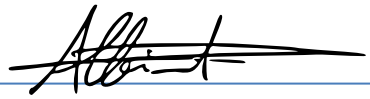
Review of Analysis Ready Data Specification

Author:		***Samuel Hunt*** *NPL, Task Technical Expert*
Approval:		***Nigel Fox*** *NPL, Task Lead*
Accepted:		***Clément Albinet*** *ESA, Project Technical Officer*

EDAP.REP.062 *Issue: 1.0* *31 October 2021*

AMENDMENT RECORD SHEET

The Amendment Record Sheet below records the history and issue status of this document.

ISSUE	DATE	REASON
1.0	19/11/21	Submission to ESA

TABLE OF CONTENTS

1. INTRODUCTION

Defining and preparing Analysis Ready Data (ARD) product is crucial to expanding the utility and adoption of Earth observation data, particularly by non-experts in remote sensing. Ideally, ARD product design could be underpinned by industry-wide standards, maximising potential product interoperability. In this report a draft standard for ARD is proposed, which expands the scope of the existing CARD4L (CEOS Analysis Ready Data for Land) to apply broadly for the optical satellite data products, particularly those developed in the commercial sector.

1.1 Scope

In Section 2 of the this document a brief review of the current ARD landscape is described, outlining the motivation for this activity. Section 3 describes the draft ARD specification for optical satellite data products prepared. Section 4 illustrates how evaluations against this standard could be reported in the form of a maturity matrix.

1.2 Acronyms & Abbreviations

ARD	Analysis Ready Data
CEOS	Committee on Earth Observation Satellites
CARD4L	CEOS Analysis Ready Data for Land
EDAP	Earthnet Data Assessment Pilot
EO	Earth Observation
IVOS	Infrared and Visible Optical Sensors
JACIE	Joint Agency Commercial Imagery Evaluation
LSI-VC	Land Surface Imaging – Virtual Constellation
QA4EO	Quality Assurance Framework for Earth Observation
VH-RODA	Very High-resolution Radar & Optical Data Assessment
WGCV	Working Group on Calibration and Validation

1.3 Reference Documents

[RD-1] CEOS LSI, “CARD4L Product Family Specification - Surface Reflectance,” 2020. [Online]. Available: https://ceos.org/ard/files/PFS/SR/v5.0/CARD4L_Product_Family_Specification_Surface_Reflectance-v5.0.pdf.

[RD-2] CEOS LSI, “CARD4L Product Family Specification - Surface Temperature,” 2020. [Online]. Available: https://ceos.org/ard/files/PFS/ST/v5.0/CARD4L_Product_Family_Specification_Surface_Temperature-v5.0.pdf.

[RD-3] A. Piro, P. Castracane, S. Scifoni, and F. Niro, “2nd VH-RODA 2021 Workshop - Summary Report,” 2021.

[RD-4] S. Hunt, N. Fox, and C. Albinet, “Mission Quality Assessment Guidelines,” 2019.

2. REVIEW OF ANALYSIS READY DATA LANDSCAPE

Analysis Ready Data (ARD) are satellite products that have a well-defined format, where common data preparation operations are handled prior to distribution to users. For land surface imagery, this can include operations such as image clipping, geometric correction, and atmospheric correction amongst others. The objective is to reduce the burden of pre-processing on the data user, which can require significant effort, infrastructure, and remote sensing expertise.

In recent years, there has been significant discussion within the EO community aimed at working towards an industry-wide agreement on how to practically implement ARD. This includes some work to develop of the required underpinning definitions and standards to achieve this.

The engagement in these discussions typically comes from two groups – traditional "institutional" space entities, such as space agencies and public research institutions, and "new space" entities, such as commercial satellite vendors. Though the motivation and priorities of these institutions can be different, there is a desire to find a consensus in this area to progress as an industry.

In the remainder of this section, recent progress in this area is outlined.

2.1 Institutional Space and ARD – CARD4L

The Committee on Earth Observation Satellites (CEOS) Analysis Ready Data for Land (CARD4L) initiative constitutes the most concrete practical definition of ARD to data. It defines a set of Earth Observation product specifications, for different "product families", that look to ensure that compliant data:

> *"have been processed to a minimum set of requirements and organized into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets."*

The initiative is led by CEOS' LSI-VC (Land Surface Imaging – Virtual Constellation) group.

