CEOS Future Data Access & Analysis Architectures Study

Interim Report
Version 1.0 – October 2016

CEOS Committee on Earth Observation Satellites

CEOS Future Data Access & Analysis Architectures Study

Interim Report

Version 1.0, 13th October 2016

Issues for Plenary Discussion and Decision

1. Approval for the Ad-hoc Team on Future Data Access and Analysis Architectures to continue for a further year to complete the mandate. And confirmation of the Co-Chairs.

2. Agreement for the proposed pilot project to be progressed in parallel with the ongoing report work, with oversight by the FDA team and contributions from LSI-VC, SEO, and SDCG.

3. Invitation for further proposals for practical demonstrations in the area of FDA for 'lessons learnt' evaluation by CEOS Principals at CEOS-31.

4. Action for CEOS and SIT Chairs to confer with the FDA Team to ensure necessary CEOS Principal engagement on the strategic issues arising from the 2017 Report, in support of identifying common ground as the basis for a long-term CEOS strategy.

This report was achieved thanks to all agencies and their representatives who participated in the Ad-hoc Team process, in particular the writing team: Robert Woodcock (CSIRO), Tom Cecere (USGS), Andrew Mitchell (NASA), Brian Killough (NASA/SEO), George Dyke (CEOS Chair Team), Jonathon Ross (GA), Mirko Albani (ESA), Stephen Ward (CEOS Chair Team), and Steve Labahn (USGS).

1.Introduction

Overview

With each passing year, new generations of Earth observation (EO) satellites are creating increasingly significant volumes of data with such comprehensive global coverage that for many important applications, a 'lack of data' no longer needs to be the limiting factor.

Extensive research and development activity has delivered new applications that offer significant potential to deliver great impact to important environmental, economic and social challenges, including at the regional and global scales necessary to tackle 'the big issues'. Such applications highlight the value of EO to Ministers and others who ultimately adjudicate on investment in programmes and missions

The challenge is in providing the right settings so the potential can translate to reality both for individual CEOS members through to global initiatives.

For EO, the gap between data, application and user needs to be bridged. Currently, many applications fail to successfully scale up from small-scale research to global or regional operations because of a lack of suitable data infrastructure. Even today, much archived EO satellite data sit under-utilized on tapes. Despite multiple examples of big data analytics across application domains, significant development remains consigned to prototypes, pilot projects, exemplars and test-beds.

Addressing this challenge is difficult for advanced economies. It is simply not technically feasible or financially affordable to consider traditional processing (e.g. local desktop workstation) and data distribution methods (e.g. scene based file download) to address this 'scaling' challenge in many economies, as the size of the data and complexities in preparation, handling, storage, analysis and basic processing remain significant obstacles. This challenge is already holding back key GEO/CEOS initiatives such as the Global Forest Observations Initiative (GFOI), Disasters, Water Resources and the GEO Global Agricultural Monitoring initiative (GEOGLAM).

Addressing this problem by individual users working on their desktop workstations has not resulted in an optimal solution and misses the opportunities offered through collaborative environments that bridge data providers, intermediary value-adders (such as researchers and industry) and users to work together across domains, and across geographic boundaries, to co-create solutions.

Fortunately, just as satellite Earth observation technology has advanced significantly, so to has information and communication technology. The data management and analysis challenges arising from the explosion in free and open data volumes can be overcome with the high-performance ICT infrastructure, technologies and architectures now available. These solutions have great potential to streamline data distribution and management for providers while simultaneously lowering the technical barriers for users to exploit the data to its full potential.

Purpose

In response to these changes, the CEOS Future Data Access and Analysis Architectures Ad-hoc team (FDA-AHT) has been tasked by the CEOS Chair team to assess the potential of new technologies and approaches, identify key issues and opportunities, and propose a plan of action for consideration by CEOS.

This report has:

1. Reviewed relevant initiatives and plans being undertaken by CEOS and related agencies;
2. Reviewed lessons learned from the early CEOS-led prototypes currently underway with the governments of Kenya and Colombia;
3. Identified key issues and opportunities resulting from the trend towards Big Data, Analysis Ready Data, EO application platforms, etc;
4. Made recommendations for the way forward for CEOS and its agencies, including in relation to standardisation, interoperability, and how current CEOS priorities might be advanced through a set of proposed activities.

This study is anticipated to be of value both to CEOS Agencies as data providers and to existing and prospective users and beneficiaries of EO satellite data. The full potential of EO satellite data will not be realised with the obstacles that users face in current data handling and analysis approaches. Global initiatives such as GFOI and GEOGLAM exemplify the difficulties that countries without developed national spatial data infrastructures face in terms of lack of capacity to handle EO satellite data. This capacity gap is a major hindrance to the uptake of EO data in the types of global initiatives and agendas CEOS has stated as important in recent years. Moreover, even many developed countries are struggling to determine how best to capitalise on large and rapidly growing EO data collections and would appreciate guidance on both best practice they can adopt themselves, and approaches that can bring CEOS Agencies together to support approaches that maximise value from their collective constellation of over 130 Earth observation satellites.

