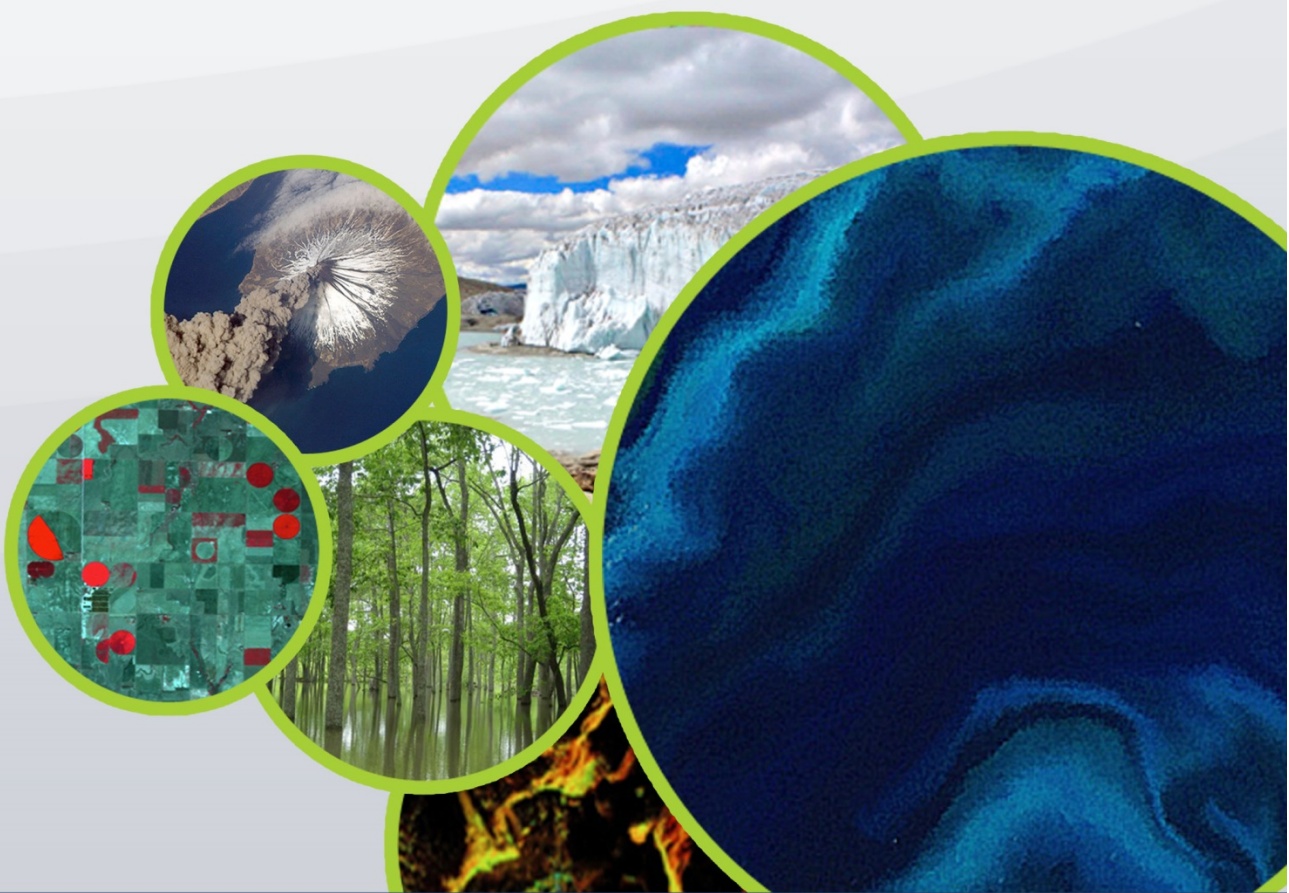




Committee on Earth Observation Satellites



Future Data Access & Analysis Architectures Strategy for CEOS

Discussion Document for the 31st CEOS Plenary

Version 1.0, October 2017

A Future Data Access & Analysis Architectures Strategy for CEOS

Discussion Document for the 31st CEOS Plenary

Version 1.0, October 2017

EXECUTIVE SUMMARY

CEOS agencies have spent billions acquiring Earth Observation (EO) data and developing ground segment architectures to make these data available to users. To realise the full societal benefits of these investments, however, users must be able to access and utilise EO data with maximum efficiency.