The CARD4L product family specifications target geophysical variables derived from land surface imaging, currently including surface reflectance [RD-1] and surface temperature [RD-2] amongst others. Each product family specification defines a "Threshold" and "Target" level of compliance.

LSI-VC coordinates a process that assesses the compliance of products to CARD4L product family specifications. This process begins with a self-assessment by the product provider, which is peer reviewed CEOS-WGCV (Working Group Calibration Validation). Approved products may advertise themselves as CARD4L compliant and are listed on the CARD4L website[1]. To date the Landsat Collection 2 surface reflectance and surface temperature products are the only CARD4L compliant products (to Threshold level), though several more are under peer-review/development, include ESA's Sentinel-2.

2.2 New Space and ARD – Commercial Perspective

Commercial "new space" entities are, of course, particularly incentivised to prepare products that are the most usable for the widest range of customers. As part of this effort, designing products that follow ARD principles is a significant component. Further, it is widely understood that such ARD implementation can be most effective if the approach is standardised, where possible, across

[1] See https://ceos.org/ard

the EO industry, enabling dataset interoperability. Such a standardisation activity requires significant collaboration between new and institutional space entities.

Many new space companies have participated in international workshops for a number of years to foster this collaboration, including JACIE[2] (Joint Agency Commercial Imagery Evaluation) in the US and VH-RODA (Very High-resolution Radar & Optical Data Assessment) in Europe (e.g. [RD-3]). Further, Planet and Maxar have taken a particular lead by initiating an annual ARD Workshop[3] that aims to "advance interoperability and collaboration in the Earth observation and remote sensing industry" and is attended by representatives from a mix of commercial and institutional space entities.

In such workshops, the development of an agreed way forward has so far proved a challenge, in part due to variety of satellite systems and applications any standard should cover. What constitutes "analysis ready" can vary widely across the breadth of these domains. So, although CARD4L is typically discussed as a potential standard that could be adopted more widely, it is focused on a specific set of product families and applications that may not be suitable for many commercial missions. For example, in the optical domain the CARD4L product family specification is primarily aimed at quantitative analysis of land surface reflectance time-series – the needs of which are not particularly relevant for many very high-resolution imaging missions, where calibration may not even be necessary for some applications.

To make progress on defining a standard that can more apply more widely to the span of missions in the optical domain the CEOS-WGCV-IVOS (Infrared & Visible Optical Sensors) subgroup led a series of teleconferences with representatives from commercial space companies (including Maxar, Planet, Urthecast, Deimos, Space Flight Industries, and European Space Imaging) in early 2020. The resulting draft ARD specification prepared by this group is described in detail in Section 3.

[2] See https://www.usgs.gov/core-science-systems/eros/calval/jacie

[3] See https://www.ard.zone

3. ANALYSIS READY DATA DRAFT SPECIFICATION

The ARD specification defined in this section aims to expand the scope of the CARD4L surface reflectance product family specification (see Section 2.1) to cover the span of optical satellite mission products more fully. This definition activity was undertaken by representatives of institutional and commercial space entities in early 2020, led by CEOS-WGCV-IVOS, with the specification currently in a draft state.

The approach taken was to adopt the same specification structure defined by CARD4L, while expanding the previously defined "Threshold" and "Goal" levels (for surface reflectance) to four new classes that cover the typical application areas encountered across the optical domain. These application classes are defined in Table 1. It is important to note that these application classes are not intended to be hierarchical, i.e., missions are not expected to aim to progress from Class 1 to Class 4. Each mission is designed with its own design trade-offs for a particular class of applications.

The relevant requirements defined for General Metadata (Table 2), Per-Pixel Metadata (Table 3), Radiometric and Atmospheric Corrections (Table 4) and Geometric Corrections (Table 5) for each application class is defined in the remainder of this section.

NB: Wording that deviates from the original CARD4L requirements are indicated in the specifications in red.

