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

**DISCUSSION PAPER**

 Version 0.1, 22 Mar 2018

**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, and on the extensive engagement of data suppliers, of Big Data hosters and aggregators, and of data users for maximum impact and benefit, LSI-VC has initiated this short discussion paper ahead of the SIT-33 meeting - in order to stimulate discussion on the need for, and nature of, a *CEOS Strategy for Analysis Ready Data.*

**Objectives**

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

* to ensure continued competitiveness of public EO programme data and information, leveraging the availability of all relevant CEOS agency missions to meet user needs for information;
* 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;
* to establish a broad understanding of, and participation in, CEOS efforts to define, produce, and apply ARD in support of societal needs;
* 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;
* 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;
* to establish priorities for which products and applications might follow on from the current CARD4L PFS;
* to ensure appropriate organisational responsibilities across the CEOS structure for the definition and execution of the way forward on ARD.

 **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 might consider this example as a model for further study.

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.

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.

1. **Promotion**

Some CEOS agencies are already seeing the benefit of their migration to provide ARD products. USGS reported to LSI-VC-5 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 like Google and Amazon 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-VC-5 agreed to prepare sample outreach flyers targeted at each of these communities for CARD4L. CEOS can build on these first steps as feedback and need dictates.

**Others?**

**Next steps**

LSI-VC has taken this initiative as part of its own efforts to progress the CARD4L activity and to help establish a common understanding and planning within CEOS on a broader context for CARD4L and for the strategy on ARD within CEOS. The discussion paper will be shared with CEOS agencies ahead of the SIT-33 meeting and referred to in the LSI SIT-33 report with an invitation for all CEOS groups and agencies to comment and contribute to the refinement of the ideas herein. Ideally, CEOS might agree actions to ensure appropriate organisational responsibilities across the CEOS structure for the definition and execution of the way forward on ARD.