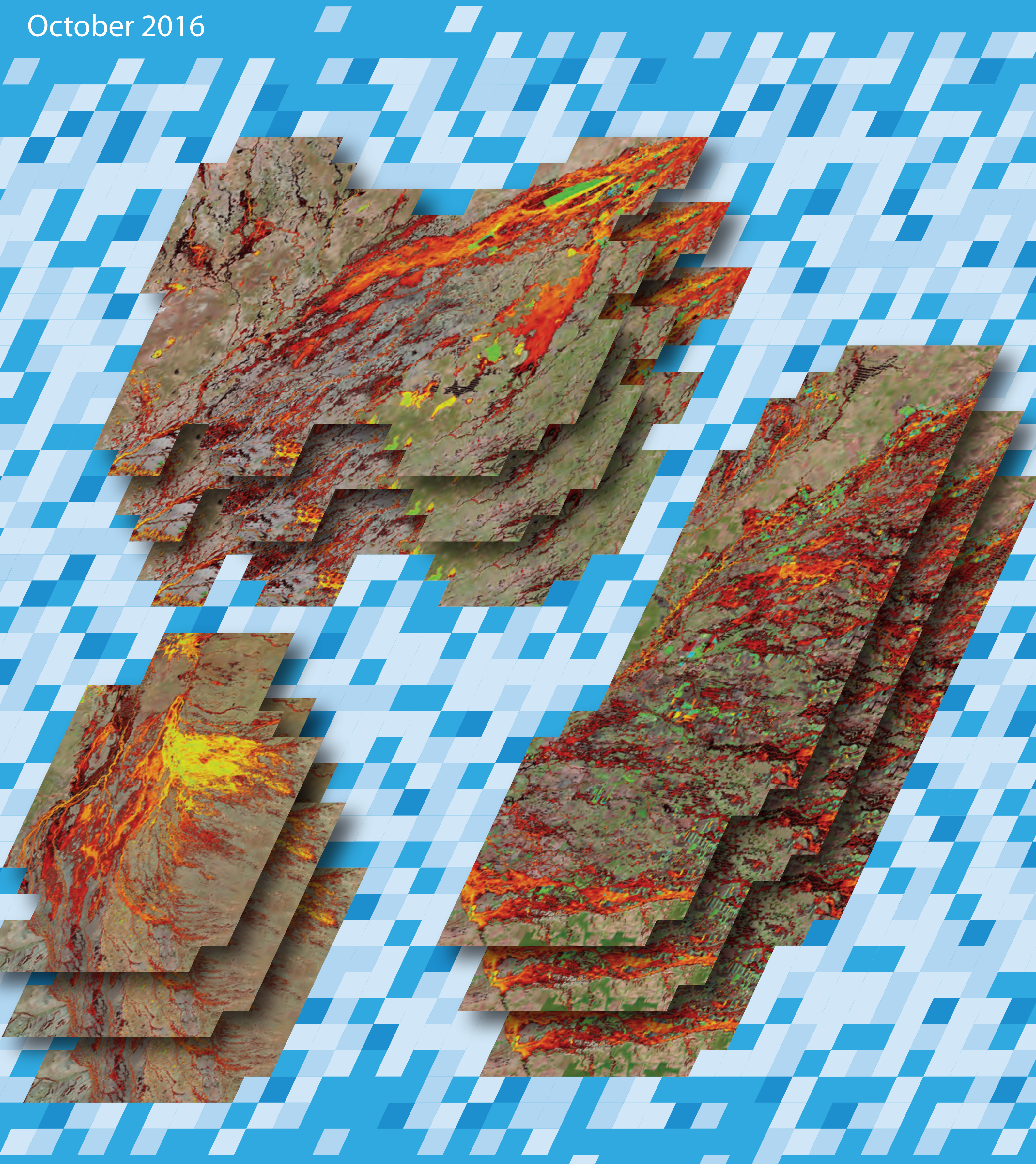
The CEOS Data Cube

Three-Year Work Plan 2016 – 2018

Version 1.1

October 2016



CEOS Committee on Earth Observation Satellites

Table of Contents

1 Introduction

1.1 Purpose

There is significant interest across multiple CEOS agencies and groups in the development of the CEOS Data Cube concept and multiple new activities and dimensions have been proposed over the last year. This Work Plan has been developed in order to provide focus, consistency and clarity in direction for these efforts. This document defines a 2016 – 2018 Work Plan for the CEOS Data Cube initiative, to encourage a CEOS-wide top-down strategy for its further definition, development and application, and to provide a reference for both internal and external communication of the activity and its components. It is currently an informal document to help stimulate discussion among the various relevant CEOS subsidiary groups and interested agencies. It is hoped that the document will evolve and serve as a focus for further development of the Data Cube strategy and implementation within CEOS.

1.2 Background

Recent years have seen an increase in CEOS interest in the uptake and application of EO satellite data in support of pressing global issues like food security, deforestation, and disasters - through significant investment in activities in support of GEOGLAM, GFOI and WCDRR, amongst others. It is assumed that the trend in this direction will only increase and that major changes in both space segment (with many and varied non-governmental participants) and ground segment, will have a significant impact on the expectations of data users for the contribution from CEOS and its space agencies. Multiple studies underway by CEOS groups (e.g., the Global Data Flows work by SDCG, and the Future Data Architectures Team work) are highlighting the importance of removing the obstacles of data size and complexity from individual user agencies if CEOS agency EO programmes are to realise their full potential and achieve societal impact, particularly within developing countries.

The data volumes involved in new missions mean that the old 'come and get it' data system model is simply not sustainable and new paradigms are required. This realisation is reflected in the Analysis Ready Data (ARD) strategy of USGS which will drive its entire Landsat ground segment design in future. It is also demonstrated in the provisional conclusions of the Global Data Flows work which advocates that considerably greater emphasis be placed on ARD and Future Data Architectures (such as the Data Cube) by CEOS as the peak EO coordination and standardisation body for civil space agencies. These developments are also seen as having the potential of reducing the dependence of CEOS and space agencies on third party partners, e.g. in relation to capacity-building. Significant new applications and users can be realised from the newly-achieved continuity and coverage of data supply in some sectors, but only if 'interoperability' and interchangeability is achieved among the core data streams and data handling burdens are removed from users.

At the same time, commercial cloud storage and processing service providers have developed potentially 'game changing' capabilities for users of EO data. As a part of their broader activities, they also have the infrastructure to deliver a more streamlined user experience, and are increasingly involved in serving users in sectors of interest to CEOS. These providers have ingested much of the free-and-open data available from CEOS agency programmes, and in some cases have invested in their own sources of satellite data (upstream on-orbit hardware, ground segment storage, and downstream data processing capabilities). This context presents a number of opportunities for CEOS programmes, including

- the opening of significant new data uptake channels, leveraging the strong and increasing supply of free-and-open data, combined with capabilities and investment of these service providers, to reach existing, and potentially new users;
- to rapidly demonstrate, prototype, and deploy government agency developed capabilities against those same free-and-open data streams utilising commercial infrastructure; and,
- to provide a non-commercial, enduring, exploitation platform for users which allows satellite data processing and information extraction without necessarily building dependency on commercial providers and applications.

