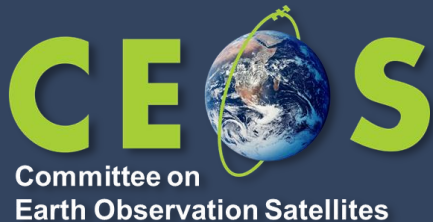
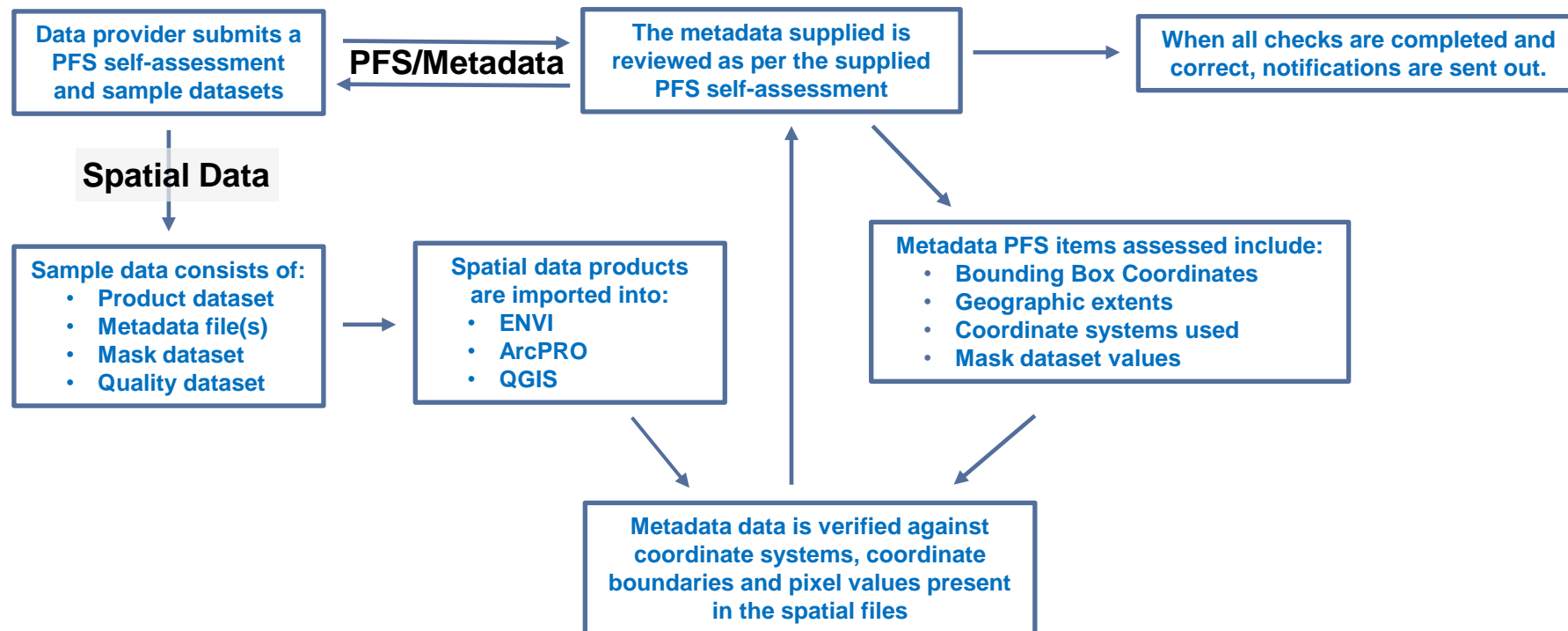


# CEOS-ARD Data Quality Issues



**Peter Harrison**  
**Medhavy Thankappan**  
**Geoscience Australia**  
**Agenda Item 4.1**  
**LSI-VC 17**  
**14 APRIL 2025**