A clear, shared vision of Future Data Access & Analysis Architectures (FDA) for EO is essential to enhance and encourage the use of EO data for science, operational services, and commercial applications. Joint initiatives undertaken by CEOS agencies with a common understanding of FDA will have a significant impact by more quickly and effectively addressing users' increasing, and increasingly diverse, needs.

CEOS agencies must be aware that:

- The EO user base is expanding rapidly, and includes a growing number of non-experts whose main interest is the easy transformation of data into information.
- Data volumes and diversity are also increasing rapidly, which requires alternatives to traditional data download dissemination schemes. These alternatives include coordinated efforts to develop easy-to-use data analysis platforms and tools, such as Data Cubes, by which users can access processed and organised data for the transformation of Analysis Ready Data (ARD) into relevant information.
- Data platforms can co-exist or interact with an increasing number of EO data service environments maintained by large commercial cloud providers that have enormous capacities for data storage and processing.
- Only full commitment to user needs and expectations can assure that CEOS agency initiatives will be successful. Therefore, user acceptance and user behavior need to be rigorously measured in any FDA activity.
- In addition to providing ARD—including optical imagery and other data types—CEOS agencies can facilitate greater use of EO data by applying free and open principles for easy tailoring of software as well as coordinating and prioritising initiatives for standardisation and interoperability between similar sensors.

The FDA Ad Hoc Team (AHT) proposes that CEOS agencies implement a number of activities that reflect the insights listed above, and that fall into five core initiative areas, which this document discusses in detail. The five initiatives are:

1. CEOS Analysis Ready Data (ARD)
2. Interoperable Free and Open Tools
3. Data, Processing, and Architecture Interface Standards
4. Analytical Processing Capabilities
5. User Metrics

The outcome of these initiatives should be a comprehensive and well stewarded EO data offering, against which organisations can confidently invest their resources and leverage their distribution channels to users in an effort to realise benefits that go far beyond traditional applications.

INTRODUCTION

The CEOS Future Data Access & Analysis Architectures (FDA) Ad Hoc Team (AHT) has prepared this document for discussion and endorsement by the 31st CEOS Plenary in October 2017. It outlines a strategy for a FDA approach for CEOS agencies. Consecutive CEOS Chair agencies—the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in 2016, U.S. Geological Survey (USGS) in 2017, and European Commission (EC) in 2018—have identified this topic as a priority for space agencies and for CEOS. The FDA strategy for CEOS addresses:

- The need for new paradigms in information systems to deal with big data (i.e., increased data volume, velocity, and variety) in Earth Observation (EO);
- The need to integrate different satellite EO datasets;
- The opportunities provided by advances in high performance computing and in cloud storage and processing systems that are changing user community expectations for accessing and using EO data.

This document also supports keeping the FDA AHT in place for the foreseeable future as it works to integrate the activities of several key entities within CEOS and sustain the momentum that has been built to date. These key entities are the Land Surface Imaging Virtual Constellation (LSI-VC), Systems Engineering Office (SEO), Working Group on Information Systems and Services (WGISS), and Working Group on Capacity Building and Data Democracy (WGCapD). However, in order for this FDA strategy to be truly effective and have the desired impact, it is critical that all CEOS agencies actively support the principles and activities outlined below.

CONTEXT AND STATUS

CEOS Principals are referred to the comprehensive October 2016 report, *CEOS Future Data Access and Analysis Architectures Study*, prepared by the FDA AHT. It provides a thorough review of the context and evolution of user requirements and technical possibilities. The report also suggests that government-sponsored satellite EO is at a key juncture, where observation techniques have advanced to the point where they have significant potential for addressing society's grand challenges while the number and diversity of data users has increased dramatically. EO data users are increasingly non-technical, expect to easily access and use the data, and require that information derived from EO data be easy to integrate with other information sources. Notwithstanding the progress made in recent years, difficulties in finding and manipulating EO data still represent a considerable barrier to realising their full potential and to properly harnessing EO big data for its societal benefits.