Table 1 - Proposed Classes of ARD for commercial land imaging applications

	Description	Driver	Comment
Class 1	High spatial resolution imagery	Imagery analysis, such as object detection (e.g., car counting)	
Class 2	Land Imaging	Imagery analysis, less demanding quantitative analysis	Based on current CARD4L "Threshold" level, minimum for quantitative analysis
Class 3	Time-series observation	Change detection, agriculture	
Class 4	Climate	Climate observation, litigation	Based on/extension to current CARD4L "Goal"

3.1 General Metadata

As described by the CARD4L specification, General Metadata are:

"These are metadata records describing a distributed collection of pixels. The collection of pixels referred to must be contiguous in space and time. General metadata should allow the user to assess the overall suitability of the dataset, and must meet the following requirements"

Table 2 - Adaptation to CARD4L *General Metadata* specification for application to commercial missions

#	Item	CLASS 1 High spatial resolution imagery	CLASS 2 Land Imaging	CLASS 3 Time-series observation	CLASS 4 Climate
1.1	**Traceability**	Not required.	Not required.	Route towards evaluating uncertainty and evidencing SI-traceability should be described. Although some evidence and methods may be proprietary and not necessarily made visible and fully accessible.	Data must be fully and evidentially traceable to SI reference standard. *Note 1: Relationship to 3.2. Traceability requires an estimate of measurement uncertainty.* *Note 2: Information on traceability should be available in the metadata as a single DOI landing page.*
1.2	**Metadata Machine Readability**	Metadata is provided in a structure that enables a computer algorithm to be used consistently and to automatically identify and extract each component part for further use.	As Class 1.	As Class 1, but metadata should be provided in a community endorsed standard that facilitates machine-readability, such as ISO 19115-2.	As Class 3.

1.3	**Data Collection Time**	The data collection time is identified in the metadata, expressed in date/time, to the second, with the time offset from UTC unambiguously identified.	As Class 1.	Acquisition time for each pixel is identified (or can be reliably determined) in the metadata, expressed in date/time at UTC, to the second.	As Class 3.
1.4	**Geographical Area**	The surface location to which the data relates is identified, typically as a series of four corner points, expressed in an accepted coordinate reference system (e.g., WGS84).	As Class 1.	The geographic area covered by the observations is identified specifically, such as through a set of coordinates of a closely bounding polygon. The location to which each pixel refers is identified (or can be reliably determined) with the projection system (if any) and reference datum provided.	The geographic area covered by the observations is identified specifically, such as through a set of coordinates of a closely bounding polygon. The location to which each pixel refers is identified (or can be reliably determined) with the projection system (if any) and reference datum provided.
1.5	**Coordinate Reference System**	The metadata lists the coordinate reference system that has been used.	As Class 1.	As Class 1.	As Class 1.
1.6	**Map Projection**	The metadata lists the map projection that has been used and any relevant parameters required in relation to use of data in that map projection.	The metadata lists the map projection that has been used and any relevant parameters required in relation to use of data in that map projection.	As Class 1.	As Class 1.
1.7	**Geometric Correction Methods**	Not required. The user is not explicitly advised of the geometric correction source and methods.	As Class 1.	Indicative Information on geometric correction methods should be made available in the metadata as a single DOI landing page, ideally including reference database and auxiliary data such as elevation model(s) and reference chip-sets. However,	As Class 2, except detailed information must be public.

				detailed information may be retained for proprietary reasons.	
1.8	**Geometric Accuracy of the Data**	Estimated/specification value provided	Measured value with statement of method	Evidence of verification by an internationally recognised method (ideally, but not necessarily, by an independent team)	The metadata includes metrics describing the assessed geodetic accuracy of the data, expressed units of the coordinate system of the data. Accuracy is assessed by independent verification (as well as internal model-fit where applicable). Uncertainties are expressed quantitatively, for example, as root mean square error (RMSE) or Circular Error Probability (CEP90, CEP95), etc. *Note 1: Information on geometric accuracy of the data should be available in the metadata as a single DOI landing page.*
1.9	**Instrument**	The instrument used to collect the data is identified in the metadata.	As Class 1.	As Class 1.	As Class 1, but information should be available in the metadata as a single DOI landing page with references to the relevant CEOS Missions, Instruments, and Measurements Database record.
1.10	**Spectral Bands**	Information to allow the effective wavelength and bandwidth to be described e.g. central wavelength, upper and lower limits, FWHM etc expressed in SI units and identified in the metadata	As Class 1.	As Class 1, with instrument spectral response details (e.g., full spectral response function) also included or directly accessible using details in the metadata. Central wavelength and bandwidth at full-width half	Instrument spectral response details (e.g., full spectral response function) included or directly accessible using details in the metadata. *Note 1: Information on spectral bands should be available in the*