Structure of the Report

In the creation of this report submissions were made by a number of CEOS members regarding current trends and their specific development responses in EO systems architectures and applications. As each agency has different terminology, operational methods, language, policy context and business drivers the submissions appear different in the detail. Careful analysis though shows common trends and responses that are particularly relevant to the mission of CEOS. In addition, the report has been limited to only those aspects of the EO systems architecture that are high priority or impact directly on the CEOS mission.

The report is structured as follows:

Chapter 2 consolidates contributions and identifies trends and priorities in EO systems architecture development across CEOS agencies. It serves as a baseline of current architecture and near future development responses.
Chapter 3 discusses the challenges faced in EO system architecture design and development for the medium to long term future. It serves to identify the key challenges that must be addressed in future data architectures
Chapter 4 describes key architectural responses that seek to resolve the challenges identified in Chapter 3. It is not a complete architectural description and focuses on the essential elements necessary for the CEOS mission leaving details to future projects or Agency developments.
Chapter 5 summarises the outcomes of the report and presents recommendations on Future Data Architectures and activities for CEOS.

2. Current trends and developments in EO systems architecture and applications

Earth Observation (EO) programmes of space agencies are facing a number of trends which, taken together, are driving the need for change in the ways in which data are processed, accessed, distributed, and analysed. The magnitude and speed of these changes is determining the importance and urgency with which change is required in future EO system architectures. This chapter will attempt to summarise some of those key trends and develop an assessment of the state of these systems architectures and their ability to meet these user needs and user applications.

Maximising the Value of Earth Observations

Maximising the value of EO is a fundamental driver for all CEOS agencies and a key part of the CEOS Strategic Guidance. There is an expectation that publicly funded EO agencies should maximise the value returned to the country through the application of national data holdings. As a fundamental driver most agency systems architectures have been designed to deliver calibrated observations and produce value added products for use by other Government agencies on predominantly national and global societal, environmental, and scientific problems. In recent years there has been a steady trend across all agencies towards greater integration of diverse EO data holdings for land, inland water, coastal, climate and ocean purposes with other data types held by more diverse Government agencies - "Comprehensive collection and integration of ... information independently controlled by governmental agencies should be promoted and such information should be disclosed appropriately to increase the convenience for users to access and handle such information." (JAXA, Ocean Policy, 2013). The increased integration produces a more diverse range of applications and leads to complexity in the value chain as observations are combined with analytics to meet multiple user requirements (Figure 2-1).

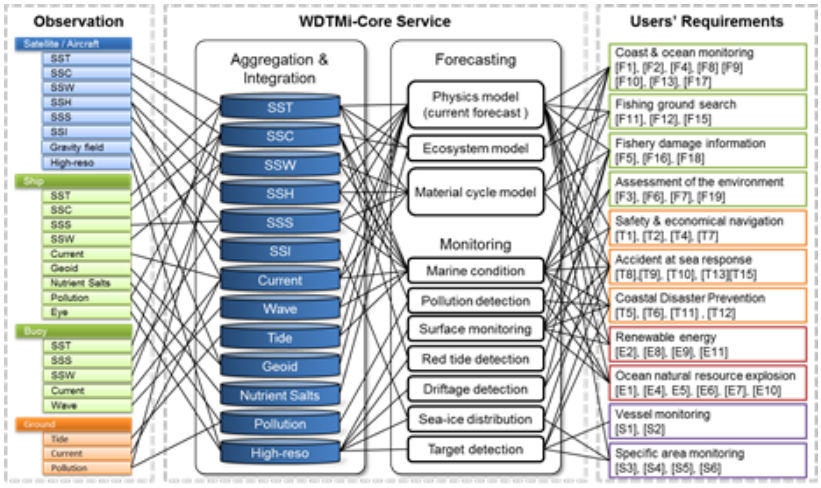

Figure 2-1: One example of the complex inter-dependencies that exist in meeting user requirements from diverse EO data sources. Courtesy of JAXA Ocean Data Infrastructure.

Increasingly EO data are valued not only for its scientific and technological value but as a potential field for economic growth through new commercial ventures and industry development. Agencies are being asked to promote and strengthen an EO industry whilst continuing to maintain their strong scientific and technological foundation.

Additionally, a recent USGS report (Miller et al, 2012) attempted to evaluate the benefits of Landsat data to its users. The report concluded that more than 80% of the users saw environmental benefits and more than 90% saw improvements in decision-making. The estimated annual economic benefit of this free/open data is greater than US$2 billion per year. Though this is one example, there are likely many more similar examples among CEOS missions and datasets. Overall, there is an increased relevance of EO missions for resource management and decision-making.

Open Data Policies

Beginning with the Instituto Nacional de Pesquisas Espaciais (INPE's) move toward free and open data policies in 2004, data policy changes have been critical and are influential in leading to a significant trend across all CEOS agencies. Another important policy change was the USGS adoption of free and open Landsat data in 2008. This allowed international Landsat collaborators to change business models and to move from being image resellers to being data scientists and providers. Other agencies including ESA and JAXA also championed these changes, however the global reach of Landsat data meant that the US decision had the widest implications. NASA has been operating under an open data policy since 1990 and other organizations are moving towards such policies in recent years. We have seen the recent movement from Europe (e.g. Copernicus Sentinel data, ESA Earth Explorers and Heritage missions' data) and Japan (e.g. mid to coarse resolution data) to provide free and open data.