The AHT has continued the analysis through 2017 to prepare final recommendations on the way forward for a decision at the 31st CEOS Plenary. In parallel, a number of recommendations from the interim report have been progressed by CEOS groups:

1. SEO-led pilot projects to elicit real-world experience of novel architectures, including the CEOS Data Cube (CDC), to provide input for further strategic decisions;
2. LSI-VC definition of CEOS standards for Analysis Ready Data (ARD) using CEOS ARD for Land (CARD4L) as the pathfinder;
3. LSI-VC progress on the Moderate Resolution Sensor Interoperability (MRI) initiative, which both proposed an interoperability framework and conducted a case study utilising a merged Landsat and Sentinel-2 dataset;
4. WGISS continuation of work on discovery search engine optimisation, common standards for interoperability of product formats and application programme interface (API) for analytics

and data access services, and exploration of emerging big data services, including cloud computing.

The 2017 analysis by the FDA AHT has focused on defining the substance for a collaborative strategy within CEOS, and how this would best be accomplished within the CEOS Work Plan and structure. This included a survey of CEOS agencies and groups to establish a common set of priorities from the membership. The most common responses and requests that were suggested as ways to facilitate user uptake for inter-agency collaboration through CEOS were:

- Simplification of EO data uptake and application for many users;
- Definition and automated delivery of interoperable ARD (e.g., CARD4L);
- Lessons learned on the use of cloud computing for storage and processing (e.g., exploitation platforms and Data Cubes);
- Algorithm portability;
- Solutions that enable scalability, from local (e.g., institutional) to global;
- Complementarity between free and open tools and commercial products.

PROPOSED STRATEGY

The 2016 report recommended that the FDA strategy for CEOS address a number of key objectives, including de-risking and simplifying EO data for users, enabling users to make use of ALL available and relevant CEOS agency data, and supporting CEOS ambitions in relation to its chosen grand challenges. The resulting FDA strategy that the AHT now proposes for CEOS centers around five core initiative areas, introduced here and explained in more detail in the next section:

1. CEOS Analysis Ready Data (ARD)
 - Implementation of CARD4L, noting the critical need for progress in the land domain, which is less advanced than marine and atmosphere domains;
 - A comprehensive CEOS strategy for ARD, incorporating coastal, marine, and atmospheric domains in a manner complementary to Coordination Group for Meteorological Satellites (CGMS) and World Meteorological Organization (WMO) activities.
2. Interoperable Free and Open Tools
 - Provision of interoperable free and open tools to exploit advances in technology and meet user demand, including increasing contributions to the CEOS Data Cube (CDC) initiative and utilising data platform infrastructures;
 - Ongoing programme of 'pilot projects' for new technologies, undertaken at low-risk/low-cost, to ensure feedback to CEOS agencies on real users' requirements, providing Principals with real-world evidence as to what suitable future projects might be.
3. Data, Processing, and Architecture Interface Standards
 - Standards for portability of 'algorithms' that exploit satellite EO data between different exploitation platforms, whatever technologies, whether proprietary or open, that they are based on.
4. Analytical Processing Capabilities
 - Implementation of modular and easy-to-use analytical processing capabilities in full computing environments for time series and other analysis.
5. User Metrics
 - A data use metrics framework through which agencies can contribute to a 'sector wide' view of how EO data is actually being used, not just how much is being downloaded.

INITIATIVE 1: CEOS ANALYSIS READY DATA (ARD)

A comprehensive strategy for definition and production of ARD standards by CEOS agencies is a fundamental part of the proposed way forward for CEOS, as ARD product generation becomes more of a core business for CEOS agencies and as technology and required ancillary data sources become more widely available. CEOS ARD addresses the need to remove technical obstacles for users to access and apply data, and to ensure that *every available pixel counts*, such that EO data from all space agencies, large and small, might reach full potential.

This is an obvious standardisation and coordination role for CEOS and responds to the overwhelming priority for CEOS action voiced by agencies in the FDA AHT survey. Multiple CEOS agencies have already initiated, or are on the cusp of defining, standards for ARD and it is of paramount importance, in order to maximise the interoperability of the next generation of missions and information systems, that CEOS act decisively and quickly to establish CEOS ARD standards in the context of a broad ARD strategy. Based on feedback from members to date, we anticipate strong support for and participation in this standardisation effort. CEOS agency ARD product generation is intended to be additive and is not intended to preclude other individual agency product offerings.

Different philosophies may emerge for the routine production and provision of ARD in agency product pipelines, including ‘on-the-fly’ product generation; but all will benefit from agreement on the minimum standards for fundamental corrections and formats so that space agencies of all sizes, and users of all kinds, might participate in the provision or uptake of data to the common standard. This will greatly facilitate space agency contribution to the grand challenges CEOS seeks to address—such as the United Nations Sustainable Development Goals (SDGs), disasters, and climate—and support the ambitions for mainstreaming EO into society among varied and non-technical users. Further adoption of this strategy facilitates integration of data for public good, other CEOS agency data, and commercial data all working together.