# PFS Self-Assessment Review



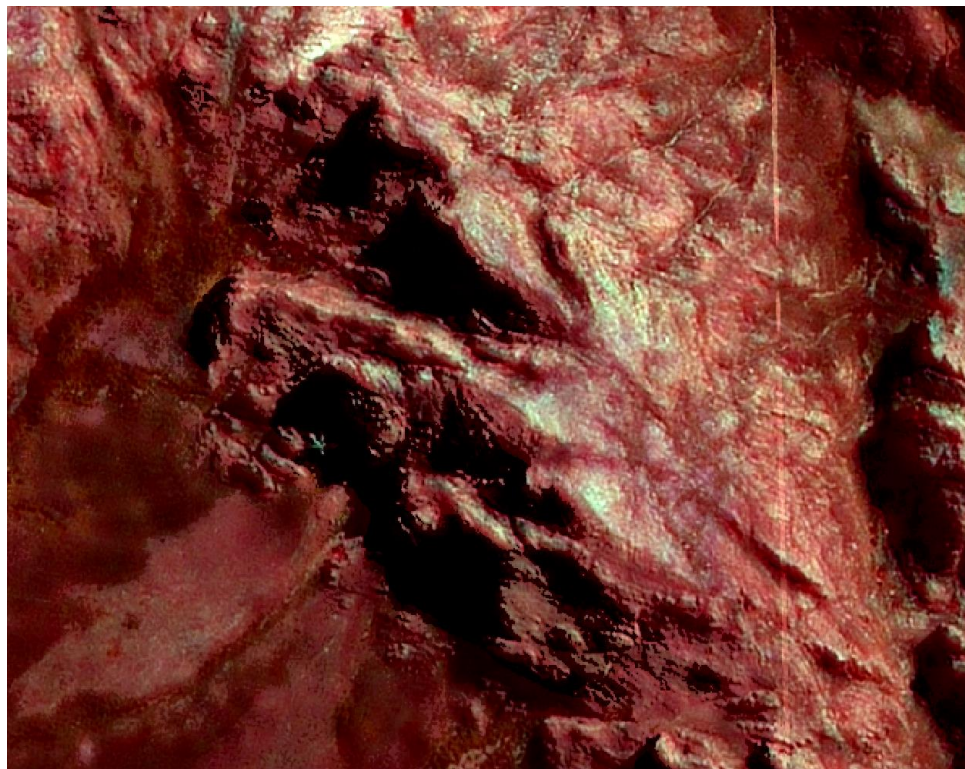
- ❖ Reviews are an iterative process
- ❖ The reviews can be complex with metadata spread over multiple files and formats
- ❖ Metadata does not necessarily follow the PFS parameter order
- ❖ Metadata in some PFS self-assessments do not follow a standard item tag (variations between Providers)
- ❖ The submitted PFS version is incorrectly identified or uses an earlier version
- ❖ The spatial data samples provided are checked to ensure mask codes, bounding coordinates, etc are correct as per the metadata. By doing so, it has flagged issues with the Providers data quality

# Data Quality Issues



## Case 1 Metadata Quality Issue

```
<PerPixelMetadata>
  <Filename>[REDACTED].tif</Filename>
  <SampleType>Mask</SampleType>
  <DataFormat>GeoTIFF</DataFormat>
  <NoDataValue>0</NoDataValue>
  <DataType>UINT</DataType>
  <BitsPerSample>8</BitsPerSample>
  <ByteOrder>N/A</ByteOrder>
  <BitValues>
    <NoData>0</NoData>
    <IncompleteTesting>1</IncompleteTesting>
    <Saturation>2</Saturation>
    <Cloud>10</Cloud>
    <CloudShadow>30</CloudShadow>
  </BitValues>
  <SolarAndViewingGeometry>
    <AngleOfSolarElevation>61.730305</AngleOfSolarElevation>
    <SolarAzimuth>56.883973</SolarAzimuth>
  </SolarAndViewingGeometry>
</PerPixelMetadata>
```