Within Australia, changes in government policy further supported the direction of free and open data. Agencies such as Geoscience Australia were able to support the development of simple but effective open licenses to apply to their EO data distribution. Resources previously committed to license management and manual distribution of products were able to be re-focused on scientific exploitation of Landsat data.

Fewer hurdles to data access implies broader use of data. This results in a much higher return on investment by organizations from their spaceborne and ground systems' assets. However, this also means higher workloads for data systems. It can also be very difficult to track the resulting impact once it leaves the confines of the agency. There is considerable business model innovation occurring across many agencies and users of EO data as the impact and value of EO data now readily available is understood.

Open Source Software

In addition to free and open data, it is also desirable to have free and open access to analysis software and tools that facilitate the use of data. Some examples of this include QGIS (an open source geographic information system), THREDDS and Geoserver (open source EO and other geospatial data delivery services). Open source software policies help with this data exploitation. A draft open source policy has just been released (March 2016) for public comment by the U.S. Federal Chief Information Officer (see

https://sourcecode.cio.gov/) emphasising that using and contributing back to open source software can fuel innovation, lower costs, and benefit the public. A specific example within CEOS is the Data Cube initiative. This project relies on open source software for the creation of data cubes (ingesting satellite data) and the interaction with data cubes (application programming interfaces). It is believed that open source software will stimulate application innovation and the increased use of satellite data since these advanced technologies can be utilised globally and even by developing countries that are not traditional users of satellite data.

Emergent EO Analysis Platforms in the Cloud

Cloud computing has had a dramatic impact on the availability of highly scalable and accessible computational and data infrastructure. With recent advances in the Cloud technology the use of scientific computing in the Cloud has grown significantly with many previously HPC only applications now running effectively in the Cloud environment amongst there are several EO offerings.

The most well-known of these would the Google Earth Engine, which "combines a multi-petabyte catalogue of satellite imagery and geospatial datasets with planetary-scale analysis capabilities...available for scientists, researchers, and developers to detect changes, map trends and quantify differences on the Earth's surface".

Amazon Web Services Cloud offerings operate via a different business model, providing access to the Cloud platform but not directly offering applications, preferring to provide a base platform upon which others build their own. Amongst the EO offerings available are open source EO technologies like GeoTrellis which make heavy use of web services architectures to provide scalable raster operations, and commercial offerings like Planet Labs providing fully automated scalable image processing pipelines downloading data, performing corrections and analytics at 5-10 terabytes per day.

Whilst AWS and GEE are mentioned similar initiatives exist on Microsoft Azure (Layerscape) and other organisations. Each of these platforms acquire CEOS agency data and places it into a managed environment with closely coupled and highly scalable analytical capabilities in the Cloud. Significantly the business models in use vary from free, to open source, through to full commercial service offerings. How well these models work both for the organisation operating the service and for their customers has not been explored in this report but the business model in use is clearly an important consideration. There are also operational issues for these organisations in acquiring the CEOS data, as one example NASA has been working with multiple cloud providers (Amazon, Microsoft and Google) to better understand how NASA data systems can better support bulk data downloads by cloud providers. The objective is to enable efficient discovery, access and transfer of large volumes of data from NASA archives to commercial clouds and to make the transition to commercial cloud infrastructure easier. Some of these are providing efficient metadata, the use of standard file formats, use of standard structured directories to hold files and use of well-defined map projections.

Increased Commercial and Non-Governmental Interactions

NASA, NOAA, and the USGS conduct a large part of their EO activities through contracts with commercial entities. ESA and JAXA conduct their activities through commercial entities as well, even though there are differences in the nature of contracts among the different countries. There is a distinction between commercial entities working under contracts with government agencies for development and operation of observing systems and data systems, and other commercial entities that apply the resulting information to some self-sustaining profit-generating activities. The Federation of Earth Science Information Partners (ESIP) is an example of both types of commercial entities collaborating with government and university organizations. From the point of view of data architecture, commercial interactions have an influence on standards for interoperability, among other things. With the increase in open data policies and open source software (see examples above), there will be an increasing need to work closer with commercial entities to expand the use of satellite data and its benefits. Though there are many examples in the commercial world, some of the greatest impacts on satellite data application have been made by Google and Amazon.

In addition to the traditional commercial entities, one must also consider the non-traditional or non-governmental groups as they also play a major role in connecting EO data to users. Some recent examples in CEOS have been connections with the UN-FAO (supporting forest management), World Bank (high interest in water management), SilvaCarbon (funded by USAID to support forest management), and the Clinton Foundation (working in central Africa). These groups, and many others, utilise satellite data and will continue to increase their demand for such data to support regional and local applications.

Pre-processed Analysis Ready Data

Countries and international organizations have expressed a desire for support from CEOS agencies to facilitate access to and processing of satellite data into CEOS Analysis Ready Data for Land (CARD4L) products. CARD4L are satellite data that have been processed to a minimum set of requirements and organized into a form that allows immediate analysis without additional user effort. Existing CEOS agency efforts include NASA's MODIS model that set the standard for ARD, Dr. David Roy's Web Enabled Landsat Data (WELD) model that has stimulated demand for more highly processed Landsat data, Geoscience Australia's efforts, as well as USGS Land Change Monitoring Assessment and Prediction ARD efforts.