The CEOS ARD strategy will build upon the work already underway within LSI-VC and embed it within a broader strategy and schedule of relevance to all CEOS missions and user communities. The CARD4L standards drafted by LSI-VC will serve as a pathfinder with which to secure significant user feedback (Figure 1). The costs involved in agencies adapting their production pipelines to include CEOS ARD standard products are significant; these costs will not be taken lightly, and certainly not without substantial evidence of the cost-benefit to individual agencies as well as the collective cost-benefit resulting from the increased uptake of EO data.

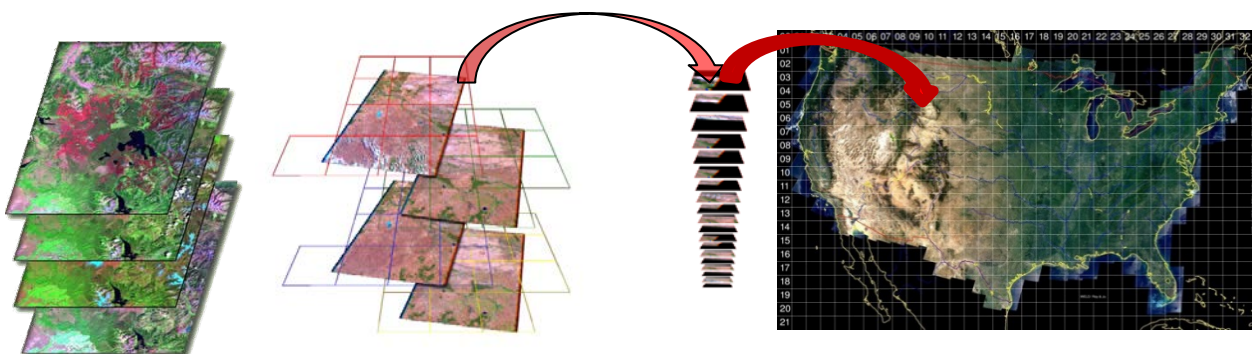


Figure 1. CEOS Analysis Ready Data (CARD4L)-compliant products improve data interoperability and enable efficient time series analysis. (Image credit: SDSU WELD, Roy et al., 2010)

Consequently, the FDA AHT proposes the following activities as the basis for this CEOS ARD initiative:

1. Identify candidate CARD4L-compliant optical data product providers. Engage agencies to encourage development and routine optical product generation, and/or sustained provision of toolsets that enable generation of such products. (LSI-VC)

2. Identify candidate CARD4L-compliant SAR data product providers. Contact agencies to encourage development and routine SAR product generation and/or sustained provision of toolsets that enable the generation of such products. (LSI-VC). Alternatively, such products could be generated on-the-fly through provided free and open tools.
3. Make ARD products or production capability available (e.g., on-the-fly and/or pre-processed). Develop a summary of ARD production status and implementation options. (LSI-VC)
4. Utilising the LSI-VC CARD4L approach, engage ocean and atmosphere VCs to evaluate the utility of similar CARD4O and CARD4A Definitions and Specifications. (LSI-VC)

To further this CEOS ARD initiative, CEOS agencies are asked to contribute in the following ways:

- Develop CARD4L compliant optical and/or SAR products;
- Produce ARD data and/or provide ARD production tools;
- Examine ARD for the ocean and atmosphere domains;
- Participate in the standardisation and review processes (including user feedback loops for evolution);
- Engage through the CEOS Working Group on Calibration & Validation (WGCV) to define Quality Assurance (QA) protocols and cross-validation projects across ARD products;
- Reach out to commercial providers to obtain their participation in ARD (at their cost);
- Promote and enable discoverability of ARD datasets.

INITIATIVE 2: INTEROPERABLE FREE AND OPEN TOOLS

Each CEOS agency will continue to develop its data and computational infrastructures consistent with its capacity and user service mandates. CEOS has a role in identifying tools to support complementarity and interoperability across CEOS agencies in support of the FDA strategy objectives of:

- De-risking and simplification of EO data for users;
- Allowing users to make use of all available and relevant CEOS agency data;
- Supporting CEOS ambitions in relation to its chosen grand challenges.

