suitable for us. TERN collects canopy LAI, whereas we need understory LAI as well. Satellite data includes both canopy and understory, so we use that to measure total LAI.
4. There are datasets in Australia but not accessible. Either the researchers are not willing to sell them or it's cost-prohibitive. Most of the research is publicly funded through Research Development Corporations and there is no mandate to publish.
5. Tech demo will look at eucalypt forests. The validation sites selected for the tech demo are going to be in western Sydney Cumberland Plains Supersite, the National Arboretum Canberra, and the Tumbarumba Supersite (UQ). We intend to use well defined site, that are relatively large. One of the most important things for us is to find places that we actually measure the canopy. Crane towers (Cumberland and Tumbarumba) can be used for canopy validation. The next iteration of the sensor, i.e. the operational product and hyperspectral sensor will collect data across a range of ecosystems and vegetation types. We need to expand the sites, focusing mainly on measuring changes in the flammability of vegetation, at this point not thinking about black and white validation targets.
6. One of our operators capture over calibration sites in the Pilbara. Not sure about other operators.
7. We are looking at overseas calibration sites. External organisations provide data for those sites, such as our launch providers and ground station operator. Doesn't matter where the site is, just need the quality. We are working off sites from USGS catalogues and what areas that overpass. We can access international sites, as we don't think there are sites in Australia that are not available elsewhere. But we need as much calibration data as much as possible for flagship missions.

E: What data quality attributes do you want or request from your satellite providers? Have you encountered issues with data from your satellite providers?

1. Our satellite providers already do provide data quality information. Metadata provided gives accuracy of what their calibration process has been helping to ensure high quality. But we also need to make sure we are confident in the data quality we're being provided with. For example, in monitoring environmental change, iterative accuracy is becoming more important, for being able to align imagery to assist machine learning.
2. Consistency over time. Radiometric processing does not seem to be important for Australian clients but is for our overseas clients.
3. There have been some issues on the satellites, happening a lot more often due to solar flares in space that effect satellites. The satellite has to do more thruster burns to realign itself, then every time it does that there is certain recalibration needed. Occasionally there are CCD camera issues. These are issues we need to think about in the processing chain and how to account for it in an automated processing chain.
4. Where there is missing data from the satellite, the data is discarded and then recaptured.
5. We do a lot of multi-temporal work so we rely on good and consistent calibration. We try to work with operators of the system to find a solution if there's a problem.
6. Most customers care about the output or result. 80-90% is a good benchmark for accuracy of the product.
7. For the technology demonstrator, 50 m GSD is sufficient for the application we are looking at. We don't necessarily need to know where the small things are; we are looking at the landscape-scale changes. Also, it being coarser resolution helps reduce the data we need to downlink.
8. For geolocation, ideally 25 m accuracy (half a pixel). Considering the camera will slew, it will be hard to get this accuracy but this is a trade-off we are willing to make in in a demonstrator order to get more data. For future sensors and satellites, we aim to be more stringent with geolocation accuracy.

9. For data that we obtain for operators – (1) if going over a calibration site, we will compare with the information available from that site, (2) assess for drift against a known standard (sentinel, Landsat) and co-register against each other to work out what is the overall offset, what is the distribution of offsets with the image in case ours is warping issue or ortho issue.

10. In terms of radiometric calibration certainty – no requirement – just the best we could do. When we started the project in house, we didn't even have the equipment to do radiometric calibration. A lot of requirements for water quality monitoring are sub couple of percent. For spectral calibration, no hard requirements defined, best case possible – our accuracy is sub nm, with a well-controlled source to calibrate pre-launch. Post-launch, we will likely look for guidance from similar work (international gold standard mission) to do that. For signal to noise ratio - designed for 80 to 100 + signal to noise, trying to increase that for the next sensor / mission. For downstream data processing product – part of the issue is that we have to build that data cube up over time, a lot of the work was accurately referencing the stacked images together data, so one of the big constraints was trying to develop on-board processing algorithms to do that based on feature finding to build the data cube, again no hard requirement but best case. For the first mission, we know the orbit of the satellite very well. We have our own ground segment software to generate time stamps.