All of this has the potential to result in improved user experience, a lowering of the bar for access and exploitation of satellite data, and can help to address increased user expectations of government-sponsored data provision, and to stimulate new applications and information services that are not currently foreseen. Investment in these areas is seen as vital for CEOS and its agencies to remain competitive and to meet user expectations in the changing context of EO satellite data information systems and services.

In response to these opportunities, and to both

- specific and immediate needs arising from prototype efforts underway by the SEO; and
- the initiative of the 2016 CEOS Chair in encouraging consideration of Future Data Architectures by CEOS to address some of the above challenges and opportunities;

the CEOS Systems Engineering Office (SEO) has taken the lead in progressing the definition of the CEOS Data Cube concept – taking forward the concepts proposed in the Australian Geoscience Data Cube (AGDC) through the collaborative framework of CEOS.

The Data Cube organises data into stacks of consistent, time-stamped geographic 'tiles', so that they can be rapidly manipulated in a high performance computing (HPC) environment. A relational database is used to track all of the tiles, pixels and associated metadata in the Data Cube. The Data Cube provides a collaborative infrastructure for many possible users and uses. Basic handling, calibration and processing of the EO satellite data are undertaken in a standardised way – once - and made openly available for the benefit of all prospective users and all prospective data providers.

1.3 Scope

This work plan spans the period 2016-2018. In its current form it includes a mixture of both existing, resourced activities which are already underway – such as the pilot projects within GFOI and GEOGLAM, and future, candidate activities which are more aspirational and assumed to be integral to the anticipated trajectory of the Data Cube initiative within CEOS. The Work Plan should be considered to be informal and an aid to stimulate the required discussion and decisions with CEOS regarding the way forward in this area.

As the lead on a number of the existing Data Cube prototype activities, and participant in the supporting strategy studies such as the Global Data Flows and Future Data Architectures, the SEO was the logical lead at the present time to take this initiative. It is recognised that there is broad interest within CEOS and its agencies and groups in the Data Cube efforts and that a number may wish to engage and support in various ways in future and these connections should be captured in future evolutions of the Plan. This document is proposed as a pragmatic step to bring focus to the multiple threads of discussion. It is hoped that this document will be updated annually to reflect both the internal and external activities supporting the CEOS Data Cube effort. There are also significant dependencies on the ongoing work on the core technology within the AGDC effort by Australian agencies, which although not identified as belonging within this CEOS-focused work plan, are nevertheless important foundation tasks that contribute to its outcomes. The tasks to evolve this core technology are included here for completeness until such time as the AGDC might develop to be recognised formally as a CEOS activity.

The Work Plan is organised according to the following *'threads'* relating to the different aspects of the activity and to the definition, development, deployment and application of the CEOS Data Cube from the perspective of both data providers and data and information users:

- **Core technology** behind the CEOS Data Cube, including in relation to data storage, ingestion and APIs and provider and user interfaces; the intention is to employ open source software as far as possible and to maintain and manage CEOS Data Cube code on the GitHub site;
- **Data preparation and formatting issues** including coordination and standardisation work that may be necessary, including in relation to data projections etc, to support ease of contribution to the CEOS Data Cube by all CEOS data providers; aspects of this will cover issues related to use of Analysis Ready Data (ARD) and the CEOS definition thereof;
- **User requirements and engagement –** for the development of a common understanding of the issues around obstacles in different sectors to EO data uptake and how the CEOS Data Cube might best address them; these activities will focus on the user experience and feedback;
- **Prototype deployments –** with an emphasis on deployment of Data Cubes in support of existing CEOS initiatives and priorities.
- **Capacity building** activities and coordination with internal (CEOS) and external partners to promote understanding and expertise in relation to use of the Data Cube by both data providers and users.

These threads are summarised in the schematic below which indicates their role within the overall Work Plan.

1.4 Contents

Section 2 summarises the vision and architecture for the CEOS Data Cube and its various supporting activities.

Section 3 defines the Work Plan outcomes.

Section 4 summarises the Work Plan tasks and high-level schedule.

Section 5 briefly describes the process of maintaining and updating this Work Plan, and also summarises some of the institutional issues that may need to be addressed.

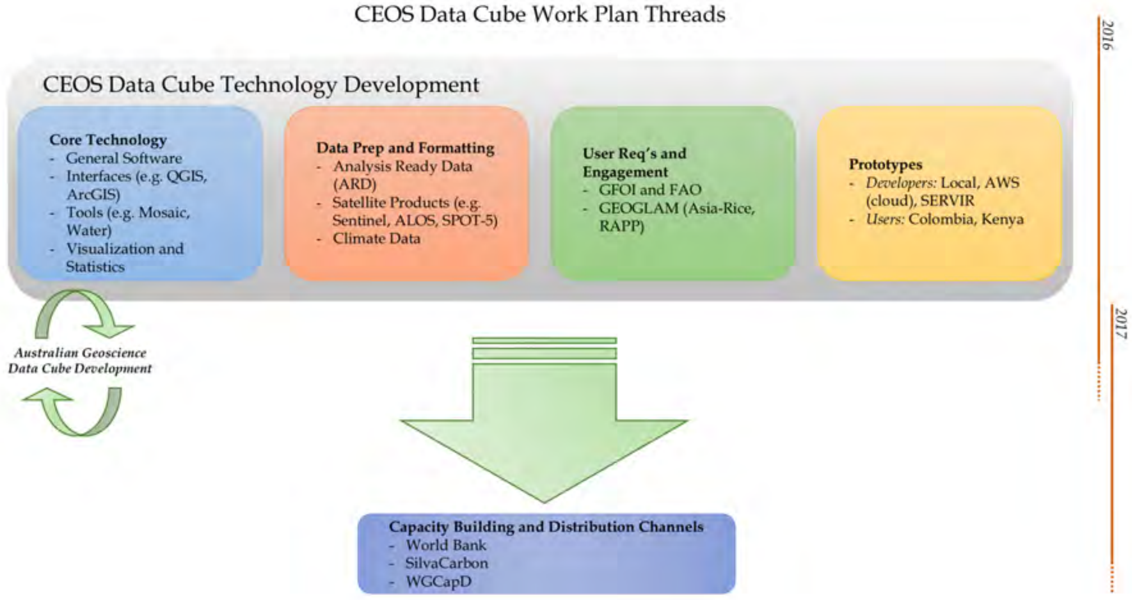


Figure 1-1: Work Plan Structure

2 Vision and Architecture

2.1 Vision

The CEOS Data Cube Work Plan is based on a vision which:

- Builds upon the substantial activities and ambitions in this area by a number of CEOS agencies and groups, including the Australian Geoscience Data Cube efforts of CSIRO and GA and the LCMAP efforts of USGS;
- Is easily communicated both internally and externally including to CEOS and its agencies, CEOS initiative stakeholders and countries;
- Can serve as a focus for contributing efforts by the various CEOS agencies and groups with expertise to contribute;
- Can provide the necessary direction and resources for the definition and execution of the activities and tasks required to realise the outcomes; and
- Informs and supports the discussion required among CEOS leadership regarding the coordination necessary for realisation of the outcomes.