Coordination among CEOS agencies has already led to the adoption of the first of these tools and to its promotion as an open source consortium: the Open Data Cube (ODC) (Figure 2). The CEOS SEO stewards its implementation within CEOS, in the form of the CEOS Data Cube (CDC), supported by USGS, CSIRO, Geoscience Australia (GA) and others, as a tool for the realisation of CEOS ambitions in relation to the grand challenges and Group on Earth Observations (GEO) flagships. The CDC seeks to simplify the access and use of EO data for users, and to provide a platform whereby all space agencies conforming to the CEOS ARD standards can contribute their data and application algorithms.



Figure 2. The Open Data Cube (ODC) consortium is targeting a formal release (Version 1) in February 2018, with the initial goals of establishing 1) an end-to-end architecture that allows user deployment and 2) a core set of documentation and training tools. (Image credit: <https://www.opendatacube.org/>)

The CDC is an important open source tool for development under this FDA strategy. A considerable suite of EO toolboxes, multi-mission tools for data processing, and data product readers and writers as well as visualisation and analysis tools have become available under open source conditions. Pilot implementations of the CDC are underway in support of the Global Forest Observations Initiative (GFOI), GEO Global Agricultural Monitoring Initiative (GEOGLAM), and other CEOS priorities. A three-year Work Plan has already been defined and will continually be reviewed and revised to ensure consistency with this proposed CEOS FDA strategy. User feedback activities will be prominent in support of the user metrics activity. Additionally, individual (or groups of) CEOS agencies are anticipated to propose further tools as their respective plans for evolution of their data and computing infrastructure and opportunities for collaboration emerge.

The FDA AHT proposes the following activities as the basis for this Interoperable Free and Open Tools initiative:

1. Further develop and communicate CEOS Data Cube governance model. Seek broad awareness and acceptance. (SEO)
2. Develop strategies for local, national, regional, and global scale implementations, ensuring maximum interoperability. (SEO)
3. Prepare a long-term plan of CEOS Data Cube sustainment. (SEO)
4. Develop 'self-help' materials on CDC implementation and testing to help establish new Data Cubes that are consistent and interoperable. (SEO)
5. Demonstrate stakeholder promotion and adoption of ODC elements in global initiatives. (SEO)
6. Promote the adoption of free and open tools, algorithms, and technologies in support of the goal of interoperability. (SEO)
7. Establish 'FDA Setup Support Scheme' whereby agencies can support projects that wish to get a Data Cube or some other future data architecture set up, to include some basic data, processing time, and storage. (SEO)

To further this Interoperable Free and Open Tools initiative, CEOS agencies are asked to contribute in the following ways:

- Promote the use of interoperable free and open tools and application algorithms to enhance the use and impact of CEOS satellite data;
- Continue to support the CEOS Data Cube (CDC) initiative and progress the development of core components, supporting documentation, and training materials;
- Support the initial deployment of the CDC architecture by providing training and capacity building to interested international users;
- Contribute application algorithms to the CDC repository along with documentation and case studies.

INITIATIVE 3: DATA, PROCESSING, AND ARCHITECTURE INTERFACE STANDARDS

As CEOS agencies are defining their processing and data dissemination standards, they seek to apply and follow international standards, including those generated by the CEOS Working Group on Information Systems and Services (WGISS). This does not only concern common standards in terms of catalogs, metadata, terminology, and semantics, but it also involves interoperability standards for data discovery and download, as well as common interface standards such as INSPIRE, OGC, and W3C, and interoperability with other data access services (e.g., European Data Portal, international, GEOSS). One of the primary FDA concerns is the portability of products and algorithms between the various future data architectures, including the CEOS Data Cube and other agency-specific architectures. This ensures that users are not locked into a single platform as new architectures emerge.