11. We are more likely to acquire validation data over Australia, where we can get people on ground to help acquire site data. Informed a lot by internal data scientist and experts.

F: What are your criteria to sign up satellite data providers or be an early adopter of their services?

1. Data needs to spatially accurate without needing to do much ground control. Big companies can get it to 3 m GSD, small satellites are at > 20 m GSD. So it makes it hard for small satellite provider to compete.

2. Noise in the data – compression from vendors can cause noise.

3. Band alignment, for examples the CCDs (sensors) haven't been aligned correctly.

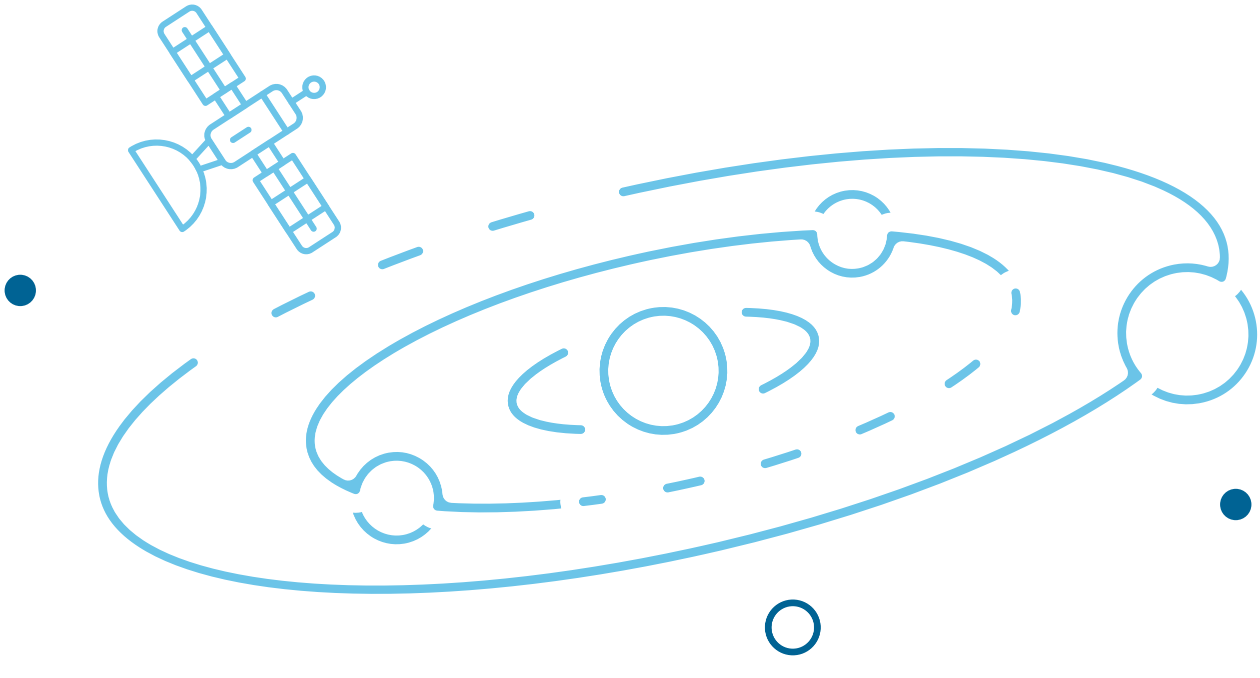
4. Vendor has changed the way they produced (calibrated) the data. We had a client come back to us for this issue, and we had to go back to the satellite provider to change their modelling.

5. Geolocation accuracy is not always the most important.

6. We have had startups approach us with a sample. If we decide not to on-sell their data we'll provide them with feedback why. So a cal/val centre and reference sites could be really useful for them.