				maximum value of the relative spectral response function are provided at least. *Note 1: Information on spectral bands should be available in the metadata as a single DOI landing page.*	*metadata as a single DOI landing page.*
1.11	**Sensor Calibration**	Not required. The general metadata does not include sensor calibration details.	As Class 1.	Sensor calibration parameters have been determined and are available on request (subject to proprietary agreements with provider)	Sensor calibration parameters and methods used are identified in the metadata, or can be accessed using details included in the metadata. Ideally this would support machine-to-machine access. *Note 1: Information on sensor calibration should be available in the metadata as a single DOI landing page.*
1.12	**Radiometric Accuracy**	Not required. The general metadata does not include information on the radiometric accuracy of the data.	The metadata states the provider estimated radiometric accuracy.	The metadata states the measured value of radiometric accuracy and the method used.	The metadata includes metrics describing the assessed absolute radiometric uncertainty of the version of the data or product, expressed as absolute radiometric uncertainty relative to appropriate, known reference sites and standards (for example, pseudo-invariant calibration sites, rigorously collected field spectra, PICS, Rayleigh, DCC, etc.). Clear evidence demonstrating the accuracy (method independently verified) and the timescale of its validity should be provided.

					Note 1: Information on radiometric accuracy should be available in the metadata as a single DOI landing page.
1.13	**Algorithms**	No requirement.	All algorithms, and the sequence in which they were applied in the generation process, are identified in the metadata, although details of functionality etc may be proprietary and not public.	As Class 2, but with Evidence of verification of performance of the algorithms For example, these may be available through high level Algorithm Theoretical Basis documents with an associated verification plan. *Note 1: It is possible that high quality corrections are applied through non-disclosed processes. CARD4L does not per-se require full and open data and methods.* *Note 2: Information on algorithms should be available in the metadata as a single DOI landing page*	As Class 3, but only algorithms that have been published in a peer-reviewed journal. *Note 1: For Climate and other high integrity applications details of the algorithm and its verification should be made available to the community.* *Note 2: Information on algorithms should be available in the metadata as a single DOI landing page.*
1.14	**Auxiliary Data**	No requirement.	The metadata identifies the sources of auxiliary data used in the generation process, ideally expressed as a single DOI landing page, where this information is not proprietary. *Note 1: Auxiliary data includes DEMs, aerosols, etc. data sources.*	As Class 2, but information may not be proprietary.	As threshold, but information on auxiliary data should be available in the metadata as a single DOI landing page and is also available for free online download, contemporaneously with the product or through a link to the source.

1.15	**Processing Chain Provenance**	Not required.	Not required.	Not required.	Information on processing chain provenance should be available in the metadata as a single DOI landing page containing detailed description of the processing steps used to generate the product, including the versions of software used, giving full transparency to the users.
1.16	**Data Access**	Information on data access should be available in the metadata as a single DOI landing page. *Note 1: Manual and offline interaction action (e.g., login) may be required.*	As Class 1.	As Class 1.	As Class 1.
1.17	**Overall Data Quality Metrics**	Not required.	As Class 1.	Machine-readable metrics describing the overall quality of the data are included in the metadata, at minimum the cloud cover extent, i.e.: Proportion of observations over land (c.f. ocean) affected by non-target phenomena, e.g., cloud and cloud shadows	As Class 3.

3.2 Per-Pixel Metadata

As described by the CARD4L specification, *Per-pixel Metadata* are:

"The following minimum metadata specifications apply to each pixel. Whether the metadata are provided in a single record relevant to all pixels or separately for each pixel is at the discretion of the data provider. Per-pixel metadata should allow users to discriminate between (choose) observations on the basis of their individual suitability for application."