The vision for the CEOS Data Cube is for:

- A next generation data architecture to 'unlock' EO satellite data – removing many of the technical challenges to its uptake and use – so that it might realise its full potential in service of society;
- An infrastructure that will provide a widely-used, free and open source software toolset for deploying local, regional or national pixel-based time-series of multiple datasets that are spatially aligned according to user needs (spatial region, time period, data layers, grid projection);
- A powerful solution for the challenge of data interoperability for the latest land surface imaging satellites, in tandem with advances in collaboration among CEOS agencies on Analysis Ready Data (ARD) for these missions;
- Ingestion of ARD products provided or facilitated by CEOS agencies;
- Users will connect free and open interface tools to the CEOS Data Cube for common analyses (e.g. cloud-free mosaics) or utilise Advanced Programming Interfaces (APIs) to connect to traditional geographic tools (e.g. QGIS) or their own tools;
- Closer connection of EO satellite data providers and end-users; knowing that basic quality assurance is undertaken by space agencies, users need not be concerned with low-level technical complexities of the data - but instead invest and focus on the application of the standard data products. More users will apply more data for more purposes;
- Efficient global data flows – the CEOS Data Cube working with ARD seeks to *'bring the users to the data'* and to provide a solution to the unsustainable movement of enormous datasets around the globe many times over;
- Facilitating ease of extraction of the information that users require, and minimising inefficient multiple distribution and processing of the same pixels from CEOS missions;
- Continued competitiveness of the user-facing aspects of government-sponsored Earth observation programmes, in comparison with the sophisticated user platforms offered by the 'big data players'.

Over time, it is expected that approaches such as the Data Cube will be used more frequently than traditional scene-based tools to exploit the large quantity of satellite data and take advantage of time series analyses. Both public good and commercial data providers can benefit.

CEOS recognises that the many traditional data services provided by space agencies are a critical need to support the Data Cube infrastructure, as the mechanism that provides the core data. Data Cubes are not meant to replace those data services, but provide an alternative method for utilising this information by enhancing interoperability with other space and non-space datasets and facilitating uptake and application of the data, in the form of information, by uses and users that would otherwise not be possible.

2.2 Architecture

The following figure and summary describe the Data Cube architecture. These elements are the major components of the architecture that connect users to satellite data. It is expected that the details of these elements will evolve over time, but the main element categories will not change.

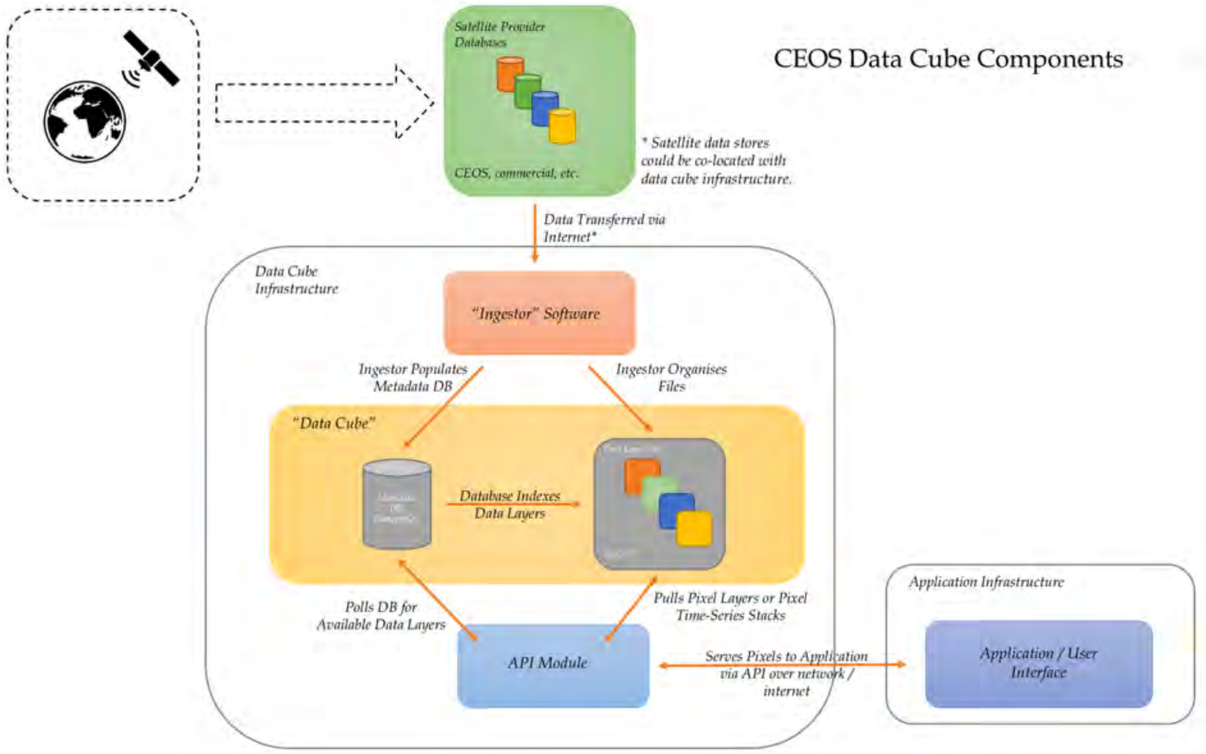


Figure 2-1: Data Cube Notional Architecture

<u>Satellite Data:</u> CEOS Space Agencies provide free/open data and make provisions to supply this data in Analysis Ready Data (ARD) format for immediate application.

<u>Ingestor Software:</u> Open Source software, developed by CEOS, is used to extract pixel-based information from traditional scene-based files and store those data into spatially aligned time series called *"data cubes"*. These ingestors support multiple datasets, multiple grid projections, and variable storage unit dimensions.

<u>Data Cubes:</u> Spatially aligned data layers, consisting of pixels, are stored in defined spatial and temporal dimensions and data formats. These storage units can be combined to represent time series of local, regional or national scale regions. Deployment is flexible and can support local computers, regional data hubs or cloud-based services.

Advanced Programming Interface (API) Modules: APIs, developed by CEOS, are specified protocols to allow users to query the content of the Data Cubes. These APIs address data storage, query and access, support analytics, and support various execution options. These APIs will also be used by external GIS tools for the integration of plugin tools.

User Interfaces: Users can interact with Data Cubes using open source reference user interface tools, supplied by CEOS, or by creating their own user interface tools. These user interface tools create analysis products from spatial and temporal analyses to support a diverse set of applications. The reference user interface tools, supplied by CEOS, will be focused on common functions such as mosaic creation and change detection.

3 Outcomes

3.1 Introduction

The outcomes envisioned from the Work Plan are described below, organised under the five activity threads.

3.2 Core Technology

CEOS is committed to sustaining the baseline Data Cube architecture through maintenance and management of the open source software site (GitHub) and facilitating the continued enhancement of the software products. This will be accomplished through donated efforts by many CEOS groups and Agencies, but mostly by the SEO, Geoscience Australia (GA) and CSIRO. This software development is focused on data ingestors, APIs, user interface tools and connections to other software tools (e.g. ArcGIS, QGIS, ENVI, ERDAS, PCI). The Data Cube open source website can be found at: http://www.ceos-cube.org

3.2.1 Data Cube General Software

Target: Ongoing with periodic software releases

The Australian Geoscience Data Cube (AGDC) version-1 had limited dataset and projection capabilities and used a GEOTIFF storage format. This version has been used by the Australian and SEO teams through early 2016. A new version-2 was released in July 2016 and includes significant advances. The team has demonstrated ingestor support for Landsat 5/7/8, Sentinel-2A, MODIS, SRTM, and Himiwari-8 datasets. Additional ingestors will be developed to support version-2 and other datasets (e.g. SPOT-5, climate data). The new version-2 code also supports UTM, Sinusoidal, Albers Equal Area grid projections, and utilises a flexible NetCDF storage format. The team will develop 3 new APIs that will be compatible with the version-2 storage format. These APIs are: "Storage Query and Access", "Analytics" and "Execution". The "Storage Query and Access" API will require testing and revision to be sure the API supports multiple datasets and metadata. Work on the "Analytics" and "Execution" APIs will be commenced in later 2016. All of this development will also require sufficient documentation and training materials (see Section 3.6.1) to allow local Data Cube installation at local site or on a data hub.