7. We get signed up as early adopters to satellites to help quality check sensor or see if the satellite data is fit for market. We work with the big primes as well as start ups.

8. We've gotten samples of data from other suppliers, and the data has not been great. We're playing the re-seller role because we didn't have good revisit over specific areas and needed data from another operator. The data from the other supplier had terrible band alignment, really bad georeferencing and almost needed to reprocess the data entirely, but couldn't get the info needed from the supplier.



G: What data quality or cal/val attributes do you aim to achieve? What data quality or cal/val attributes do your clients request?

1. Positional accuracy may not be as important as relative accuracy. Consistency over time is important.
2. For carbon projects, client algorithms are certified by the energy regulator, so the satellite data output has to meet certain regulatory requirements for our clients.
3. Clients don't have much interest in accuracy, but they want the final product to look "good", i.e. filter out noise, clouds, etc. We had many issues in the beginning but have become in the last two years in getting to the quality that clients expect. For the applications we're in, there are people on the ground who can judge our final product based on their experience.
4. We haven't don't have requirements yet. They will be defined within individual research projects using the satellite data [as a research platform].
5. Generally, our mission data should be comparable over time. Bands should be registered accurately – standard requirements. We should be able to compare it with Landsat and Sentinel 2 data.
6. Geolocation accuracy is a client priority, target of 5 m or better. Band alignment issues are a major issue that end user customers cite.
7. Clients are happy with our Top of Atmosphere corrections. We can also provide additional metadata such as sun angles if clients request it. Envision clients will need increasing satellite revisit time, consistency of data, eg. good alignment of imagery over time for change detection analytics.
8. We aim to vertically-integrate; we have access to all the files needed from the satellite, sensor model, operational parameters. Satellites have jitter no matter how stable. So fluctuation and temperature profiles of the sensors are all additional variables and taken into account to get the best quality outcome. If you don't have additional oversight of the data and you're just reselling, you don't have that additional level of insight on data issues. We've built the workflow that captures all these potential issues.
9. Clients have requirements like wanting specific spatial resolution, specific spectral bands, geolocation accuracy. Some might ask about band alignment.
10. Customer quality requirements are provided by the original imagery provider. Some customers have stricter requirements for location, for example a government client wanted better than 5-10 m georeferencing accuracy. So far we've had no issues with our upstream providers not being able to deliver what is needed. We just sell the data and haven't heard complaints or concerns.
11. In some instances, we process collected data. In other instances, clients use their own algorithms and corrections.
12. We know we can't calibrate our pasture model, as granularity of calibration data is not enough. On the cropping side, we don't know. We would use the same data sources as used in the US, which is MPP for forage production and is very accurate. But this doesn't exist in Australia. The closest equivalent is MODIS.
13. Desire for radiometric resolution but no hard requirements on that yet because there is still a lot of testing and validation to do. Signal to noise ratio of 100 or better, but that may not be achievable in all wavelengths, especially in the SWIR. We need to refine these requirements and targets.
14. Will use Sentinel 2 data and Landsat 8, 9 and Next data, so we know we are in the correct range compared to other sensors. Something that the tech demo will do that is not ideal is that the camera will slew up to 20 degrees. The number of acquisitions will be very low in the order of 1 every three months, because it's a single sensor, orbit parameters etc. Not ideal to slew but we decided it was worth it to get as much data as possible. By comparing to Sentinel or Landsat it will tell us, are we pointing where we think we are pointing? Are we getting the data, how much contamination are we get from adjacent sites etc.?

15. Our target clients are reluctant to use our products straight away. If we can demonstrate that the flammability product has a basis on the ground – we have burned something, how quickly it catches fire and energy it gives – if we demonstrate the link between the leaf burning and the satellite product that is when they are more likely to start adopting the products and using them in their fire models. The main inputs into our clients' models are based on topographic, soil moisture, wind speed, meteorological factors. But very little related to vegetation goes into their models, like canopy height, leave density or area. We want our products to give our clients something they can use as further input for their models. So our data can be used for modelling, resource allocation, soft containment lines, prescribed burns etc. The key is to demonstrate the link between what we see on the ground and satellite data. If we can do that, the users will be more convinced that this data brings something to the table.