Table 3 - Adaptation to CARD4L *Per-Pixel Metadata* specification for application to commercial missions

#	Item	CLASS 1 High spatial resolution imagery	CLASS 2 Land Imaging	CLASS 3 Time-series observation	CLASS 4 Climate
2.1	**Metadata Machine Readability**	Metadata is provided in a structure that enables a computer algorithm to be used to consistently and automatically identify and extract each component part for further use.	As Class 1.	As Class 1.	As Class 1.
2.2	**No Data**	Pixels that do not correspond to an observation ('empty pixels') are flagged.	As Class 1.	As Class 1.	As Class 1.
2.3	**Incomplete Testing**	The metadata identifies pixels for which the per-pixel tests (below) have not all been successfully completed. *Note 1: This may be the result of missing ancillary data for a subset of the pixels.*	As Class 1.	The metadata identifies which tests have, and have not, been successfully completed for each pixel.	As Class 3.

2.4	**Saturation (Exceeded Linear Response Range)**	Metadata indicates where one or more spectral bands are saturated.	As Class 1.	Metadata indicates which pixels are saturated for each spectral band.	As Class 3.
2.5	**Cloud**	Not required.	Metadata indicates whether a pixel is assessed as being cloud.	As Class 2.	As threshold, information on cloud detection should be available in the metadata as a single DOI landing page.
2.6	**Cloud Shadow**	Not required.	Metadata indicates whether a pixel is assessed as being cloud shadow.	As Class 2.	As threshold, but information on cloud shadow detection should be available in the metadata as a single DOI landing page.
2.7	**Land/Water Mask**	Not required.	As Class 1.	As Class 1.	The metadata indicates whether a pixel is assessed as being land or water. Information on land/water mask should be available in the metadata as a single DOI landing page.
2.8	**Snow/Ice Mask**	Not required.	As Class 1.	As Class 1.	The metadata indicates whether a pixel is assessed as being snow/ice or not. Information on snow/ice mask should be available in the metadata as a single DOI landing page.
2.9	**Terrain Shadow Mask**	Not required.	As Class 1.	As Class 1.	The metadata indicates pixels that are not directly illuminated due to terrain shadowing.
2.10	**Terrain Occlusion**	Not required.	As Class 1.	As Class 1.	The metadata indicates pixels that are not visible to the sensor due to terrain occlusion during off-nadir viewing.

2.11	**Solar and Viewing Geometry**	Provide average solar and sensor viewing azimuth and zenith angles.	As Class 1.	Provide per-pixel solar and sensor viewing azimuth and zenith angles.	As Class 3.
2.12	**Terrain Illumination Correction**	Not required.	As Class 1.	As Class 1.	Coefficients used for terrain illumination correction are provided for each pixel.
2.13	**Aerosol Optical Depth Parameters**	Not required.	As Class 1.	As Class 1.	To be determined.

3.3 Radiometric and Atmospheric Corrections

As described by the CARD4L specification, *Radiometric and Atmospheric Corrections* are:

"The following requirements must be met for all pixels in a collection. The requirements indicate both the necessary outcomes (3.1-3.3) and the minimum steps necessary to be deemed to have achieved those outcomes (3.4 onward). Radiometric corrections must lead to a valid measurement of surface reflectance."

Table 4 - Adaptation to CARD4L *Radiometric and Atmospheric Corrections* specification for application to commercial missions

#	Item	CLASS 1 High spatial resolution imagery	CLASS 2 Land Imaging	CLASS 3 Time-series observation	CLASS 4 Climate
3.1	**Measurement**	Processing to surface reflectance not required.	Pixel values that are expressed as a measurement of the Surface Reflectance of the land. This is a dimensionless value.	As Class 2.	Surface Reflectance measurements are SI traceable (see also 1.1).
3.2	**Measurement Uncertainty**	Not required. *Note 1: In current practice, users determine fitness for purpose based on knowledge of the lineage of the data, rather than on a specific estimate of measurement uncertainty.*	As Class 2.	An estimate of the uncertainty of the values is provided in measurement units. Although some evidence and methods may be proprietary and not necessarily made visible and fully accessible. *Note 1: This is a requirement for SI traceability. See also 1.1.* *Note 2: Information on measurement uncertainty should be available in the metadata as a single DOI landing page.*	Uncertainty information provided per pixel divided into different error-correlation components (e.g. random, systematic).