3.2.2 QGIS Interface

Target: Multiple tasks and completion dates (see below)

Data Cube APIs will be used to allow users to connect traditional GIS tools to the Data Cube. The SEO team hosted a 2-day meeting on June 7-8, 2016 with Chris Holden from Boston University to train the SEO team on how to use QGIS and develop an approach to connect this open source GIS tool to the Data Cube. This tool is commonly used by the community as it is open source and it also has many plugins for change detection, including the CCDC algorithm. The primary plugin is called "TSTools". Here are the detailed tasks and objectives:

December 2016: Chris Holden will update the TSTools plugin to enable it to source image inputs from a Data Cube "raster brick" directly using the Data Cube API. This plugin enhancement will allow viewing of data layers by selecting variables (bands, time, location). The current QGIS NetCDF browser does not provide this capability.

TBD: The SEO team will consider development of a "Custom Mosaic" plugin that is similar to the current web-based user interface for creating mosaics over a specific time and region. Users will be allowed to select scenes from any Landsat mission (5, 7, or 8). The analysis will use FMASK to display statistical results and show the resulting mosaic as a "virtual raster" image. The plugin should also output a layer(s) containing metadata describing the pixel provenance (e.g. acquisition date, sensor)

TBD: The SEO team will consider development of a "Water Detection" plugin that is similar to the current web-based user interface for detecting water extent over time. The tool will use Landsat data (5, 7, or 8) and the WOFS algorithms from Australia. A time slider will be used to view portions of the time series results.

Future (TBD): The SEO will investigate VCT, LandTrendr and BFAST algorithms as additions to TSTools for change detection. These methods have a large user community. The team will work with Chris Holden on potential integration into his TSTools plugin.

3.2.3 ArcGIS Interface

Target: Ongoing testing into 2017

Data Cube APIs will be used to allow users to connect traditional GIS tools to the Data Cube. The SEO team will investigate options for connecting the Data Cube to the commercial ArcGIS tool, developed by ESRI. This is the most commonly used GIS tool and is also utilised by several large groups via an enterprise license (e.g. SERVIR, Colombia IDEAM). Initial discussions with ESRI have begun (May 2016). Future plans will be developed once the SERVIR partnership is defined, as SERVIR uses ArcGIS at all of its hubs.

Recent contacts with ESRI (August 2016) suggest that NetCDF formats are easily connected to the tool. Demo testing will ensure that Data Cube layers can be explored. Additional discussions with ESRI will explore the use of the online ArcGIS tool to run Web Mapping Services (WMS) at remote locations. This will be the model used by many of the SERVIR hubs.

3.2.4 Custom Mosaic Tool

Target: August 2016 (release 1)

The SEO is developing a web-based user interface for custom mosaics to support Landsat and Sentinel optical datasets. This tool will allow users to define custom spatial regions and time periods (month to years) and develop cloud-filtered images. The output image can be viewed in true color and several false color band combinations. The tool will be released in increments to allow user feedback and addition of features over time. Initial work will be focused on Landsat with the addition of Sentinel-2 when ARD products are ready. Future updates will allow interoperable results whereby multiple datasets can be screened together to assess cloud-filtered results. This effort is different, yet similar, to the custom mosaic QGIS plugin described in Section 3.2.2.

October 2016: The Custom Mosaic tool will also be connected to the SEO AWS (cloud-based Amazon server) account to allow testing of the tool via the internet.

3.2.5 Water Detection Tool

Target: August 2016 (release 1)

The SEO plans to develop a web-based user interface for time series water detection. Water detection and water resource analysis is a high priority GEO and country application. Initial development has utilized the simple Landsat QA water flag (very coarse and low accuracy), but future versions will implement the 23-step water detection algorithm used for the Australian Water Observations from Space (WOFS) project and possibly the 5-step water detection algorithm used by USGS (per John Jones, researcher). These tools will be valuable for the assessment of water cycle dynamics, historic flooding extent and flood risk. Many users have expressed a desire for a water classification tool to support a large number of applications with minimal analysis effort (e.g. World Bank, Colombia). In addition, the SEO will produce water quality Jupyter (Python) notebooks. This effort is different, yet similar, to the water detection QGIS plugin described in Section 3.2.2.

October 2016: The Water Detection tool will also be connected to the SEO AWS (cloud-based Amazon server) account to allow testing of the tool via the internet.

TBD 2017: The Water Detection Tool will be expanded to include the Australian water quality (TSM) algorithm, and other possible algorithms for water quality (e.g. CDOM, Chlorophyll-A) or coastal erosion. In addition, climate datasets will be considered for the data cube including precipitation and surface temperature.

3.2.6 BFAST Change Detection Algorithm

Target: TBD 2017

The SEO has been in contact with Ben DeVries (Univ of Maryland) and Jan Verbesselt (Wageningen Univ.) regarding the BFAST change detection algorithm (http://bfast.r-forge.r-project.org/). The code is currently written in "R" and needs to be connected to the Data Cube through the API. BFAST is similar to CCDC-YATSM in that it performs interative change detection on long time series. There are two versions: Original BFAST looks at entire time series and finds all "break points" using harmonics to detect change. This code takes much longer to run. The newer BFAST Monitor (2012) compares a stable history (selected by user) to a monitoring period, which is similar to the CCDC process. Jan Verbesselt will propose a task to connect his BFAST code to the CEOS Data Cube and develop a set of instructions for users. He will also investigate methods for visualizing results using R, or QGIS.

3.3 Data Preparation & Formatting

Countries and international organisations have expressed a desire for support from CEOS to facilitate access and processing of satellite data into Analysis Ready Data (ARD) products to support a diverse set of applications. Systematic and regular provision of ARD will greatly reduce the burden on global satellite data users, and in particular, developing countries. These data include, but are not limited to, optical and radar at various spatial resolutions (coarse, medium and fine). There is also a need to utilize multiple datasets together through interoperable methods, where the data remains separate but takes advantage of complementary benefits, or through merged products, where the data is combined to improve temporal sampling.

3.3.1 CEOS Analysis Ready Data (ARD) Definition

Target: November 2016 (initial draft release)

The provisions for ARD and merged data products is seen as a responsibility of the CEOS Land Surface Imaging Virtual Constellation (LSI-VC). This group is currently working on an ARD description and specification document and will also track the progress of merged and interoperable datasets. The Data Cube implementation will take advantage of these outcomes and utilize them directly.