Systematic and regular provision of CARD4L will greatly reduce the time and technical burden on global satellite data users, who have up to this point needed to invest significant efforts in preparing EO data for further analysis. The provision of this data is possible through many options including systematic processing and distribution, processing on hosted platforms, and processing via toolkits provided to users. The CEOS Land Surface Imaging Virtual Constellation (LSI-VC) team is developing CARD4L definition and specification documents that will define the details of CARD4L. These documents intend to improve the current and future provision of EO data and to maximise the value of the data to users and address the needs of the majority of global users. In addition, CEOS is developing a Data Cube (spatially aligned time series stack of

pixels) architecture that depends on CARD4L to allow immediate creation of Data Cubes and subsequent analyses. The result of this effort will be improved interoperability among land based datasets, facilitating time series analyses and enhanced global use and scientific value of satellite data. There is a similar desire to develop ARD specifications for ocean, inland water, and coastal environments as well for non-optically based instruments like SAR.

Advanced Storage and Distribution Architecture for Growing Data Volumes

Growing data volumes will continue to place new requirements on advanced storage and distribution architectures. With the increase in CEOS missions and sensors and increased spatial and temporal resolution, the amount of data available to the world will be orders of magnitude greater than in the past. In a recent CEOS Ad-hoc Space Data Coordination Group (SDCG) Global Data Flow (GDF) Study for the Global Forest Observations Initiative (GFOI), it was found that expectations must be managed to achieve sustainable solutions that can be adapted as country capacity increases. Significant risks are associated with maintaining infrastructure and expertise though there is always a desire to utilise the latest modern technologies for storage and distribution of key satellite data.

Through the efforts of the Working Group on Information Systems and Services (WGISS), CEOS is developing new approaches to ease data discoverability and developing standards for metadata and distribution formats (i.e. spatial or temporal tiles vs. standard scene files). The CEOS Working Group on Calibration and Validation (WGCV) is working on new approaches for product validation to improve data quality. Finally, the CEOS Data Cube initiative is investigating new advanced storage formats (pixel-based data cubes versus scenes) that will allow for subsetting of data (spatial and temporal relevance), significant data compression and facilitate distribution to non-expert users.

Time Series Analyses and Change Detection

The extended lifetime and success of many missions (e.g. Landsat and MODIS) have allowed a new ability to exploit information from long time series. Following the open release of Landsat data, there has been a significant number of these analyses focused on land use change (including inland water and coastlines). These analyses have utilized a variety of change detection tools (e.g., Continuous Change Detection and Classification (CCDC), Breaks For Additive Season and Trend (BFAST), Hansen et al.'s global forest gains and losses) to find trends on the data or identify periods of significant change. One example is the Australian Water Detection from Space (WOFS) algorithm that calculates time series pixel-level water observations. These results provide critical information for water management that will allow users to assess water cycle dynamics, historical water extent and the risk of floods and droughts.

The availability of time series data will continue into the future as programs such as Sentinel develop plans for sustained long-term measurements. To efficiently and effectively use these large time series datasets, there will be a need to use new technologies and data architectures such as Analysis Ready Data, Data Cubes, Data

Provenance, and advanced databases. With such advancements, we will be able to assess the impact of climate and land change on people and natural resources over time.

Advanced User Requirements

Advanced user requirements are those requirements that extend beyond the typical use cases and place significant demands on future data architectures. These include, but are not limited to: real time applications including rapid monitoring of land and water changes; diverse applications and output needs for monitoring, assessments and projections; integration of multiple datasets (climate, in situ, economic, demographic) and sensor types (e.g. optical and radar); fusion of datasets (e.g. combining Landsat and Sentinel-2); access to lower level products for "power users"; the use of high performance computing for complex analyses; and Climate Data Records (CDRs) and Essential Climate Variables (ECVs). As more complex data becomes available, there will be an increasing need for new technologies and data architectures to meet those needs.

As an example of high performance computing needs in Australia, High Performance Computing (HPC) became available in 2011 for the management and analysis of Australia's Landsat data collections under the "Unlock the Landsat Archive" project. This work applied the automated production systems to the entire Australian Landsat collection, which was moved to the Australian National University National Computational Infrastructure (NCI) to complete this work. Geoscience Australia became a partner in the NCI in 2012. The HPC environment allowed the data to be held on disc, rather than tape, and directly attached to large computing resources.

An example of advanced user requirements in Japan is found in the WDTMi-Core project. The goal of this project is to provide integrated data of both international ocean-observing satellites and in-situ observations. The projects intend to develop a common infrastructure which integrates, manages, and provides data, models, and analytical results from ocean related satellites and In-Situ observation.

Increase in the Number and Diversity of Users

In the past, satellite data users were traditional scientists and researchers with the capacity to obtain and analyze the data for decision-making. Due to trends in open data and cloud-based hosting (e.g. Google and Amazon), there has been a significant increase in the number and diversity of global EO data users. No longer can we say that the users are only technical, directly manipulating EO images. We must now consider hundreds and thousands of non-expert users that rely on EO derived products like land cover, vegetation condition, etc. Examples of these non-expert users include local decision-makers utilizing Google Earth, "crowd sourcing" projects to compile EO data, and the use of common smart phones to access EO data.

