



# CEOS Analysis Ready Data Strategy

August 2019

CEOS SIT Technical Workshop 2019 Draft

## **CEOS Analysis Ready Data Strategy**

CEOS ensures international coordination of civil space-based Earth observation programs and promotes exchange of data to optimize societal benefit and inform decision making for securing a prosperous and sustainable future for humankind.

The rapidly evolving domain of information technology has led to greatly enhanced computational capability and much increased expectations of a broadened user community familiar with Big Data systems and the benefits of cloud storage and computing, machine learning, and artificial intelligence. Satellite data is now just one of many sources of information available to results-driven users and data hungry, automated systems.

Additionally, Earth observation missions are no longer the sole domain of government programs and space agencies. Miniaturization, standardization and innovation have led to significant new Earth observation and data access capabilities in the commercial 'Space 2.0' sector. This new presence poses a challenge for CEOS agencies to identify, and communicate, the distinct role for government funded missions and to adapt programs to avoid direct competition with the commercial sector.

Earth observation data from CEOS Agencies have several distinctive characteristics which ensure their ongoing relevance. Those features include:

- Highly superior platforms and instrumentation, allowing accuracy of observations that are not possible with systems that are developed with orders of magnitude less cost;
- Robust calibration and validation;
- Long-term data records and measurement continuity;
- Their potential to be free and open for all of society, allowing benefits to be delivered over many end users.

For these benefits to continue to be realised however requires that agencies adapt their methods of operation to ensure that data is easy to access – preferably in locations already favoured by the established IT and user community – and ready to use with minimal additional processing.

One of the measures proposed to support the objective of simplified data handling was the pursuit of CEOS standards for Analysis Ready Data (ARD) – which involves data suppliers removing many of the fundamental data correction and processing tasks from the users, so that more users and more uses of the data are possible.

There have been efforts in this direction from CEOS Agencies, both independently and through the LSI-VC and the development of the Product Family Specifications (PFS) for CEOS Analysis Ready Data for Land (CARD4L). However, noting the need for broad consensus on the evolution of this work, on the prioritisation of future products as the basis for new ARD standards in different thematic areas (including oceans, atmosphere, etc.), and on the extensive engagement of data suppliers, of Big Data hosts and aggregators, and of data users for maximum impact and benefit, CEOS has agreed the need for a top-down, coordinated strategy.

The Analysis Ready Data Strategy for CEOS is presented in the following pages, and is based around four pillars:

1. CEOS ARD User Needs & Specifications
2. Assured Production and Access
3. Pilots and Feedback
4. Communication & Promotion

#### Objectives

- a) to ensure continued competitive performance of public EO programme data and information, leveraging the availability of all relevant CEOS agency missions to meet user needs for information, amidst an explosion in provision of commercial smallsat constellations providing EO data;
- b) to remain flexible and adaptable to the evolving expectations of the user base, which is increasingly non-technical and more accustomed to simplicity in geospatial data sourcing, integration and application; thus to support the mainstreaming of EO data in society through removal of user burden and data complexity;
- c) to allow CEOS agency data to benefit from the increasing interaction with machines – through machine learning and artificial intelligence methods – which can handle large amounts of data, and allow these data to be integrated and interoperable with minimal intervention;
- d) to establish a broad understanding of, and participation in, CEOS efforts to define, produce, apply, and promote ARD in support of societal needs;
- e) demonstrate the benefit of ARD through pilots for a range of use cases, illustrating specific applications;
- f) to ensure effective engagement of the three key stakeholder groups: EO data providers (both public and private); Big Data hosts and aggregators who stage increasing amounts of CEOS Agency free and open data; and data users;
- g) identify the challenges faced by users in accessing and working with remote sensing data;
- h) to manage expectations of all stakeholders as to the status and outlook for ARD availability – so that all might plan and invest with confidence in capabilities to best exploit the CEOS Agency ARD;
- i) to establish priorities for which products and applications might follow on from the current CARD4L PFS – forming a rational and orderly queue for future efforts including ocean and atmosphere data products as appropriate;
- j) to ensure appropriate organisational responsibilities across the CEOS structure for the definition and execution of the way forward on ARD;

A well executed strategy will ensure CEOS Agency missions: remain competitive, increase the users and uses of data, and continue informing decision-making for a prosperous and sustainable future for humankind.