3.3.2 ALOS PALSAR Mosaics Training Manual

Target: October 2016

The SEO is working with Ake Rosenqvist (SDCG for GFOI) to investigate JAXA L-Band SAR (PALSAR) annual mosaic data from ALOS-1 for the Data Cube. Ake will obtain datasets over Kenya and develop a training manual to support interoperable optical data analysis in support of land classification. Once this is complete, the SEO will begin developing an ingestor for this dataset and for the more general ALOS-2 dataset in support of Asia-Rice. Future work may include developing training manuals for interoperable radar (C-band and L-band) analyses to support forest applications.

3.3.3 SPOT Data Products

Target: TBD

CNES is currently processing a large amount of historic SPOT archive data including data over Colombia and Kenya. Once available, the SEO will develop an ingestor for this data so that it may be used in the Data Cube along with other optical datasets. To date, Australia has not developed an ingestor for these data products. In addition, the SEO is investigating processing options to allow a simplified atmospheric correction. This work depends on the user needs of Colombia and Kenya, which will be defined in early 2017. A recent (August 2016) message from Steven Hosford (CNES) stated that the SPOT World Heritage website now has all of the SPOT 1-5 holdings available for search and download. The website shows a mixture of Level-1C and Level-2A data products. The SEO completed analysis in August 2016 of the available data over Colombia and Kenya. Colombia has 69 level-1C scenes and 56 level-2A processed scenes. Kenya has 29 level-1C and 29 level-2A processed scenes.

3.3.4 Ground-based Climate Data

Target: TBD

Many users are interested in using climate data (e.g. precipitation, surface temperature) within a Data Cube. The SEO will investigate options for adding this data to the Data Cube via a new ingestor. A demonstration of interoperable satellite and ground data will add a significant capability to the Data Cube.

3.3.5 Sentinel Dataset Ingestor Development

Target: TBD

The SEO will develop Sentinel-1 and Sentinel-2 data ingestors for the version-2 CEOS Data Cube. The Sentinel-2 ingestor will leverage the progress of GA, as they have already completed an ingestor for the AGDC. Sentinel-1 will require a more stable definition of ARD so that an ingestor can be developed for the most common Sentinel-1 data products. The completion of this task depends on a solid definition of the Analysis Ready Data (ARD) product for these two missions. To date, that has not been resolved.

3.3.6 Landsat-Sentinel Combined Product Development

Target: TBD

The SEO is communicating with Jeff Masek (NASA GSFC) to consider the use of two combined products in the Data Cube. Once complete, these products may be tested within the Data Cube as potential opportunities for users to utilize both datasets in a common approach.

There is an effort to combine Landsat and Sentinel-2 into a HLS (Harmonized Landsat Sentinel) 5-day merged product with atmospheric correction comparable to the USGS Landsat-8 version of LEDAPS (http://hls.gsfc.nasa.gov). The current focus is on single acquisitions that are radiometrically corrected (atmosphere, BRDF, spectral) and then resampled to a common grid (the S2 UTM grid at 30m resolution). The next step will be to take those single data products and select "best pixel" at 5 day increments.

3.4 User Requirements and Engagement

3.4.1 GFOI and FAO

Target: Ongoing, see Prototypes section for implementation

The SEO is working with the GFOI Project (Tom Harvey), FAO (Anssi Perkkarinen) and SDCG (Stephen Ward) to plan a future end-to-end country demonstration where the groups show a collaboration effort to support GFOI objectives. GFOI and FAO took an action at SDCG-9 to identify a country on the future FAO plan. This could also include SilvaCarbon and the Methods and Guidance Document (MGD). The latest plan is to conduct this demo in Colombia as they have an active forest monitoring team and initial knowledge of the Data Cube architecture.

One of the objectives of the SDCG's Global Data Flows report is to help inform the user requirements from the forest monitoring sector for the CEOS Data Cube and to poll user countries and key intermediaries such as FAO and World Bank as to the key challenges that the Data Cube can help address and what preference countries have for solutions.

Work is also underway with FAO to understand the requirements by their SEPAL system. At the request of FAO, the SEO team will demonstrate a connection between the Data Cube Custom Mosaic Tool and the FAO SEPAL data system. This would use the Amazon connection between the SEO and FAO accounts. The first test will be a NetCDF result that is directly sent to the FAO system for further analyses using OpenForis. This NetCDF sample data cube was sent to FAO in June 2016. FAO is interested in testing the capabilities of the Data Cube to shorten their in-country data processing and analysis time. For CEOS this activity provides valuable insight into the requirements of the user countries and intermediaries such as FAO.

3.4.2 GEOGLAM

Target: Ongoing, see Prototypes section for implementation

Early discussions have indicated a strong interest in the application of a Data Cube implementation in support of the Asia-RiCE project and the SAR monitoring of the Asian rice crop. Further exploration and scoping will be necessary, starting with a visit of JAXA to CSIRO to discuss this approach in July 2016. Discussions will establish the need and possibility of a Data Cube instance with initial applications around crop mapping in the Mekong, thus aligning with existing programs their by U.S. (SERVIR), ESA (GEORice) and

Japan (Asia-RiCE). Contributions from Australia are expected along with potential funding and contributions from ADB, USAID, EC, and DFAT. Details will be defined with the Australia team (leader, Alex Held) to define a schedule and planned outcomes.

The SEO has created a sample data cube over Carimagua, Colombia in order to test the Australian fractional cover algorithm (from Juan Gerschman) in support of GEOGLAM-RAPP (rangelands). The SEO will consider development and testing of a suer interface tool for rangelands fractional cover that will develop maps and plot time series changes.

In addition to the projects above, the University of Maryland (Chris Justice, UMD) is exploring options to fund a project to use a Data Cube connected to the web-based crop monitoring system developed by UMD. The current system using MODIS data, but future options will explore the use of Landsat for higher spatial resolution.

3.5 Prototypes

The Data Cube project is pursuing three deployment options to create, store and analyse Data Cubes. These include local computers, data hubs and cloud services. Prototype project details and progress are described below.

CEOS is currently supporting two primary applications to demonstrate the Data Cube in countries: forest and agriculture. These are supported through the Space Data Coordination Group (SDCG) for GFOI and the Ad Hoc Team for GEOGLAM. During past SilvaCarbon training workshops and SDCG meetings, the CEOS team was approached by representatives from Kenya and Colombia concerning an interest in using Data Cubes to support static land classification maps and time series change detection analyses. These two countries were selected as the first prototypes for CEOS Data Cube testing.

3.5.1 Local Deployment

Target: Ongoing

The SEO team, through its support contractor, Analytical Mechanics and Associates (AMA) will test Data Cube deployment on a small local computing system. The group plans to expand this local computing system in 2017 to add more storage space and more processing capability to allow online testing and evaluation of ingestion software and user interface tools. The results of this testing will inform countries on the requirements for local deployment.

3.5.2 Amazon Web Services (AWS) Deployment

Target: March 2017

The SEO submitted a research grant proposal to Amazon Web Services (AWS) in January 2016 to obtain free storage and processing "credits". This proposal was recently accepted on March 19, 2016 and awarded 45,000 credits. The proposed project will investigate the use of block-based (EBS), object-based (S3) and glacier archive storage of data products and Data Cubes and the use of "on-demand" and "spot" processing instances (EC2) to optimize operations and costs. In addition, the team will test remote connections to a Data Cube using and API to allow QGIS or ArcGIS functions. It is expected the proposed demonstration will be completed within one year, though additional data expansion and country testing may require more years.