The rapid development of the earth observation industry and the limited availability of scientific expertise it led to has fostered a paradigm of start-to-finish science delivery, starting with earth observation data selection, to data preparation, to conducting earth observation analytics, and concluding by relating resulting earth observation patterns and trends to implications to specific application domains. However, the growing awareness of the power of advanced earth observation analytics has substantially

increased demand from the end-user community for accessible and usable value-added products built from data-cube technologies which remain in the domain solely of the earth observation expert community. This increase in demand for the earth observation scientific community to deliver on needs and requirements domestically and internationally cannot be met under the traditional paradigm of science delivery. Thus, the paradigm of start-to-finish science delivery is rapidly diminishing in favour of the development of a new paradigm with end-users taking ownership of the final stages and earth observation scientists working to develop rich and diversified analysis ready data that enables scientists and decision-makers across multiple application domains.

As the number of users and their diversity increases, there will be increased questions over the control of EO information and its application to decision-making. As stated by Karen Litfin (MIT Press, 1998), the relationship between satellite technology and state sovereignty continues to become more complex. Today, users include multinational corporations, scientists, policymakers, grassroots environmental groups, and indigenous peoples." With the widespread use of cloud storage and computing services the traditional geopolitical barriers no longer exist. Users are now free to interact and download data from cloud-based servers, though they must "trust" these providers and adhere to their service level agreements. For many international users, this is a very large "leap of faith" and many are quite reluctant to use cloud-based services and prefer to store and analyze the data locally to ensure "ownership" and control of the information. With increasing data volumes and complexity of data, this approach is not sustainable for the future, and this paradigm must shift to enable these users to take full advantage of EO data.

Successfully abstracting the complexity of sensor data will free users to focus on the development of algorithms which can run across sensors. This is a key area for further development and is now only possible due to the advent of datacube technologies which can apply techniques such as machine learning to model the variability in target spectral response between sensors. With the range of new sensor missions on the horizon, removing the need for deep understanding of each sensor will be integral to the effective utilisation of these new data.

Limited Internet in Developing Countries

Though developed countries depend on the internet for most of their data access and analyses, this is not always the case for developing countries. In a recent CEOS Global Data Flow (GDF) Study for GFOI it was found that 50% of the studied countries had internet speeds below 5 Mbps, which would require ~19 days to download 1TB of data. Even with speeds improving, the cost of downloads is often a prohibitive factor (http://www.tandfonline.com/doi/full/10.1080/01431160903486693). Without consistent and cost effective internet, there must be other options for users to obtain data or at least interact with data. As part of the CEOS Data Cube project, the use of regional data hubs (e.g. SERVIR) that are close to users and have improved internet performance are being investigated. Other approaches are considering hosting of data on larger cloud-based hubs, such as Google or Amazon Web Services, to take advantage of web mapping services (WMS) that allow users to work with the data remotely, use the advanced cloud computing capabilities for analysis and then only download small resulting products over limited internet bandwidth. This same approach is also being

implemented in Europe for the Copernicus Services to “bring users to the data” for interacting with Sentinel data over limited bandwidth internet while avoiding download of large datasets.

3. The challenge and opportunity of changing user expectations and increasing EO data volume, variety and velocity on EO systems architecture

The impact of volume, velocity and variety

With individual missions producing 10's of Petabytes per year by 2020, and 100's of CEOS missions, there is no question data volumes are undergoing a step change in growth rate. In general, individual CEOS agencies factor in operational data requirements as part of mission design and this is unique to that organization's needs (sharing of design information and lessons learned clearly occurs). Whilst volume and growth rate are clearly major challenges for a mission, it is both velocity and variety that pose a major challenge for EO systems architectures when viewed from a CEOS community perspective:

- Large data volumes imply a need for very accurate and structured search services so that users are not overwhelmed with the volumes they need to access. File discovery is no longer sufficient.
- Higher acquisition rates and the new near-real time applications that benefit from them (e.g. Disaster monitoring for bushfires, Flood monitoring; Agricultural monitoring, etc.) require the entire acquisition, calibration through product generation pipeline in its entirety to be completed prior to next acquisition. Automation is essential at all stages and includes third-parties.
- The benefit of increasing variety can only be realized if the barriers to accessing multiple collection types simultaneously and consistently are reduced significantly so that the burden of discovery and integration does not scale along with volume, velocity and variety.
- Data volumes and velocity are such that in an increasing number of cases, the volume is too large to move data to a local analysis platform on the available networks.
- Many local analysis platforms (e.g. PCs, mobile platforms, departmental clusters) are not large enough to benefit from the increasing volumes of data available
- With increased volumes, automation and third party application development conveying data quality information become increasingly important

Earth Observation (EO) data systems today face challenges from two directions. On the "push" side, new instruments and models are producing ever greater volume, velocity and variety of data. On the "pull" side, user expectations of the data, and the systems that serve them, are expanding.