16. Some customers don't care about calibration at all – some of this is from customers who don't have the awareness to ask about calibration and some for some of them, we are surprised that they are not asking about it. Some of our customers have one-off purchases and they often view it as a picture from space and not a scientific dataset. When suppliers format their tiff in a way that windows image view can't open this causes some confusion.

17. We get customers who want to do machine learning but they don't ask about calibration and don't overly care when we bring it up. The main customers who care about accuracy and calibration are government customers and research/academic customers who understand of how to work with the data. We have mixed results from the larger enterprise customers who for example have very mature GIS department and will ask for it.

18. Geolocation – we provide a target within 10m but ideally within 5m. Generally trying to keep it under 3m but we can't guarantee due to some of the providers not enabling regular calibration.

19. A lot of customers come from the machine learning space don't take the time to understand the Earth Observation and remote sensing side. We try to point customers towards good references and more information. Over the medium term we are aiming to set up resources to help train our customers. For the customers that should or do care about it, they largely have the expertise. The ones that don't care are generally doing very large features that they are trying to capture and have an enough of a margin of error in the area that they have requested that it's not going to affect them.

20. The end market is after high resolution data. What they expect is geospatial accuracy, then local validation provided by the client to 1 pixel accuracy. We are looking to diversify from big primes, would have large impact if anything happened to the primes.

21. Customers want specific thresholds for cloud cover, frequency of data, quality versus quantity. Some customers might want high frequency images even if there is a lot of cloud.

H: Are your clients aware the data could come from different operators?

1. Clients don't generally ask for specific operators, for we do name the operator for end licence agreements

2. 5 years ago clients might have asked for specifics on sensors and accuracies, now they don't.

3. We make it clear to our clients that we provide synthetic data and they're happy with that. As long as they have the numbers [insights / analytics].

4. 90% don't care, and the 10% who ask are geeks.

5. Yes, it is annotated what supplier and what provider the imagery is from. We clearly label what data customers are buying in the result card, in the shopping cart, in the add-to cart so it's visible in multiple places. For those group that are "I want a picture from space" they don't care as much. They often like Maxar, their BOA products have very high colour saturation in way that isn't realistic but creates a nice picture, more popular choice but not always the cheapest.

I: What data quality issues have you encountered from providers, and any impacts on your business?

