**A CEOS STRATEGY FOR ANALYSIS READY DATA (ARD)**

**DISCUSSION PAPER FOR SIT-34**

 Version 0.2, 11th Mar 2019

1. **Introduction**

The CEOS Future Data Architectures (FDA) study team addressed a broad range of issues impacting public Earth observation programmes, particularly in relation to the data and information systems linking data holdings to users and their applications. The Team’s report included an excellent review of the changing landscape and context for CEOS agency EO programmes and the way in which they service their user communities.

Of all the concerns raised by the many CEOS agencies participating in the FDA study activity, the most common was the need for significant simplification of data handling and uptake by users, and the changing expectations of a much-broadened user community familiar with Big Data systems and the benefits of cloud storage and computing. One of the measures proposed to support the objective of simplified data handling was the pursuit of CEOS standards for *Analysis Ready Data* – 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.

The Land Surface Imaging Virtual Constellation (LSI-VC) has taken on the responsibility for the first CEOS coordination in this area through the development of the Product Family Specifications (PFS) for the *CEOS Analysis Ready Data for Land (CARD4L)* set of products. 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, SIT Vice-Chair (CSIRO/GA) has initiated this short discussion paper ahead of the SIT-34 meeting - in order to stimulate discussion on the need for, and nature of, a *CEOS Strategy for Analysis Ready Data.*

1. **Objectives**

The proposed objectives for a CEOS Strategy for Analysis Ready Data are:

1. 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;
2. to meet the changing 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;
3. 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;
4. to establish a broad understanding of, and participation in, CEOS efforts to define, produce, apply, and promote ARD in support of societal needs;
5. to ensure effective engagement of the three key stakeholder groups: EO data providers (both public and private); Big Data hosters and aggregators who stage increasing amounts of CEOS agency free and open data; and data users;
6. 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;
7. 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;
8. to ensure appropriate organisational responsibilities across the CEOS structure for the definition and execution of the way forward on ARD; the *Strategy* is seen as a fairly informal grouping of relevant CEOS activities which SIT Vice Chair proposes to oversee and promote during their coming SIT Chair term – with new tasks identified as needed.
9. **CEOS ARD Strategy Components**

We foresee a number of activities as essential to a comprehensive and successful strategy on ARD for CEOS, identified below.

1. **CARD4L**

As the first CEOS steps towards ARD capabilities, CARD4L will include many of the activities identified in this section, including: technical specifications; trial production of data; pilot programmes for stakeholder engagement and feedback; and continued refinement of the standards.

1. **ARD Stocktake and outlook**

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 should 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 should include a current snapshot and forecast for 1, 2, and 3 years hence.

LSI-VC-5 in Tokyo in Feb 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.

1. **Technical Specification Development and Maintenance**

The first steps for any new ARD standards across CEOS will be the development of the Product Family Specification as has been undertaken for CARD4L for 3 initial products:

* Surface Reflectance (optical);
* Surface Temperature (optical);
* Backscatter (Synthetic Aperture Radar).

As specified in the documentation:

*CARD4L products 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. These products would be resampled onto a common geometric grid (for a given product) and would provide baseline data for further interoperability both through time and with other datasets.*

*CARD4L products are intended to be flexible and accessible products suitable for a wide range of users for a wide variety of applications, including particularly time series analysis and multi-sensor application development. They are also intended to support rapid ingestion and exploitation via high-performance computing, cloud computing and other future data architectures. They may not be suitable for all purposes, and are not intended as a ‘replacement’ for other types of satellite products.*

LSI-VC has specified that the CARD4L branding is applied to a particular product once:

* *that product has been assessed as meeting CARD4L requirements by the CEOS agency responsible for production and distribution of the product;*
* *that assessment has been peer reviewed by the CEOS Land Surface Imaging Virtual Constellation in consultation with the CEOS Working Group on Calibration and Validation.*

And that *Agencies or other entities considering undertaking an assessment process should contact the co-leads of the Land Surface Imaging Virtual Constellation.*

A product can continue to use CARD4L branding as long as its generation and distribution remain consistent with the peer-reviewed assessment.

Assuming that the CARD4L model and process is deemed to be effective by CEOS and suitable for replication to other products and disciplines, a systematic and effective process will be required to ensure consistency and performance across the relevant standards and datasets. These areas may be broad thematic measurement categories – such as oceans or atmosphere – or might be specific applications such as agriculture or forest monitoring. They might also be ARD in support of increased machine access and learning for CEOS agency data.

The interest and capacity of the various thematic communities represented within CEOS (across different VCs and AHTs) will be the rate-determining step for prioritisation of product groups to follow in the steps of CARD4L. But it will be helpful for SIT Chair to ensure an orderly queue of effort – in particular if WGCV or others are limited in their capacity to serve a supporting role in the assessment process.

1. **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. In the case of CARD4L, LSI-VC has an action underway to establish reference groups within the GEO Flagships of GFOI and GEOGLAM to ensure diverse experience and opinion is fed back into the evolution of the standards to keep pace with these communities.

In addition, a number of pilots related to the Open Data Cube (in Colombia, Vietnam, and elsewhere) are already yielding useful experience as to future evolution for CARD4L. Equivalent effort and engagement will be essential for any CEOS standards to thrive in the wild.

An important driver in these trials will be to ensure data discovery, access and integration is optimal - such that the user can make the most of the CEOS ARD standards. Our 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.

1. **Promotion**

Some CEOS agencies are already seeing the benefit of their migration to provide ARD products. USGS has reported, for example, that uptake of their Landsat-8 SR product is outpacing uptake of the equivalent level 1 product by a ratio of about 3:1. But each agency has different mandates, different circumstances and different constraints and must be convinced as to the cost-benefit of the additional production overhead to supply ARD. Active promotion of the benefits of ARD and experience gained by the pioneering agencies will be an essential part of achieving that critical mass for the standards to succeed. This must include:

* data providers - both CEOS space agencies and private EO data providers; eg, at the recent GFOI Plenary in Colombia, Planet indicated interest in having their SR product evaluated with a view to being identified as meeting CARD4L standards; the first priority will be CEOS agencies but the adjunct benefits of an even broader standardisation and interoperability by having private providers seek to comply to the CEOS standard should not be over-looked or under-estimated;
* data hosters and aggregators: the so-called Data Giants already stage substantial amounts of free and open data such as Landsat and Sentinel-1; such players invest substantial resources into the pre-processing and provision of that data; CEOS can help such organisations distribute CEOS agency data through their participation in standardisation efforts – efforts which will further simplify their task in providing data to their user base in ways which they consider essential to its appeal;
* data users as the ultimate consumers of the embedded information and whose needs must be addressed if the concept of standardisation for ease of use is to be truly successful.

LSI has already prepared outreach materials targeted at each of these communities for CARD4L. CEOS can build on these first steps as feedback and need dictates.

1. **Next steps**

SIT Vice-Chair has prepared this short paper to support discussion on the topic of a CEOS ARD strategy at the SIT-34 meeting in Miami in April 2019. 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.

Assuming a positive response from Principals at SIT-34, SIT Vice Chair Team will bring a fleshed-out *Strategy* and tasks to the SIT TW in September, and a refined version to the CEOS Plenary in Vietnam in October - where CSIRO/GA will assume the SIT Chair role. The team will consult across the CEOS structure in the process, in order to determine interests and priorities of CEOS agencies and groups. A short questionnaire has already been provided to VCs, WGs, and AHTs and all are encouraged to provide a reply if they have not already. We look forward to the discussion at SIT-34 and feedback is welcome at any stage.