The FDA AHT proposes the following activities as the basis for this Data, Processing, and Architecture Interface Standards initiative:

1. Review standards among different FDA approaches to promote interoperability of tools. Approaches might include Data Cube, standard API(s), as well as other FDA approaches. (WGISS)
2. Update CEOS data discovery and download standards, moving from discovery of files to discovery and download of pixel level data across multiple sensors as user-defined ARD packages. (WGISS)
3. Promote awareness of new FDA paradigms and systems. (WGISS/WGCapD)

To further this Data, Processing, and Architecture Interface Standards initiative, CEOS agencies are asked to contribute in the following ways:

- Identify key data and metadata standards experts to join WGISS;
- Support WGISS in the development of standards that ensure interoperability among one or more FDA platforms;
- Support prototype testing of data and application standards to ensure successful implementation;
- Identify specialists to join WGISS with expertise in their data holdings to provide guidance on how their holdings are changing to accommodate pixel level access;
- Identify key system engineers, applications liaisons, and communication/outreach liaisons to engage with WGCapD in developing strategies to promote FDA paradigms and systems.

INITIATIVE 4: ANALYTICAL PROCESSING CAPABILITIES

Future data exploitation environments should provide a variety of functionalities that allow analytical data processing, including:

- A full computing environment allowing users to process data and information with high efficiency (i.e., CPU, memory, storage, bandwidth, access to other data);
- Easy-to-use programming environments (based on free and open software, if possible) with tools to perform generic operations (e.g., Python, R);
- Implementation of modular service tools, allowing multiple users to invoke, use, add, and 'chain' the tools and functionalities as per user needs;
- Algorithm implementation;
- Time series analysis (Figure 3);
- Access to tool libraries from third party providers;
- Data generation service that allows generation of sub-sets of products per user specifications of area of interest and format;
- Projection tools, allowing users to convert data into different projection systems.

The FDA AHT proposes the following activities as the basis for this Analytical Processing Capabilities initiative:

1. Develop and review a best practice white paper (i.e., reference architecture) for EO data analysis pipelines involving CEOS agencies and third-party systems (e.g., cloud, commercial providers) from a user analysis perspective with an emphasis on global agendas. (WGISS)
2. Develop and review a best practice white paper on interoperable EO data analysis APIs (i.e., how can a user-supplied algorithm be made portable between different EO data analysis platforms? How can such analysis packages be made discoverable and shareable?). (WGISS)
3. Develop a best practice single-sign-on (SSO) authentication white paper to support machine-to-machine authentication for analysis services. (WGISS)