3.5.3 SERVIR Deployment

Schedule: TBD

The SEO has interacted with the SERVIR team on several occasions. SERVIR has several regional hubs that may allow prototype testing of the Data Cube. These include the East-Africa, West-Africa and Mekong hubs. It is expected that the data cube architecture and the associated user interface tools could support many regional applications (e.g. agriculture, land management, water management, and weather/climate). The SERVIR team would provide the necessary computational storage and processing infrastructure along with its associated maintenance during the demonstration period. In addition, these teams use ArcGIS for many of their applications so there is a need to investigate Data Cube connections to this GIS tool.

The most recent interaction with the SERVIR team (August 2016) resulted in several specific actions. In general, the SERVIR team would like to evaluate Data Cubes as a mechanism for data management and simplifying delivery and processing. This will require that the SERVIR team and the NASA-SEO work closely on testing these functions in Huntsville (as a first step) and then migrating such functions to one of the SERVIR hubs (second step). Here are the near-term actions:

(a) The NASA-SEO team has asked the SERVIR team for a "sandbox" account on the Huntsville servers to allow testing of Data Cube components. This may include installing sample Data Cubes, Virtual Machines (VM), and user interface tools.

(b) The SERVIR team asked about the required "capacity" requirements for a Data Cube installation. The NASA-SEO team will define the minimum computing resources needed for installation and testing of Data Cubes. This would apply to the SERVIR Huntsville testing and SERVIR hub installation and testing.

(c) The SERVIR team was interested in the NASA-SEO interactions with the World Bank in the Niger Africa region. The NASA-SEO will discuss water management projects with World Bank and communicate that information with SERVIR. It may be possible to leverage the work of World Bank with the SERVIR hub to allow improved data management for that region.

(d) The SERVIR team discussed an automated landslide application called "SLIP-DRIP" led by Ahamed Kirschbaum. The SEO plans to investigate creating a prototype Data Cube application using the open source GitHub code. Integration of Landsat data, precipitation data and elevation data within a single Data Cube would be an excellent application example.

3.5.4 Colombia Prototype

Target: October 2015 to December 2017

The SEO is leading a prototype Data Cube project in Colombia to evaluate the user experience and the advantages of Data Cubes. The project is supported by the Colombian Government, CSIRO and IDEAM (Institute of Hydrology, Meteorology and Environmental Studies). A mini-cube (4 Landsat path-row regions) including Landsat 7 and Landsat 8 surface reflectance data back to 2000 was delivered in Oct 2015. The cube includes 20 time-series stacked tiles (1-deg square). Colombia has already demonstrated a significant growth in capacity as they have used the open source ingestor software to expand the data in their cube and they have modified the open source user interface to add median mosaics, PCA change detection and NDVI-based forest/non-forest maps. The SEO plans to evaluate the new additions from Colombia and work with their team to expand the supported datasets, convert the cube to version-2, and assess support for more applications. In addition, this prototype may also be expanded to include interactions with GFOI and FAO.

The 3rd Data Cube Workshop was held in Bogota, Colombia on August 30 to September 1, 2016. Colombia has made considerable progress in learning how to create and use Data Cubes. Land change detection, forest management and water management are the primary application needs. The SEO and CSIRO teams will continue to work with Colombia into 2017.

Outcomes from the meeting included:

- Discussed options for creating, storing and accessing Data Cube content
- Solved several issues with Data Cube ingestion
- Discussed "projections" at length and interoperable data use
- Initiated and tested MODIS data ingestion for one product (MCD43A4)
- Trained local users how to use the Custom Mosaic user interface tool and the Water Detection Python notebook tool

Several future needs were expressed by the Colombia team, including:

- Support with transition to version-2 and ingestion of a new country-wide Data Cube by the end of 2016. To date, they are testing a version-1 Data Cube with PCA change detection algorithms.
- Help with parallelization of ingestion and analysis processes to improve time.
- More ingestors to support MODIS (several products), SRTM, Sentinel-1, and Sentinel-2.
- Data Cube connections to ArcMap and QGIS.
- New and modified user interface tools to support user applications.
- Ability to modify an existing Data Cube by removing or adding data layers.
- Training and capacity building modules for other country organizations to use the Data Cube.
- Automation of new data identification, ordering, download, and ingestion into existing Data Cubes.

3.5.5 Kenya Prototype

Target: TBD

The SEO is planning a prototype Data Cube project in Kenya to evaluate the user experience and the advantages of Data Cubes. Initial discussions were held with the Australian Government and the Clinton Foundation (CCI and SLEEK). Two early operating versions of the Kenya Data Cube exist on the Amazon Cloud and a local SEO computer. A full country data including Landsat 7 and Landsat 8 surface reflectance data back to 2000 has been developed (11.5 TB, 7500+ scenes, 42 path-row regions). The cube includes 68 time-series stacked tiles (1-deg square). The Kenya team is currently utilizing scene-based methods to develop historic forest maps. The SEO has been facilitating data analysis and access to scene-based products during this period.

As of August 2016, Kenya is not ready for a Data Cube, so testing will be delayed until 2017. This initial lack of readiness is due to internal schedule pressures to complete their initial set of historic forest maps for UNFCCC submission. In addition, the Australian support and the Clinton Foundation support has ended and the local Kenya government leadership has changed. Future Data Cube work in Kenya is uncertain and will be considered again in early

2017. Without a significant interest from the local government and users, this project may not commence.

3.5.6 Asia-Pacific Prototype

Target: TBD

Early discussions with CSIRO and JAXA suggest that an Asia-Pacific Data Cube is desired in the Mekong region to support Asia-Rice. This prototype project would require development of data ingestors for several SAR datasets and regional hosting (e.g. SERVIR).

3.5.7 Balkans Prototype

Target: September 2018

Early discussions with Dr. George Zalidis of the InternBalkan Environment Center (I-BEC) and Argie Kavvada (NASA HQ, Booz Allen & Hamilton) have begun to plan a project in the Balkans with a focus on Albania. This project would support the Sustainable Development Goals (SDG 2 and 15) and focus on agriculture and inland water. Data would include Landsat, Sentinels and climate data. The proposed time period is March 2017 through September 2018. Funding would come from World Bank, if approved.

3.5.8 Canada Prototype

Target: TBD

The SEO and CSIRO are working with Ian Jarvis (Agri-Food Canada) to explore the potential for a Data Cube demonstration project in Canada to support agriculture monitoring. This prototype project would require development of a data ingestor for Radarsat-2 and regional hosting. This project may be preceded by a scene-based data management system to support JECAM. Discussions in 2017 are needed to define options. To date, there are no plans.

3.5.9 Switzerland Prototype

Target: TBD

The SEO is working with UNEP-GRID Geneva (Gregory Giulliani) to explore the potential for a Data Cube demonstration project in Switzerland. This prototype project may require training for local users to install the Data Cube software and guide the team toward implementation for local applications. Further discussions are needed to define the details and implementation options.

3.5.10 GEONETCast Testing

Target: TBD

At the request of GEO, the SEO will investigate the use of GEONETCast and DevCoCast (GEONETCast for developing countries) to supply newly acquired satellite data to countries using local data cube deployment. For example, if a local instance of a data cube is deployed at a site with poor internet then there is always the issue of how they receive and add new satellite data to their baseline Data Cube. Rather than sending them data updates using traditional (hand delivery) methods, it may be possible to use GEONETCast to get Landsat or Sentinel data streamed to their small ground station such that they can ingest it into their Data Cube. Due to other priorities, it is unlikely that this task will make siginificant progress until 2018, or later.