1. We've had radiometric issues for Pacific projects using Sentinel. But generally not an issue with the high-res data we work on. It's easy to co-register at high resolution, and we have the software and expertise to correct any issues.
2. Yes, early in the business when we were not ready with our processing chain. We have now fixed this and are delivering the expected standard of data to clients, but this is a continuously improving process.
3. We know the data is bad when we try to run our algorithms on the data and they spit out junk. There are issues like haloing, bands with strips of errors in them. If the issue is geometric, sometimes we can deal with it, but if the accuracy is way off, particularly with client supplied data then it can be problematic. Business-wise, sometimes we've had to say to clients there is no data available that meets the needs of their application.
4. We've tried Planet data as a small-scale trial, but the quality, jitter etc. is much harder to correct, and there is lots of noise. We haven't experimented with using our product algorithms on it. We're more interested in Sentinel 3 as a replacement for MODIS. Sentinel 3 is lower resolution but higher frequency.
5. Main issue is around the normalisation of data – need a spec sheet to go with the data showing the expected value range.
6. Difference in spatial resolution and differences in interoperability, ie. radiometric calibration / alignment. Current practice is to normalise data from different sources to same range.
7. We've had a client that refused to work with one of our specific providers because they had a bad experience with them previously [in terms of data quality].
8. We don't intend on using commercial imagery. It's too expensive on an ongoing basis. Open data is suitable for us for upscaling, and we fill the gaps with drone data. We do always find problems even in open data. This either gets solved upstream or we do it ourselves.
9. We use open data but buy commercial if there are frequency gaps and we buy different dimensions of the data.
10. Yes, once or twice – early on when we had less experience. For example early on in our processing chain, our georeferencing model was out by 1.5 km because we forgot to account for rotation of the earth. But that has all since been refined.
11. None or most of their commercial operators won't provide chain of provenance and processing information. They might say "It's a BOA product", maybe will say this it accuracy on the locating and radiometrics but their methods are proprietary. Some researchers choose not to use this data because of unknown processing applied. For metadata the biggest issue is that is not easily consumable by users (non-experts).
12. Experience with data in the past with Sentinel 2 dataset, where the algorithm on the data produced artifacts from the data and in that case, it was quite evident because they were getting high value pixels in an area they shouldn't have. The follow-on effects can be bad when looking at long time trends, especially if we don't have the time or resources to thoroughly look at the data. The issue is how this translates to people doing cal/val and deploying them, they do their best they can and no one can check every single image. We would like to see a more transparent, and layman's person description of what they do and how they do, rather than "here's the paper", as not everyone can interpret these papers to the same degree. Cal/Val is crucial – if we make procedures more transparent, then the probability of uptake by others (students etc) would be higher
13. Potentially going to see a lot of issues transitioning from MODIS to VIIRS. There may be issues due to de-orbiting of the satellite.
14. With regards to MODIS becoming decommissioned in the future, we haven't noticed orbit drift being a noticeable impact.

J: Do / will you validate your products or models? Do you undertake validation on derived products you use as input to your own products?

1. We have a merged product from Landsat and MODIS to get the highest spatial-temporal resolution.
2. Validation of our models or products using other datasets, for example derived-products such as atmospherically-corrected data from Geoscience Australia. But can be a manual process as processes like cloud-masking is not automated. We build internal confidence in our products to the point where we can say to the client we are within 95% confidence range.
3. Geometric accuracy and consistency. We spend a lot of time doing registration, i.e. get the data and register the timeseries back sitting over the top of it. Consistency is important for timeseries work. Doesn't matter if it's 50 m out as long as it's all 50 m out, so it stacks up on top of one another when doing timeseries work.
4. For data from Digital Earth Africa and (global) Sentinel archives, we need to do additional BRDF (Bi-Directional Reflectance Distribution Function) correction because ESA do not provide that. Ideally, these types of data are provided.
5. GEDI (LIDAR) data is still experimental and [sensor] algorithms can cause issues with the calibration. For our business, we also process a lot of airborne LIDAR. Clients will provide field site to validate the products, mainly when we're not sure about the quality of the satellite data. Clients are most often not experts. Clients will ask about the interpretation of the data and we explain how it should and shouldn't be interpreted or used in the future.
6. We have used lidar as ground truth to classify imagery against for a client product / project.
7. Manual validation is typical. We do a lot of object identification so it's easy to manually validate when just identifying objects. We wants to see what's available in terms of cal/val infrastructure in the future, would be keen to see outputs. If we knew what was available, such as for Lidar, then we can request imagery over the calval sites.
8. Combination of continuously operating ground sensors with field campaigns and drone campaigns.
9. Leaf area index measured on ground with a variety of methods, also manually capturing LAI with accepted gold standard and measured under perfect conditions. We quality check automatically-corrected data on the ground.
10. Yes we also undertake validation on derived products. We run validation models, i.e. algorithms to run on top of surface reflectance products. We want to know the reflectance is acceptable so we use Geoscience Australia products to do statistics to gauge performance of our models. We train our models on existing datasets such as fractional cover data existing product (eg. Landsat 30 m fraction cover). Surface accuracy is important to us, so consistency between sensors is important.

K: Would a federated calval infrastructure / service be of value to your business and how?