The push side challenges are those of the wider Big Data movement. Volume growth stems from improvements in such factors as sensor resolution, space-to-ground bandwidth, retrieval algorithms and the computing power for processing them.

Within NASA, for instance, the Earth Observation data volume of 15 PB as of January, 2016, is expected to increase by a factor of ten by 2024. In addition to finding affordable

space for all of the data, the volume increase also manifests either in more data files, or larger data files, or both, compounding the data management problem.

The amount of data and information being generated by the Copernicus space and service components are also challenging traditional dissemination channels. For the space component alone the combined archives of units A and B of Sentinels -1, -2, -3 will amount to approximately 6 PB per year from 2018 onwards. Extrapolating from current usage patterns, each product may be downloaded 10 times on average, amounting to 60 PB downloaded per year, or 160 TB per day.

The sources of the variety growth lie in the development of more processing algorithms extracting yet more information from the raw data and the increasing availability of model data alongside Earth Observation data.

The change in user expectations on the pull side has two sources: technology advancement causes users to expect more modern capabilities and interfaces, such as broadly searching for data across many federations of systems and then retrieving data directly from the disk on which they are archived and retrieving only the specific piece of the data needed, or highly interactive, responsive-design user interfaces. However, another source of change in user expectations is the expansion of the user communities to include more different kinds of users in such diverse areas as interdisciplinary research, business and government applications and education. With the broadening and diversification of the user communities, information about data quality and data provenance will increase in importance as users access data from many more data providers. Increasing free and open data policies will reduce the hurdles to data access and facilitate broader use of data. This will result in a much higher return on investment by organizations for their space-borne and ground systems assets.

Section 2 has already shown the current trend towards more non-EO specialist users, the ubiquity and diversity of geospatial applications and changing role of participants in application development. In the subsections below, we look at how these challenges manifest in and are handled by the various areas of EO data systems.

Data Discovery

The biggest challenge in Data Discovery is presented by the increase in Variety of data. The chief sources of variety are:

- Sensor / instrument
- Platform
- Spatial footprint
- Spatial aggregation
- Temporal aggregation
- Level of processing
- Retrieval algorithm

Given the large variety of sources and types of data, inevitably, similar data products meeting a given application's need become available from various data archives. This makes it difficult for end users to sift through to find the most appropriate data products to meet their needs. A metadata clearinghouse can simplify the search tools' task of

querying the necessary sources for data product information. However, such a clearinghouse requires a common metadata model, such as ISO 19115, which can provide some levelling to allow for data product comparisons. Many metadata clearinghouses standardize their metadata to a single, interoperable metadata format, such as ISO 19115. However, system designers are now becoming aware that they need to continue supporting multiple metadata standards in their clearinghouse. This is mostly in response to concerns expressed by the data provider community over the expense involved in converting existing metadata systems to systems capable of generating a new metadata format. As an example, in order to continue supporting multiple metadata standards, NASA designed a method to easily translate from one supported standard to another and constructed the Unified Metadata Model (UMM) to support the process.

Likewise, it is also helpful to provide an Application Program Interface (API) to allow the development of a variety of search clients, ranging from simple data search-and-fetch scripts to full-featured web user interfaces. There are two common standards for such APIs used in the community: the Catalog Services for the Web (CSW) is a highly structured API, while the OpenSearch API is lightweight and based primarily on keyword search. Note that supplying a "flagship" search client, or reference implementation, can provide not only a useful search tool for the community in its own right, but also a platform for adding new search engine features and a starting point for prospective application developers. One area requiring more attention in the future is dealing with the growing diversity of user communities. For example, the highly technical, detailed data product descriptions demanded by the science research community are often not appropriate or useful to, say, a citizen scientist or a business application owner.

Data Access

To use the data and information, users have to be able to access them. Two types of access are considered:

A. either the data and information are moved to the users' premises (download/pull/push/broadcast reception) so that the users can process them and obtain from them the results they want;

B. or the data stays in the data centre and the processing occurs next to the data with the obtained results either sent to the users or stored online for further use by other users down the line ("bring the user to the data" scheme).

Both types of access are possible but the costs involved are different: either the costs consist of higher network bandwidth to move the data from the storage facility to the users' own processing infrastructure or the costs consist of more processing capacities next to the data and of less network bandwidth for moving the results to the users (if the users still need to download them, if not, the resulting value added products may well be stored in the same data centre and published from there to be further used by others).

To reach overall optimal efficiency of model (b), it should be reinforced by the guarantee that the data will always be available, otherwise users will be tempted to download the

data and archive them as an insurance of permanent availability without necessarily immediately knowing what to do with the data/ information downloaded. This aspect of availability also incorporates latency whereby data assets that require significant lag periods for delivery will also encourage users to duplicate the data holdings.

Both models may scale, however model (a) only would only scale if the bandwidths are scalable, while model (b) scalability may play on different facilities with more load balancing possibilities (storage, processing, bandwidth availability). The higher the volume of data and information to be accessed, the more advantageous model (b) is likely to become.