3.3	**Measurement Normalisation**	Not required.	As Class 1.	Measurements are normalised for solar and viewing conditions (i.e., nadir view angle and average solar angles). This may include terrain illumination and/or Bi-Directional Reflectance Function (BRDF) correction. *Note 1: Information on measurement normalisation should be available in the metadata as single DOI landing page.*	As Class 1.
3.4	**Directional Atmospheric Scattering - Molecular**	Not required.	Corrections are applied for molecular (Rayleigh) scattering.	Metadata contains a single DOI landing page with references to: • a citable peer-reviewed algorithm • technical documentation regarding the implementation of that algorithm • the sources of ancillary data used to make corrections *Note 1: Examples of technical documentation include an Algorithm Theoretical Basis Document, product user guide, etc.*	As Class 3.
3.5	**Directional Atmospheric Scattering - Aerosol**	Not required.	Corrections are applied for aerosol scattering.	Metadata contains a single DOI landing page with references to: • a citable peer-reviewed algorithm	As Class 3.

				• technical documentation regarding the implementation of that algorithm • the sources of ancillary data used to make corrections *Note 1: Examples of technical documentation include an Algorithm Theoretical Basis Document, product user guide, etc.*	
3.6	**Water Vapour Corrections**	Not required	Corrections are applied for water vapour.	Metadata contains a single DOI landing page with references to: • a citable peer-reviewed algorithm • technical documentation regarding the implementation of that algorithm *Note 1: Examples of technical documentation include an Algorithm Theoretical Basis Document, product user guide, etc.*	As C
3.7	**Ozone Corrections**	Not required	As Class 1.	As Class 1.	Data is corrected for ozone. Relevant metadata must be provided under 1.8 and 1.9. Metadata contains a single DOI landing page with references to: • a citable peer-reviewed algorithm

					• technical documentation regarding the implementation of that algorithm

3.4 Geometric Corrections

As described by the CARD4L specification, *Geometric Corrections* are:

"Geometric corrections must place the measurement accurately on the surface of the Earth (that is, geolocate the measurement) allowing measurements taken through time to be compared."

Table 5 - Adaptation to CARD4L Geometric Corrections specification for application to commercial missions

#	Item	CLASS 1 High spatial resolution imagery	CLASS 2 Land Imaging	CLASS 3 Time-series observation	CLASS 4 Climate
4.1	**Geometric Correction***	Not required.	Geometric corrections applied.	Metadata contains a single DOI landing page with references to: • a citable peer-reviewed algorithm • technical documentation regarding the implementation of that algorithm • the sources of ancillary data used to make corrections *Note 1: Examples of technical documentation include an Algorithm Theoretical Basis Document, product user guide, etc.*	As Class 3.

*Comment – here the quantitative requirements from CARD4L removed. It was unclear why this section was the only section with quantitative requirements, and they were set at a level particularly unrealistic for very high resolution missions and unnecessary or for climate missions.

4. MATURITY MATRIX CONCEPT FOR ARD

Maturity matrices are a simple, visual method for reporting evaluation results. They are effectively tables where each cell represents an aspect of the evaluation. Cells are colour-coded to report the results of the evaluation. Maturity matrices are increasingly widely used within EO, for example, the mission quality assessment framework developed in the EDAP project is based around the maturity matrix concept [RD-4].

A similar approach could be taken for reporting evaluations of the compliance of a particular satellite data product with an ARD standard. The obvious approach to apply this to the CARD4L framework is to include grid cells for each of the categories of specification criteria (i.e., general metadata, per-pixel metadata, radiometric and atmospheric corrections and geometric corrections), with a key that includes grades for full and partial compliance with the specification (see Table 6). Separate maturity matrices would be made for different product classes.

Table 6 - Maturity Matrix concept for ARD

ARD 4 [Product Class] Compliance			
General Metadata	Per-pixel Metadata	Radiometric & Atmospheric Corrections	Geometric Corrections

Key
Partial Compliance
Full Compliance