The Strategy is foreseen as an informal collection of ARD-related activities across the CEOS structure, with light oversight and coordinated reporting supported by SIT – as a headline topic

for the CSIRO/GA Chairmanship period of 2020-2021. Measures to sustain ongoing activity across the structure will be taken to ensure continuity beyond this period.

### **1: CEOS ARD User Needs & Specifications**

Specification development – to be carried out by CEOS thematic groups – is central to the CEOS ARD effort, however broad consensus is first required on the necessity and prioritisation of further CEOS ARD products.

<b>1.1: Continue development of the CARD4L Product Family Specifications</b>	<b>LSI-VC</b>	<b>Ongoing</b>
CEOS Analysis Ready Data for Land (CARD4L) have been developed in response to an identified need and are the exemplary case for the development of CEOS Analysis Ready Data for other domains. Development will continue in parallel, and should a broader, sustained CEOS ARD effort establish itself, the CARD4L work can be incorporated as a component.		

<b>1.2: Identify the need for and prioritise development of future target products as the basis for new CEOS ARD standards.</b>	<b>SIT Chair with VCs, LSI-VC support</b>	<b>2019 Onwards</b>
In the case of land surface imaging, ARD was identified as a need through the CEOS Future Data Architectures Study (e.g., greatly expanding non-expert user base) as well as interactions with user communities through SDCG-GFOI and the CEOS AHWG on GEOGLAM. The need for CEOS-shepherded ARD in other user communities is less clear, and so further effort is needed to establish this need – starting with a survey of the communities served by CEOS WGs, VCs and AHTs. The goal will be to establish priorities for which products and applications might follow on from the current CARD4L PFS – forming a rational and orderly queue for future CEOS efforts including ocean and atmosphere data products as appropriate.		

<b>1.3: Develop further CEOS ARD technical specifications based on established need and prioritisation.</b>	<b>CEOS thematic groups including VCs</b>	<b>Q4 2019 Onwards</b>
Following the example of LSI-VC with CARD4L, and based on the outcomes of task 1.2, CEOS thematic groups to initiate CEOS ARD specifications for their own domains.		

<b>1.4: CEOS Interoperability Terminology Report.</b>	<b>LSI-VC, WGCV (IVOS), WGISS, CEOS Agencies</b>	<b>Q2 2020</b>
<p>In the context of Earth remote sensing, the terms Analysis Ready Data (ARD), interoperability, and harmonization are often used and, to a large extent, used inconsistently. To rectify this, CEOS will produce a report discussing and providing examples of the following five terms drawing on work previously done by a subset of the LSI-VC, WGCV (IVOS), and WGISS:</p> <ol style="list-style-type: none"> <li>1. Analysis Ready Data (ARD)</li> <li>2. CEOS ARD for Land (CARD4L) Products</li> <li>3. Interoperable Products</li> <li>4. Harmonized Products</li> <li>5. Fused Products</li> </ol>		

## **2: Assured Production and Access**

Ensuring data discovery, access and integration is optimal – such that users can make the most of the CEOS ARD standards – will be critical. CEOS ARD efforts will be in vain if individual agency ARD, although nominally free and open, has to be accessed in different ways across different information systems by users looking to establish time and space coverage for their applications. Machine-friendly discovery and access will be part of this.