# Data Quality Issues

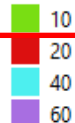


## Case 1 Metadata Quality Issue

```
<PerPixelMetadata>
  <Filename>[redacted].tif</Filename>
  <SampleType>Mask</SampleType>
  <DataFormat>GeoTIFF</DataFormat>
  <NoDataValue>0</NoDataValue>
  <DataType>UINT</DataType>
  <BitsPerSample>8</BitsPerSample>
  <ByteOrder>N/A</ByteOrder>
  <BitValues>
    <NoData>0</NoData>
    <IncompleteTesting>1</IncompleteTesting>
    <Saturation>2</Saturation>
    <Cloud>10</Cloud>
    <CloudShadow>30</CloudShadow>
  </BitValues>
  <SolarAndViewingGeometry>
    <AngleOfSolarElevation>61.730305</AngleOfSolarElevation>
    <SolarAzimuth>56.883973</SolarAzimuth>
  </SolarAndViewingGeometry>
</PerPixelMetadata>
```

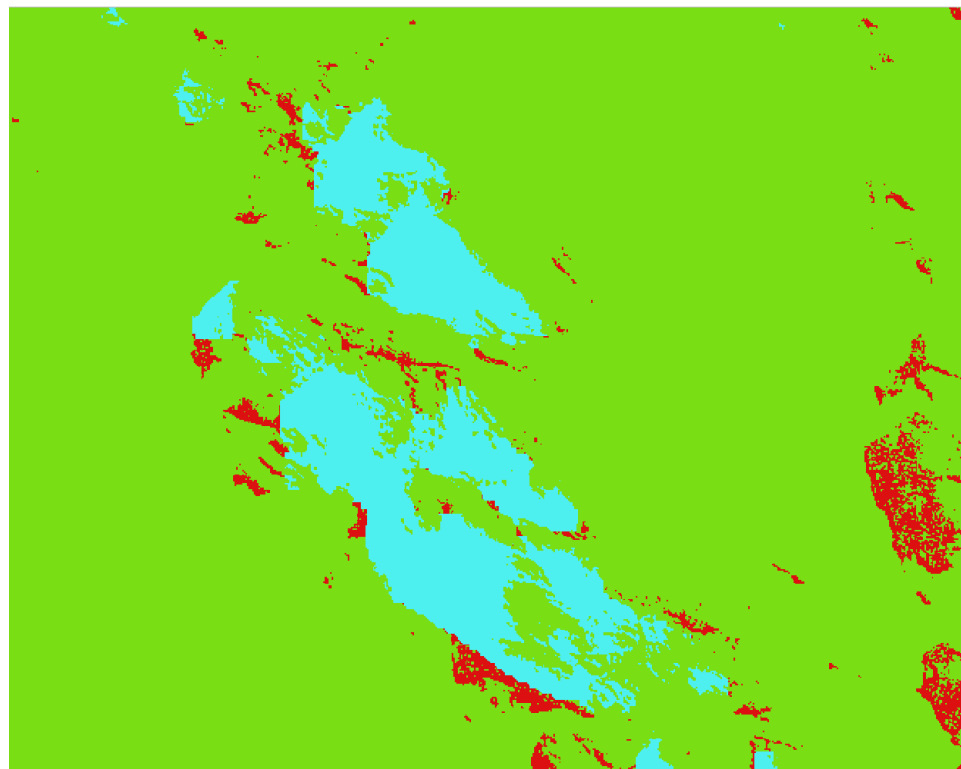


Band 1: Pixel Quality (Gray)



Pixel Value misidentified

Missing Pixel Values in the Metadata





## Case 1

```
<PerPixelMetadata>
  <Filename>[REDACTED].tif</Filename>
  <SampleType>Mask</SampleType>
  <DataFormat>GeoTIFF</DataFormat>
  <NoDataValue>0</NoDataValue>
  <DataType>UINT</DataType>
  <BitsPerSample>8</BitsPerSample>
  <ByteOrder>N/A</ByteOrder>
  <BitValues>
    <NoData>0</NoData>
    <IncompleteTesting>1</IncompleteTesting>
    <Saturation>2</Saturation>
    <Clear>10</Clear>
    <Water>20</Water>
    <CloudShadow>30</CloudShadow>
    <TerrainShadow>40</TerrainShadow>
    <Cloud>50</Cloud>
    <SnowIce>60</SnowIce>
  </BitValues>
  <SolarAndViewingGeometry>
    <AngleOfSolarElevation>61.730305</AngleOfSolarElevation>
    <SolarAzimuth>56.883973</SolarAzimuth>
  </SolarAndViewingGeometry>
</PerPixelMetadata>
```

\* Originally misidentified as 10

## Summary

In the initial submission, several mask pixel values were not referenced in the metadata.