3.5.10 Landsat International Cooperator (IC) Ground Station Network

Target: TBD

The SEO plans to work with USGS to investigate the use of the Landsat International Cooperator (IC) ground station network as a data hub deployment approach for the Data Cube infrastructure. Access to Landsat data in these regional hubs may facilitate rapid creation of Data Cubes for regional users. The team will investigate local processing into ARD, creation of Data Cubes using ingestor software and connection of GIS and analysis tools to these data. Use of these regional hubs may have significant advantages. This task may be considered in 2017, but there are no current plans.

3.6 Capacity Building

Since the Data Cube concept is a rather new concept for the satellite and user community, there is a need to build capacity and provide training for those countries and specific users that desire to use this capability. Data Cubes are an alternative method for storing and analyzing satellite data and are not meant to replace traditional scene-based methods, which are quite popular and the current basis for most common data services. In addition to capacity building within countries, the Data Cube team also recognizes the need for interactions with other external groups, such as global banks, to assess potential mechanisms for support and application in relation to development programmes aimed at national spatial data infrastructures.

3.6.1 Data Cube Deployment

Target: TBD

The SEO will develop easy to use Data Cube deployment training modules and documentation and a dedicated website to streamline user initiation and use of the Data Cube architecture. These resources will get users “up and running”, download tool versions, configure their environment, and gain access to compiled and source code. The vision ranges from something as simple as “click here to download a working image of a Data Cube” to “what you need to build your own local version on your own infrastructure”. Future development of this material will consider deployment on computing clouds (e.g. Amazon) or regional data hubs (e.g. SERVIR). The training modules can be used for workshops and will include topics such as: Python, iPython, NetCDF, Jupyter notebooks, User Interface tools, and QGIS.

3.6.2 World Bank

Target: TBD

The SEO met with the World Bank (WB) in Washington, D.C. on April 13, 2016 and again on July 13, 2016. They expressed interest in the use of satellite data for resource management with a particular focus on water. The Data Cube provides them a mechanism to assess the progress and impact of their own programs. The SEO will continue discussions with WB to define plans for a future country demonstration.

The WB has several projects in the west Africa region (Chad, Cameroon, Nigeria, and Niger) focused on water management near Lake Chad. Data access and internet are particularly challenging for this region, so the use of a regional Data Cube would allow improved knowledge of historical water extent (floods and droughts), water quality and better near-term water management decisions. The WB leaders are working on their next fiscal year plans (start in July 2017) and would like to include the use of a Data Cube for this region of Africa. This requires discussion with local Africa government officials and projects leads that control the disbursement of loan funds. If successful, it is expected that we might start training some of the local people in early 2017 and then utilize the Data Cube for analysis and management after July 2017. In addition to the use of Landsat and Sentinel data, it we discussed the inclusion of precipitation data.

In addition to work in Africa, there was also interest from Keiko Saito regarding the use of a Data Cube to help her projects in the Pacific islands. It is believed the focus of the projects is disasters (flood risk) and coastal erosion. The SEO has contacted ESA about expanding S2 coverage over the WB list of islands and with some modifications, they will all be covered. The SEO will also work with CSIRO and GA to obtain their coastal erosion algorithms.

The SEO plans to visit WB again in late September 2016. This meeting will demonstrate the latest water detection tool. Future work in central Africa, future work in the Pacific Islands and linkages with SERVIR will be discussed.

3.6.3 CEOS WGCapD

Target: TBD

The SEO participated in the CEOS Working Group on Capacity Building and Data Democracy (WGCapD) meeting in Hampton, Virginia in late March 2016. The SEO proposed a plan to explore approaches to support future Data Cube capacity building within countries. WGCapD and the SEO agreed to develop a Capacity Building plan for the Data Cube initiative whereby WGCapD identifies regional representatives that could receive training on the implementation of the Data Cube architecture, train those individuals, and support the planning and implementation of workshops that utilize Data Cube technologies for regional applications.

3.6.4 SilvaCarbon

Target: TBD

The SEO has attended several past SilvaCarbon workshops. The SEO has continued a dialogue with Sylvia Wilson (USGS) and Doug Muchoney (USGS) regarding future support of workshops with a plan to attend 3 or 4 in 2017. Once the Data Cube architecture is in an early “operational” state, the SEO plans explore opportunities to use these workshops to build country capacity to use Data Cubes to support forest monitoring. To date, these workshops have been focused on COVE tool training and helping countries with data acquisition planning.

4 Task Plan and Schedule

The planned tasks, status and schedule summary is shown in the table below. The tasks in this list represent activities with current progress or those with a good potential for future implementation. Several outcomes listed in Section 3 are not shown in this table due to a lack of commitment or plans.

Task Name	**Responsible Group**	**Task Description**	**Status and Schedule**
Core Technology			
Cloud-based User Interface Tool	SEO	Continue improvement of the cloud-based Amazon Web Services (AWS) User Interface (UI) tool to demonstrate custom mosaics and water detection applications	Updated release of new custom mosaic methods (NDVI min/max, median, and oldest/newest pixel) and TSM (water quality) features to UI by Oct 17. Additional development and testing planned in 2017.
RAPP Fractional Cover algorithm	SEO, CSIRO	Demonstrate Juan Gerschman's (CSIRO) fractional cover algorithm on a Data Cube at a RAPP (rangelands) site in Colombia.	Initial release planned for Oct 21 as part of the UI. Additional development and testing planned in 2017.
Coastal erosion algorithm	CSIRO, SEO	Obtain coastal erosion algorithm from Australia and test on several SEO sample data cubes.	Require papers and/or algorithms from CSIRO. Plans to add to UI in 2017.
Water quality algorithms	CSIRO, SEO	Obtain water quality algorithms from Australia. These include CDOM, and Chlorophyll-A.	Algorithms needed for CDOM and Chlorophyll-A (if available). Plans to add to UI in 2017.
WOFS comparison testing with Australia	CSIRO, GA, SEO	CSIRO/GA to identify a location in Australia to perform WOFS testing.	SEO awaiting identification of location in Australia for WOFS testing. Once identified, the SEO will develop a Data Cube and start testing. Planned for 2017.
QGIS integration	SEO	Develop QGIS plugins to allow Data Cube reading and common analyses (e.g. cloud-free mosaics, water detection)	In discussion with Monty Kickert (BoundlessGEO). Telecon planned for Oct 6. Second telecon on Oct 20. Task likely to begin in Nov-Dec 2016.
ArcGIS integration	SEO	Develop ArcGIS steps to allow Data Cube reading. May pursue other common analyses (e.g. cloud-free mosaics and water	In discussion with Peter Becker (ESRI). He is writing a set of instructions to guide users on how to read a NetCDF CEOS Data Cube into