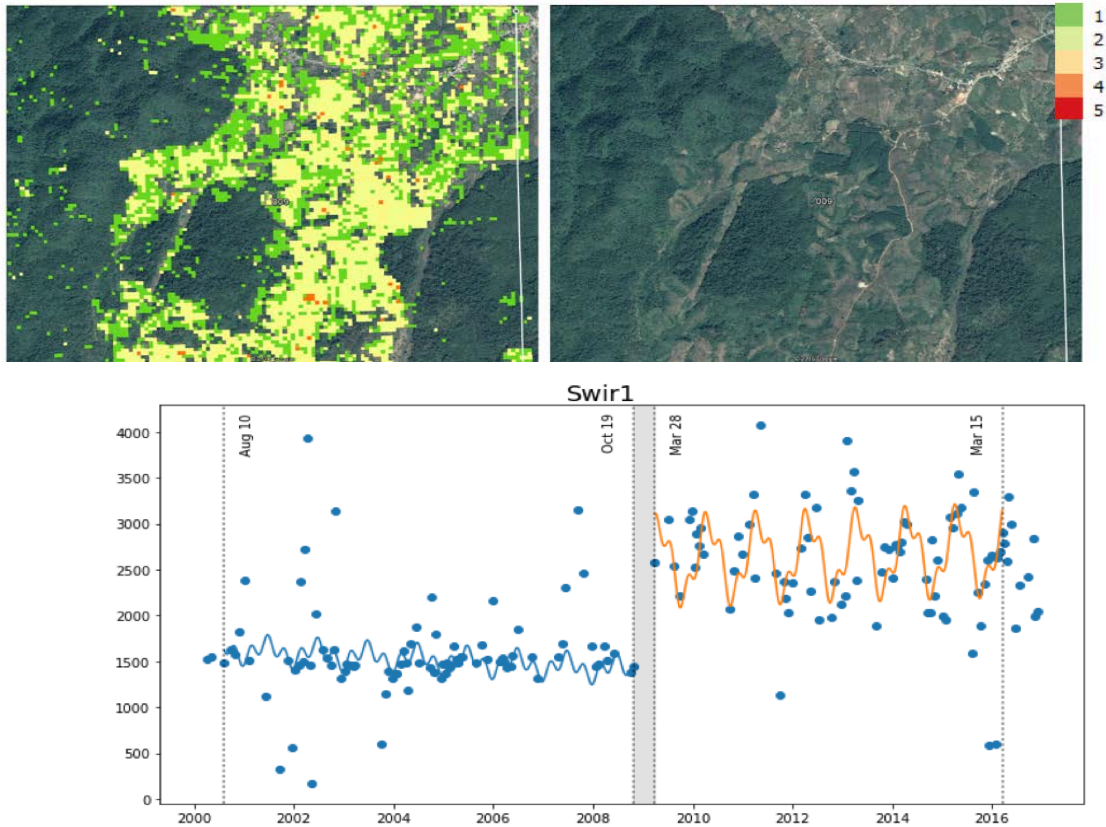


Figure 3. Continuous Change Detection and Classification (CCDC) algorithm (Credit: Zhu and Woodcock, 2012) was converted to Python by USGS, tested by the CEOS SEO on the Vietnam Data Cube, and is now called PyCCD. PyCCD time series model fits 7 bands to 6 weighted SINE and COSINE functions to find ‘breaks’ that equate to potential land change.

To further this Analytical Processing Capabilities initiative, CEOS agencies are asked to contribute in the following ways:

- Identify data analysts, system engineers, and system architects to join WGISS;
- Provide agency computing resources for prototype testing of application algorithms to take advantage of locally stored data or to utilise web-based protocols (e.g., WCS, APIs) for data interaction;
- Engage CEOS agencies that have implemented authentication systems to provide their lessons learned and best practices.

INITIATIVE 5: USER METRICS

All CEOS agencies are experiencing a shift in the number and nature of users seeking to benefit from their data and utilising their information systems to do so. These users are increasingly coming from a diverse range of sectors of society, sometimes non-technical, and with expectations of ease of access, use, and integration of space agency data with other information. Each CEOS agency has its own strategy for managing this change in user base, and a number have identified the necessity of accruing and exchanging information among CEOS agencies on the reality of the evolution of the user base and how FDA implementations are impacting them.

The CEOS FDA strategy is being defined as space agency information systems start to respond to the new possibilities being provided by advances in computing, networking, and storage, and we can

anticipate significant advances in the coming years. The proposed User Metrics initiative seeks to ensure planning and responsibilities are put in place for CEOS to leverage the experience being gained by individual agencies and to have an ongoing effort to collate available user metrics and to adapt the FDA strategy as these metrics are analysed beyond what is simply being downloaded.

In terms of user metrics, ambitions should be focused on the utilisation of increasingly sophisticated user management functionalities in the data access systems. Methods and tools applied in other data platforms and environments (e.g., social platforms) should be evaluated and adapted to better characterise user behaviour and means to catalyse EO data utilisation.

The FDA AHT proposes the following activities as the basis for this User Metrics initiative:

1. Develop a standard set of user interview and user experience questions. Write to users, requesting feedback. Example users may include CEOS Data Cube, Open Data Cube, and other platforms and applications, though generic inputs should also be accommodated. (WGISS)
2. Develop a database/warehouse for storing user feedback. (WGISS)
3. Develop a capability to analyse and report user feedback and a standard set of statistics that CEOS agency supported FDA platforms can report against, so that CEOS can provide summary statistics on how EO data is being used and for what. Promote results. (WGISS)
4. Analyse lessons from pilot activities, to scan the horizon and to submit proposals for new pilots to Principals. (WGISS)

To further this User Metrics initiative, CEOS agencies are asked to contribute in the following ways:

- Engage CEOS representatives familiar with user needs in analysis and metric reporting;
- Support development of universal metric capturing;
- Participate in open venues for discussion of abstract CEOS projects (webinars, conferences, meetings, etc.).

CONCLUSION

Leveraging the work accomplished by the FDA AHT in 2016, the team in 2017 has prepared a comprehensive and focused strategy for efficiently pursuing interoperable free and open Future Data Access and Analysis Architectures within CEOS for the foreseeable future. The strategy consists of five key initiative areas, which on their own have tremendous merit, but when combined form a powerful end-to-end approach for maximising CEOS agency resources and exploiting valuable EO data holdings for its user communities. As data volumes and diversity continue to rapidly increase at the same time that the EO user community expands and changes, it is imperative that CEOS agencies collaboratively invest in these products and architectures to maintain a viable, critical role in providing data and information that support decision making and meet user needs and expectations.