The missing mask values were “Not Required” as they were Goal parameters as per the SR PFS.

Technically, the product achieved Threshold status as it met the requirements for Threshold but was missing essential metadata for users.

## Case 2 Product Quality Issue

### PFS Self-assessment as submitted

1.13	Algorithms	<p>All algorithms, and the sequence in which they were applied in the generation process, are identified in the metadata. For example, these may be available through Algorithm Theoretical Basis documents.</p> <p><i>Note 1: Information on algorithms should be available in the metadata as a single DOI landing page.</i></p>	<p>As threshold, but only algorithms that have been published in a peer-reviewed journal.</p> <p><i>Note 1: It is possible that high quality corrections are applied through non-disclosed processes. CEOS-ARD does not per-se require full and open data and methods.</i></p> <p><i>Note 2: Information on algorithms should be available in the metadata as a single DOI landing page.</i></p>	Yes	Yes	<p>Surface reflectance retrieval algorithms have been published in the peer-reviewed journals.</p> <p>Note: See Item 1.9 for Complete [REDACTED] instrument DOI listings.</p>	
------	------------	--	--	-----	-----	---	--

No references or further details given

Linked to a totally different satellite product landing page

# Data Quality Issues



## Case 2

```
<GFSR_metadata>
  <global_metadata>
    <data_provider>[REDACTED]</data_provider>
    <satellite>[REDACTED]</satellite>
    <instrument>[REDACTED]</instrument>
    <acquisition_date>2023-05-22</acquisitic
    <scene_center_time>02:36:09Z</scene_cent
    <cloud_cover>0</cloud_cover>
    <solar_angles azimuth="121.512310" zenith
    <satellite_angles azimuth="101.374341" z
    <OrbitID>054279</OrbitID>
    <top_of_atmosphere_radiometric_rescaling
      <radiance_mult b1="0.1889" b2="0.1495"
      <radiance_add b1="0" b2="0" b3="0" b4=
    </top_of_atmosphere_radiometric_rescalin
    <central_wavelength b1="0.485" b2="0.555
    <latitude_longitude_corner_information
      <corner_longitude="114.957999" latitude
```