		detection) in the future.	ArcGIS. Due in late 2016.
BFAST change detection integration	SEO	Demonstrate a connection between BFAST and the Data Cube for the calculation of time series change. Consider display of results in QGIS and/or ArcGIS.	In discussion with Jan Verbesselt (Wageningen Univ.). Plans TBD.
CCDC-YATSM change detection interaction	SEO	Demonstrate a connection between CCDC-YATSM and the Data Cube for the calculation of time series change. Consider display of results in QGIS and/or ArcGIS.	Under development by Chris Holden (Boston Univ.). Chris is updating his TSTools plugin to allow reading of the CCDC-YATSM results and the Data Cube layers. Expected in early 2017.
Data Preparation and Formatting			
SAR data processing and ingestion	CSIRO, ESA	Demonstrate processing of common SAR datasets (e.g. Sentinel-1) into ARD and ingestion into a Data Cube	New CSIRO task starting in late 2016. Details are under review.
Sentinel-2 ingestion	CSIRO, GA, ESA	Develop Data Cube v2 ingestion configuration files for Sentinel-2 ARD datasets	Need to develop a standard ARD development process with ESA and then progress toward ingestion. Planned for 2017.
SPOT-5 ingestion	CSIRO, CNES	Develop Data Cube v2 ingestion configuration files for SPOT-5 ARD datasets	Plans for a discussion with Stuart Phinn at CEOS Plenary in Nov 2016.
ALOS-PALSAR mosaic ingestion	SEO, JAXA	Develop Data Cube v2 ingestion configuration files for ALOS-PALSAR 25-meter mosaic ARD datasets	SEO has ALOS-PALSAR mosaics for Kenya. The next step is to create the ingestion config file. Planned for completion in early 2017.
SRTM DEM ingestion	SEO	Develop Data Cube v2 ingestion configuration files for SRTM DEM ARD datasets	Plan to create and test ingestion in support of the SLIP-DRIP demo. Due by end of 2016.
ASTER DEM ingestion	SEO, JAXA	Develop Data Cube v2 ingestion configuration files for ASTER DEM ARD datasets	Plan to create and test ingestion in support of the SLIP-DRIP demo. Due by end of 2016.
Climate data ingestion	SEO	Demonstrate ingestion of ground-based climate data along with satellite data into a Data Cube	Under discussion with Rob Waterworth (Mullion Group) as part of a proposal in Australia. Plans TBD.

Landsat enhanced ARD products	CSIRO, SEO	Investigate approaches to allow simple creation of enhanced Landsat ARD products. This would include: solar illumination, terrain illumination and Nadir BRDF-Adjusted Reflectance (NBAR). Such approaches would greatly improve the value of Landsat data prior to Data Cube ingestion.	Working with Peter Caccetta (CSIRO). AMA has identified 4 Landsat scenes in Colombia with terrain variation. Other data will include MODIS BRDF, SRTM and ASTER DEM. Plans to send to CSIRO in mid-October.
Evaluation of the Harmonized Landsat-Sentinel (HLS) product	SEO	Investigate the use of the 5-day merged HLS product and ingestion into the Data Cube	Working with Jeff Masek (LSI-VC) to obtain HLS samples for testing. Will further investigate in 2017.
Prototypes			
Colombia Prototype (in progress)	SEO, CSIRO	Demonstrate the use of a Data Cube in Colombia to support forest mapping and additional country-wide applications.	Currently working with a team at IDEAM (led by Pilar Rivera) and Andes University (Harold Castro). The team is making great progress using a v1 cube and plans to develop a v2 cube by the end of 2016. Testing will continue into 2017.
Switzerland Prototype (in progress)	SEO	Demonstrate the use of a Data Cube in Switzerland with little intervention from the SEO team.	Switzerland team (led by Greg Giuliani) is moving rapidly and making good progress. Planned meeting in Feb 2017.
FDA Pilot Project (under review)	FDA Team	Demonstrate an end-to-end CEOS Data Cube in Colombia to support GFOI. Consider topics of ARD, data interoperability, cloud-free mosaics, change detection, and GIS tool use.	Working with the FDA team to plan. This may become an expansion of the current Colombia Prototype (in progress) and go into 2018.
Asia-Rice Prototype (under review)	CSIRO, SEO	Demonstrate the use of Data Cubes in Asia to support Asia-Rice. Investigate the use of the SERVIR Mekong hub for hosting.	In discussion with Dan Irwin (SERVIR) and Alex Held (CSIRO). Plans TBD.
SLIP-DRIP Prototype (under review)	SEO	Demonstrate the SLIP-DRIP algorithm for landslide detection using a Data Cube. This demo will utilize multiple datasets	In discussion with Dalia Kirschbaum (NASA GSFC). Completed an initial telecon on 30-Sep. Next telecon planned for Oct-11. Plans TBD.

		(Landsat, GPM, SRTM DEM, ASTER DEM)	
Balkans Prototype (submitted proposal)	SEO	Demonstrate the use of a Data Cube in the Balkans region. Initial proposal (ESSENTIAL Project) will demonstrate in Albania.	Proposal submitted by George Zalidis (i-BEC) and Argie Kavvada (NASA, BAH) on 01-Sep-2016. Pending award.
User Requirements and Engagement			
FAO SEPAL tool review	SEO	Working with Erik Lindquist (FAO) on approaches to connect the Data Cube to the FAO in-country operations for forest mapping. FAO has developed a data management tool called SEPAL, which is scene-based. There may be a potential to have SEPAL connect with a Data Cube (NetCDF) format.	Waiting for next release of SEPAL v2 due by end of 2016. After review, the next steps will be planned.
Capacity Building			
ALOS-PALSAR training manual	SEO	Develop a training manual and interpretation guide for the use of ALOS-PALSAR 25-meter annual mosaics.	First report received on 30-Sep-16. Under review.
SERVIR training	SEO	The SEO Data Cube team will visit the SERVIR team in Huntsville, AL to train their team on Data Cube creation and use	Meeting planned for Jan 24-25, 2017
World Bank Lake Chad Demo	SEO	Demonstrate the WOFS water detection algorithm over Lake Chad. Report to World Bank.	Expanding dataset to include the entire lake boundary. Water analysis results will be updated and shared with World Bank by the end of 2016.
World Bank Tonga Demo	SEO	Demonstrate the WOFS water detection algorithm and coastal erosion algorithm over the main Tonga island. Report to World Bank. May be a future part of the Pacific Islands project.	Working with Keiko Saito (World Bank). SEO currently building a small Data Cube for Tonga. Plans to add the cube to the current AWS UI by the end of 2016.
World Bank Togo Demo	SEO	Demonstrate the WOFS water detection algorithm and coastal erosion	Working with Nicolas Desramaut (World Bank). SEO is currently building a small Data Cube for one

		algorithm over the west Africa coast near Togo. Report to World Bank.	path-row of this region. Plans to add the cube to the current AWS UI by the end of 2016. Will investigate coastal erosion algorithm (see other task) in 2017.

Table 4-1: CEOS Data Cube Tasks, Status and Schedule

5 Governance

This Work Plan document has been prepared to help manage and communicate the activities of the CEOS Data Cube initiative. The SEO will maintain it as a living document so that it remains a current record of Data Cube activities and plans. The SEO foresees several releases of this document throughout any year due to the changing environment and needed adjustments.

The SEO utilizes a portion of its annual budget on the Data Cube project and is therefore limited in its ability to progress on all tasks. The SEO is also supported by several other groups on a "best efforts" basis, such as the SDCG for GFOI, the Ad-hoc GEOGLAM team, and certain CEOS Agencies, such as CSIRO and GA.

Progress on specific outcomes is dependent on available SEO resources and the contributions of internal (e.g. CEOS agencies, Ad Hoc Teams) and external (e.g. FAO, World Bank, SERVIR) groups.