1. Would benefit smaller satellite data providers as opposed to big players. The trend is increasingly towards small sats. The downside is the data quality from these smallsats, whereas we can expect high geospatial accuracy from larger satellite players.
2. We have been approached by global satellite providers. We do our due diligence, look at data quality especially with smaller satellites, then decide if we want to take them on as a reseller. There could be severe consequences if the data quality is poor.
3. No, but that doesn't mean that others won't. We would promote such a capability to our partners or clients and ensure others know about it.
4. No. With the sensors from the operators we use, we've not come upon a registration or recalibration issue on our downstream projects. If we used a lot of use Landsat, Sentinel, then maybe they could see value in additional cal/val infrastructure, but not for the high-res applications we typically work on.
5. No, don't see much value. If we're spending a lot of money to process, store and calibrate data, for the same amount of money we could just buy high-resolution imagery.
6. Data isn't the problem. We are focusing too much on the Landsat and Copernicus projects and data from these should just be used to help augment some other data we're already using. But startups might benefit. The problem is startups might now understand calval requirements and how calval fits them.

7. Calval is a cubesat and smallsat problem.
8. Would be useful to have additional calval to have additional confidence on the products we're generating.
9. A hurdle is accessing different datasets from different cloud workflows. Ideally data and workflows can be interoperable in the cloud. Data in FTP servers are not as useful if we can't scale access to it. Calval infrastructure should also be about making the data and process easier for others. Ideally there is interoperability from cloud to cloud. Now if something is hosted in AWS, it can't reach clients in other clouds like Google or Microsoft.
10. We want Australian datasets that are calibrated to the conditions in Australia. Not interested in US data for example.
11. We're interest to have calibrated and validated datasets to compare our products against, so that we can give a measure of quality that clients can verify against.
12. Any support in general where we can cross contribute to a federated cal/val program and vice versa would be very helpful – we've had to do a lot of this [processing] ourselves.
13. We aim to build our own satellites and possibly get access to other satellites. We're keen to build out more capability processes and homogenise datasets from various data across Australia.
14. No value in additional calval. What we lack is representation at the international level. What we want is the Australia Space Agency to represent Australian industry, eg. talking directly to ESA but need ASA at high level.
15. It would be great to have some sites out there where you can get an idea of how well your algorithm or sensor is performing without having to set up a whole program yourself would be a win to the Australia industry. Ideally, we can test and trial new techniques, knowing that we are comparing it to well collected, reliable field data.
16. Field data is important. For example, there is a market in WA but there is limited biophysical data there for validation of products, so makes it hard to sell products in those markets.
17. TERN provides biophysical data, land cover, tree cover, fractional, biomass but the low revisit rate for TERN sites is shocking. It would be good to have quarterly updates of all those sites that would give seasonality, fire regrowth and recruitment. Strong consistency across the Australian network of data and building upon already invested-upon infrastructure would be valuable. Enables cal and val, provides more trust in EO products, sets benchmarks. If you can't produce something that can effectively translate the information to match with the data from field sites, then you know something is not right and you can unpick the problem. Consolidating what we do have more frequently would be of assistance, such as revisiting the AusPlots site more often.
18. We see the benefit of an Australia-led type of program that integrates well with global datasets. International consistency would be good. If a program like this can plug into existing protocols, there will be more uptake of cal and val.
19. A way to normalise data between datasets to make it more interoperable.
20. We work with radar from Sentinel 1 and our models are built on Sentinel 1 baselines. Having continuity between S1 and the next mission for us will be critical. Our models are built on Sentinel 1 baselines. The sensor calibration has to be spot on so that we know that the variations we're detecting are real, and we can disaggregate that into soil moisture and tree cover.
21. Main gap is knowing what data is out there that we can use. Field validation protocols needs to be updated (last published in 2012). Needs to be a national effort to update the protocols for ground validation but also developing protocols for spaceborne validation. Set of standards that we can use throughout the industry that would be useful. It is something mission developers should think more about. Do these protocols exist? How do I know my sensor measuring the same thing over and over? Need to start working towards this. Nation-wide there should an effort to make those protocols or have those conversations because if we are building the satellites, we should be ensuring we provide data that is useable, not just pretty pictures.
22. Technical documentation and workshops are needed because I don't think building the satellites or using that data is the hard part but making sure that the data is good and is key for all in the space data.