The clearest challenge to Data Access arises from data volumes. If storing the volumes on disk is unaffordable, access to the data that must be archived on tape is significantly degraded, both in latency and overall throughput. The latency in turn generally requires an asynchronous access method, with notifications to the user on data readiness. Accordingly, compression techniques are typically applied whenever possible, ideally lossless internal compression such as that available with the Hierarchical Data Format (HDF) and Network Common Data Form (NetCDF) version 4. Data Access is also complicated by larger data volumes which prompt more users to request data subsets for just the data of interest. Subsetting is complicated by the fact that data product specific subsetting tools do not scale well with variety. Preferable are standard formats (e.g., HDF and netCDF) that are tractable to general tools. Also, some data services, such as those offered by the Open Geospatial Consortium's (OGC) Web Coverage Service (WCS) or the Open Source Network for a Data Access Protocol (OPeNDAP) can provide subsetting on the fly for data in standard formats, over the Internet, along with other services such as reformatting that smooth over the Variety aspect.

Data Usage

Most of the data are stored typically in forms convenient to producers. These forms are not necessarily convenient for users to access. Different users need different access mechanisms. Some want access via the traditional granule download access; some want granule level access after some subsetting (time, space, bands); some want pixel level access, i.e., without regard to granule boundaries, and some want pre-built standardized value-added products. It is challenging to provide access to spatial subsets of long time series of data. How can we provide enough descriptive information to the users to enable the types of access they need?

Given the variety of potential applications, 'right formats' mean that data would be needed in several different formats and processing levels as needed by the different user community served. This includes long time series of homogeneous data to monitor changes and long term trends (e.g. Climate), lower level data (i.e. Level 0) to allow scientists to contribute to algorithms development, higher level products (i.e. Level 1, Level 2 and higher) for research and application activities and operational services. Data need to be continuously upgraded and valorised to ensure continuity of observations and comparability with new mission data (e.g. Sentinels) and fitness for purpose for an effective utilisation and exploitation. As indicated in the "EO Science Strategy for ESA", "Long-term, carefully calibrated and documented data sets of the Earth system derived

from EO satellites will become a legacy of the highest importance for science, policy makers and society".

Traditionally, formats have been the most problematic, but custom ASCII and binary formats have largely given way to standard, self-describing formats such as HDF and NetCDF. This trend in turn has given rise to the development of a number of versatile data tools that work on large numbers of datasets, such as Panoply, IDV, GrADS, as well as finding their way into support by commercial tools such as ArcGIS, IDL and Matlab.

However, while the data format has become less of an issue, it is still important to follow conventions on data structures and attributes (such as the Climate Forecast convention) to enable these tools to be effective. Furthermore, there are some areas, such as diverse map projections, that remain challenging; even where tools support transformations to other projections, naive users applying these re-projections may unknowingly introduce significant artefacts into the result.

Data Volume represents a second challenge to data usage by the end user, who may be faced with finding enough space to store the data or enough processing power to analyze it in a reasonable amount of time. To address this, some systems offer a certain amount of processing at the archive, which may range from sophisticated subsetting schemes to running the user's algorithm at the archive.

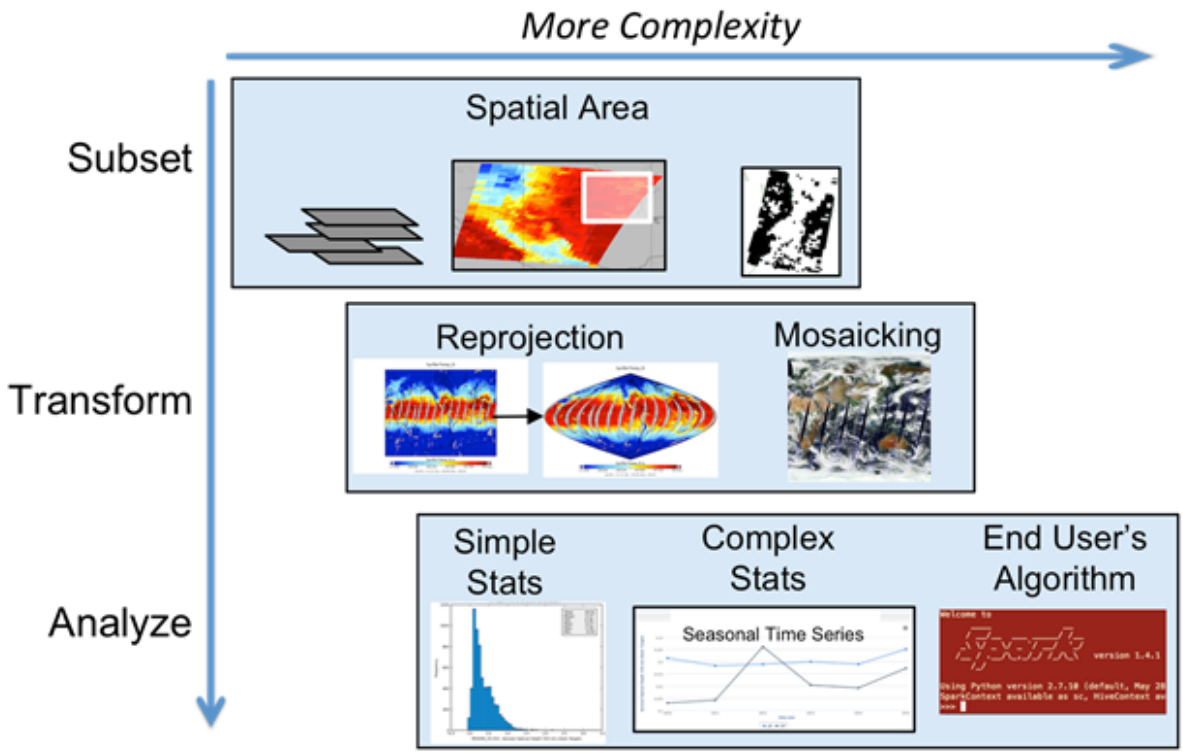

Fig. 3-1 User analysis of data ranges from fairly simple variable subsetting to both more complex content subsetting (e.g., quality filtering) and to user-provided algorithms.