Metadata states 0 Cloud  
Cover verified by visual  
inspection of the image



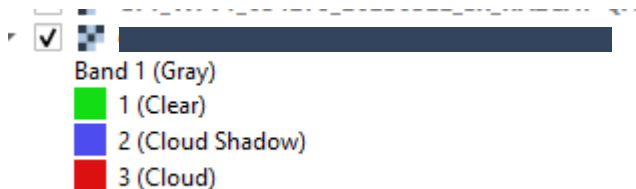


# Data Quality Issues

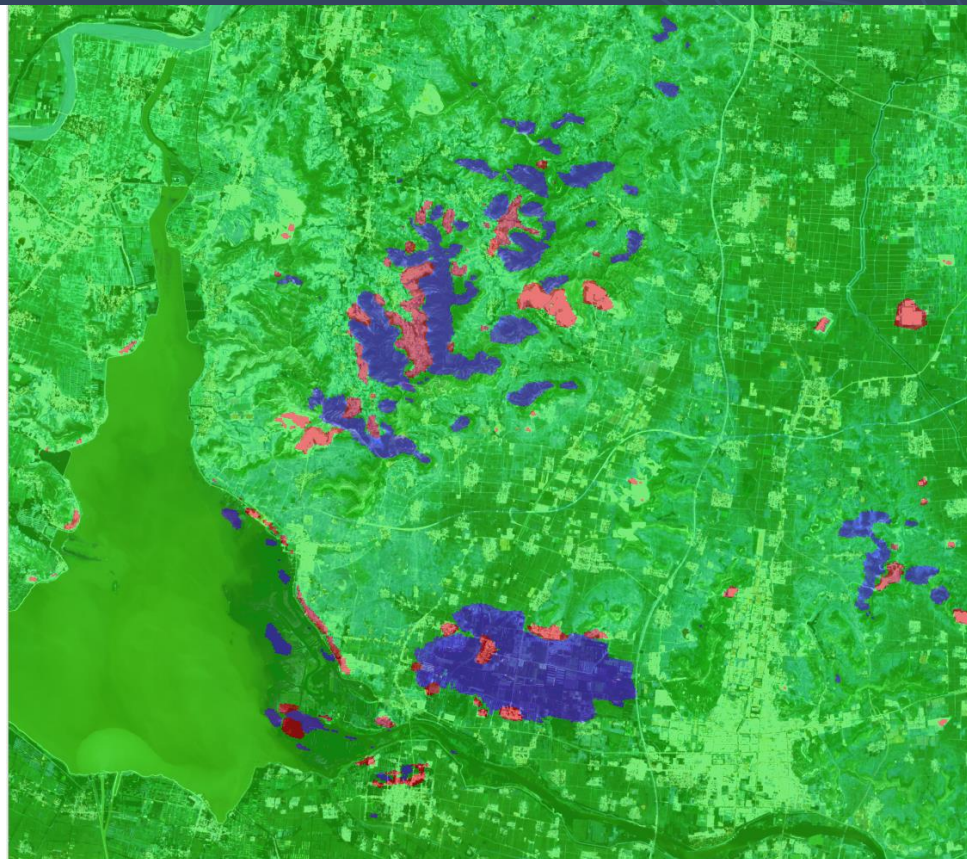


## Case 2

```
<band name="PIXEL-QA" category="qa" data_type="UINT8" nlines:
<file_name>[REDACTED]TIF</file_name>
<pixel_size units="meters" x="16" y="16"/>
<data_units>quality/feature classification</data_units>
<bitmap_description>
  <value num="1">clear</value>
  <value num="2">cloud shadow</value>
  <value num="3">cloud</value>
</bitmap_description>
<app_version>[REDACTED]</app_version>
<production_date>2024-05-11T14:44:23Z</production_date>
</band>
<band add_offset="0.000000" scale_factor="0.000100" saturate
```



The mask identifies approximately 1% of pixels as cloud and a further 1% of pixels as cloud shadow when the image is cloud free.



## Case 2

### Summary

- ❖ The product meets the Threshold level metadata requirements as per the PFS self-assessment (possible mask pixel values correctly identified in the metadata). This is not in conflict with a dataset having no cloud coverage.
- ❖ If the self-assessment was being reviewed solely on the PFS metadata requirements, the product technically achieved Threshold status as it met the requirements as set down by the PFS.
- ❖ However, on checking the mask, it was quickly realised that a number of mask pixels (2%) have been mis-identified as being cloud and cloud shadow.

## Conclusion

- ❖ The Self-Assessment Review process is both an iterative and parallel process with cross referencing required between both the metadata file(s) and sample datasets supplied. Can be complex process when metadata is spread over several files.
- ❖ Case 1: If just assessing the metadata, the product met CEOS-ARD metadata compliance as per the Threshold specifications set out in the PFS. However in the case of the Mask spatial file, close inspection revealed several issues:
  - Mask values were incorrectly identified in the metadata.
  - Mask values were omitted from the metadata as it was not required for Threshold level assessment under the PFS.

## Conclusion

- ❖ Case 2: The review flagged concerns over product quality. Whilst it would meet CEOS-ARD metadata compliance as per the PFS metadata specifications, the quality of the physical product is highly suspect. In this instance, the metadata states the SR data is cloud free, verifiable with the image. However, the mask dataset erroneously identifies the presence of cloud and cloud shadow pixels.
- ❖ This raises questions on how to deal with submissions. That is,
  1. Continue as we are, or
  2. Do we just solely assess the metadata to the PFS requirements, or
  3. Extend the process to ensure the quality of products associated or expected with the CEOS-ARD brand.



# Thank you

*medhavy.thankappan @ga.gov.au*  
*peter.harrison @ga.gov.au*