L: Would you be willing to pay for outsourced or commercial cal/val services?

1. If the calibration sites are being maintained and satellites are going over then you want to make sure they are accurate etc so yes, we would support that.
2. We looked at the buy option early on but was really expensive. Went to a lot of companies who could build out a process chain but what they found the cost of build and then ongoing cost per sq km / the cost model became prohibitive. The commercial model became prohibitive. We wanted to be competitive in terms of data price. We have pushed down the road of "build", and while it has its challenges, we were able to resolve this at half the cost of buying. Building has been really worthwhile in Australia. We an Aussie company that has this capability and not reliant on foreign vendor for those models. But still need further help to build out the chain further.
3. We try to do everything in house mainly because of budget. But if we had someone we could ask for advice (e.g. Geoscience Australia on their data processing methods) that would be useful to make the connections.
4. Paying would go upstream to customers. In this scenario, we see ourselves acting as a reseller of calval data, where we would have an agreement with the supplier that they will get the calibration data from some source, acquire the input and use it.

M: Accessing Australian cal/val infrastructure and datasets

Do you undertake or will you undertake cal/val before passing EO data, whether acquired by you or by satellite operators, onto clients?

- Yes: 83-92%. Reason for the variation is that a couple of organisations undertaken cal/val to check data quality but do not apply any corrections or further processing.
- Reasons for 'No' include:
  - Cal/val is undertaken to check the quality of the data, however data is passed on un-interfered to clients as part of the business model.
  - Data not intended to be used or was not usable

Do you have a cal/val plan for your intended mission/s?

- Yes: 4/7. One organisation is represented twice in this question, having launched a mission and developing the next.
- Reasons for 'No' include:
  - Data not intended to be used or was not usable.
  - Unknown at this stage

Will your mission/s calibrate over dedicated on-ground reference sites?

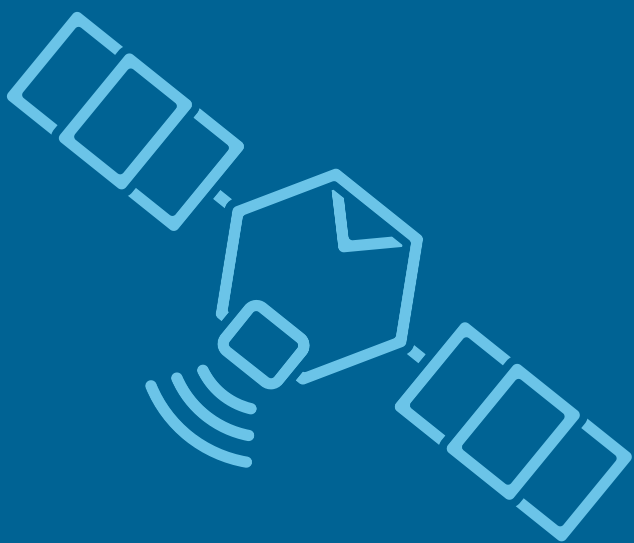
- Yes: 4/7. One organisation is represented twice in this question, having launched a mission and developing the next.
- Reasons for 'No' include:
  - Data not intended to be used or was not usable.
  - Unknown at this stage

Are you aware of Australian cal/val infrastructure and datasets?

- No: >60%

Do you know how to access the infrastructure and datasets?

- No: >80%. Those who responded 'Yes' (2/12) are part of the EO research community.



FRONTIER SI

frontiersi

frontiersi.com.au