<b>2.1: Engage Big Data hosts and aggregators and establish formal pipelines and procedures to promote CEOS ARD hosting and uptake on their platforms.</b>	<b>SIT Chair</b>	<b>Q2 2020 Onwards</b>
<p>CEOS as a collective needs to engage in a coordinated fashion with Big Data hosts and aggregators that are already re-hosting publicly available free and open data to: streamline the movement of data to global platforms, improve consistency of data and provide some means for users to verify data provenance and quality, ensure CEOS Agencies are appropriately recognized and referenced as the data source, and to promote the uptake/rehosting of CEOS ARD specifications/datasets.</p>		

<b>2.2: Survey users' needs with regard to ARD accessibility and provide feedback to all data providers.</b>	<b>SIT Chair</b>	<b>Q2 2020 Onwards</b>
<p>Providing data in a form that is tailored to users is key to promoting uptake. This task will be done in parallel with <b>2.1</b>, and will be iterative. Consider developing an inventory of existing data services that may address user needs in terms of ARD delivery or that may need to be developed where there are currently gaps.</p>		

<b>2.3: CEOS Position Paper on the Interplay of Industry and CEOS ARD.</b>	<b>SIT Chair, LSI-VC, WGISS</b>	<b>Q2 2020</b>
<p>This paper will look into the different roles industry may play (e.g., as data users, data providers, data hosts, and as providers of processing chains to produce data that meet CEOS ARD specifications) and seek to establish a CEOS position on each aspect.</p> <p>Key considerations will be machine-to-machine communication (including the role of COG and STAC technologies), machine learning, automation, and analysis platforms. The ever-growing size of EO datasets is necessitating the removal of human interaction in the processing and analysis loop. A number of 'Space 2.0' companies' strategies are built entirely on this concept. Machine-to-machine communication, machine learning methods, automation and AI are the way forward in extracting useful data from these huge datasets. WGISS would be best placed to undertake this part of the study, which should result in recommendations for CEOS in regards to these technologies and ARD, as well as recommended updates to the CARD4L PFS.</p>		

### **3: Pilots and Feedback**

Beyond the paper documents, practical experience in the production, provision and application of CEOS ARD will be essential to ensure that it is fit for purpose for the various communities and applications that we seek to serve.

<b>3.1: Production of CEOS ARD and supply to data aggregators and platforms.</b>	<b>CEOS Agencies, VCs</b>	<b>Q3 2019 Onwards</b>
<p>CEOS Agencies to commit to producing CEOS ARD as a standard product for their missions, and ensuring this data is easily accessible by the public, both through in-house, proprietary access portals but also through redistribution by third-party global IT platforms. This will be done in coordination with <b>2.1</b>.</p>		

<b>3.2: Open Data Cube/Digital Earth Africa Pilot</b>	<b>GA, SEO</b>	<b>Q3 2019 Onwards</b>
<p>Open Data Cube (ODC) is just one data analysis platform that might be used for trials of CEOS ARD on a large scale. ODC is a free-and-open option and the Digital Earth Africa instance in particular is sustainably funded and already involves a number of CEOS Agencies and the SEO. CEOS Agencies can work with Geoscience Australia, the SEO, and others to capitalise on this initiative to trial CEOS ARD supply and access. Lessons learned can be used to improve the ARD specifications and also the data platforms/delivery methods.</p>		

<b>3.3: GFOI Pilot</b>	<b>SDCG</b>	<b>Q4 2019 Onwards</b>
<p>SDCG-GFOI to coordinate ARD supply and access trials for select users in the GFOI/forest monitoring community. Lessons learned can be used to improve the ARD specifications and also the data platforms/delivery methods. There is scope for this to be undertaken with <b>3.2</b> using Digital Earth Africa as the data platform.</p>		

<b>3.4: GEOGLAM Pilot</b>	<b>CEOS-GEOGLAM</b>	<b>Q4 2019 Onwards</b>
<p>CEOS AHWG for GEOGLAM to coordinate ARD supply and access trials for select users in the GEOGLAM/JECAM agricultural monitoring communities. Lessons learned can be used to improve the ARD specifications and also the data platforms/delivery methods. There is scope for this to be undertaken with <b>3.2</b> using Digital Earth Africa as the data platform.</p>		

<b>3.5: Mekong Data Cube Pilot</b>	<b>CEOS Agencies, VNSC</b>	<b>Q4 2019 Onwards</b>
<p>Building on the effort started through the 2019 CEOS Chair Initiative, CEOS will continue working with VNSC to conduct CEOS ARD supply and access trials for users in the Mekong region using the Mekong Data Cube as the data supply, access, and analysis platform. Lessons learned can be used to improve the ARD specifications and also the data platforms/delivery methods. The CEOS Chair for 2020 (ISRO) has signalled intention to continue Data Cube studies in the region and this may be supportive.</p>		

<b>3.6: Other Pilots</b>	<b>CEOS Agencies, COVERAGE</b>	<b>2020 Onwards</b>
<p>Commit to support ARD pilots for new product families (e.g., Geostationary, LIDAR, Inland and Coastal Water Surface Reflectance, Others) once the Product Family Specifications are developed. This task will be based on the outcomes of task <b>1.3</b>.</p>		

#### **4: Communication & Promotion**

CEOS needs to communicate and promote its ARD efforts to data providers (internally and externally), data hosts, and data users. This communication must involve engaging with and showing leadership in the broader community, including the private sector and the scientific data community.

<b>4.1: GEO showcase and side event: Analysis Ready Data</b>	<b>GA, SIT Chair Team</b>	<b>Q4 2019</b>
This event will build awareness of ARD as a critical Earth observation infrastructure element. It will celebrate key achievements (including those of CEOS) and build high-level support for continued effort and increased investment in these areas at a critical time. The event will identify cross-sector synergies and explore the emerging challenge of hosting, and making accessible, growing international collections of ARD.		

<b>4.2: CEOS–Industry ARD Workshop</b>	<b>SIT Chair Team, CEOS Agencies</b>	<b>Q4 2020</b>
CEOS Agencies commit to organise a dedicated CEOS ARD workshop, inviting agencies, researchers and the private sector to continue engagement between EO data providers (public and commercial), big data hosts/aggregators, and data users.		

<b>4.3: CEOS ARD stocktake and outlook</b>	<b>LSI-VC &amp; SIT Chair</b>	<b>Q3 2019 Onwards</b>
As with all standards, a critical mass of participation is essential to achieve success. Achievement of that participation will require careful management of expectations of the agencies providing and processing data, of their partners and corporations who are hosting and aggregating the data, and of the users we all seek to serve. No group is likely to invest without the confidence that the critical mass will be achieved and the benefits realised in exchange for the additional cost. To allow investment with confidence, CEOS will define and maintain a clear statement as to the current and future availability of the different datasets produced to its ARD standards, and how to access them. This will include a current snapshot and forecast for 1, 2, and 3 years hence. LSI-VC-5 in Tokyo in February 2018 agreed an action to produce such a stocktake and outlook for the CARD4L product family. CEOS can build on this foundation from LSI-VC.		



<b>4.4: Promotion of CEOS ARD to standards organisations</b>	<b>CEOS Agencies, SIT Chair</b>	<b>Q1 2021 Onwards</b>
<p>CEOS to progress CARD4L discussions with standards organisations (e.g., OGC) to explore whether CEOS ARD specifications might be used as the basis for broader, official community standards, and, to ensure that CEOS work is recognised by others including the data research community.</p>		

<b>4.5: Promotion of CEOS ARD to data providers</b>	<b>SIT Chair, VCs</b>	<b>Q4 2019 Onwards</b>
<p>CEOS to work in a coordinated fashion to promote the uptake of CEOS ARD specifications by data providers (both CEOS Agencies and commercial entities), encourage the production of CEOS ARD as a standard output from missions, and encourage data providers to ensure that their data is pushed efficiently to data redistributors for ease of access. The CEOS LSI-VC has developed <a href="#">informational resources</a> for CARD4L that can be used as a foundation.</p>		

<b>4.6: Communication with CEOS ARD users</b>	<b>WGCapD, AHTs, CEOS Agencies</b>	<b>Q4 2019 Onwards</b>
<p>CEOS to work in a coordinated fashion to promote the uptake of CEOS ARD by data users, and to gather user experiences with CEOS ARD. This will be done through the networks of WGCapD, the AHTs, individual agencies, and the user bases of data aggregators/hosts. Targeted studies may also be conducted. The CEOS LSI-VC has developed <a href="#">informational resources</a> for CARD4L that can be used as a foundation.</p>		