Data System Functions

In addition to the user-centric aspects enumerated above, some challenges fall on the data systems serving them. Perhaps the most important is the stewardship of the data. This is complicated by both the data volume and variety. Large volumes may force a data center to resort to tape backup. Although media costs are cheap relative to disk, the tapes must be continually inspected (read) to ensure they are still readable. Furthermore, the variety of data levies a responsibility of stewarding as much of the available context as possible, from descriptions of the instrument to algorithm documents to product specifications.

In addition, science data centers require a number of back office capabilities to plan and manage evolution of the system with changing requirements, user communities and technologies. These include reliable and thorough metrics of data archived and distributed. Ideally, the science data center also maintains a repository of data citations, in order to gauge the research value of datasets, particularly when many are being managed. The data variety also affects communities of diverse science data centers that coordinate their development, which benefits from a shared development environment with such elements as wikis, ticket tracking systems, and code repositories.

Identified aspirations, constraints and open problems

Whilst data volumes, variety and velocity are clearly a major technical challenge, probably the greatest challenge to maximising value from EO data, and on the system's architecture is the changing expectations of users.

There are a number of CEOS community related issues that also need to be considered in developing Future Data Architectures:

- Cost re-allocation - there are potential opportunities, particularly with Cloud based business models, to re-allocate the cost of data distribution and computation into a more user-pays oriented model rather than having the entire burden supported by an agency.

- Economics and Performance of Cloud computing for EO storage and analysis - this remains an open problem currently. Whilst Cloud potentially has excellent scalability for analysis against local data, the economics of storing the data, sovereignty and software performance are still being assessed for EO applications.

- Capacity building - with expectations of economic growth through new EO businesses and the expanding user base there will be increasing demand for training and capacity building. User support and communication to a broader group will also be necessary placing pressure on CEOS space agencies.

- Standardisation and interoperability - Standards are key to interoperability. It is preferable to have a small number of standards to facilitate search and access to data in an interoperable manner across the CEOS community (search standards, controlled keywords, metadata). We also need a more diverse set of access

standards to support actions such as subsetting, file format conversions, spatial projection conversions, tile access, and pixel level access.

- Administrative reporting - essential to maintaining investment in EO activities is being able to measure the ROI through its use and value generation. In a future where there are many more third parties developing applications and business, along with massive automation and consumer use of open data, it will be increasingly difficult to collect metrics necessary for administrative reporting using things like user logins or agency portal access. With machine to machine connectivity it will be necessary to use alternate methods to gather such information whilst respecting privacy issues and remaining true to the principle of open data. The risk to EO data becoming an anonymous contributor to major applications outcomes is high as increasing use sees it become taken for granted.

4. The Future of EO Data Architectures

Notwithstanding the progress made in recent years, the difficulty of finding and using EO data is still a barrier to realizing their full potential and to properly harness "Big Earth Science Data" for societal benefit. This is mainly due to the inability of current paradigms in architectures to keep pace with the rapidly changing EO data management landscape and impediments to the free flow of data through analysis workflows on suitable computational hardware throughout the entire lifecycle. The simultaneous availability of complete, high-quality, trusted, ready-to-use and integratable datasets and of the enabling infrastructure allowing their effective utilisation and exploitation by an increasingly diverse user community is key to maximize the impact of EO data assets.

An enticing possibility, evident in many of the development trends, is changing the paradigm of user analysis from the current one in which users download their own copies from multiple data centres in order to perform local analysis over multiple data products. Provisioning data in the Cloud may lead to more processing in place (i.e., in the Cloud with the data), thus reducing network transfers and user data preparation and management headaches. This in turn may enable more cross-sensor and interdisciplinary analysis. This change in the user paradigm for data analysis is leading to development of processes that short-cut what the user needs to do to start the data analysis. CEOS Analysis Ready Data for Land (CARD4L) reduces the need for deep EO calibration expertise and broadens the accessible user community. Data Cube technology demonstrated in the CEOS and Australian Geoscience Data Cube projects shows how supporting pixel-based access to CARD4L will allow users to access the specific spatial, temporal and sensor ranges of the satellite data needed for science or industry applications improving the ease of use of time series and interoperable datasets.

Another example is the shift in cost of storage vs compute vs data transfer. With increasing compute capacity available the ability to process data on-the-fly will become a reality and will drive down the requirement for persistent storage of derived products. Data Cubes, or derived products could then be considered transient caches and spun up as required at small cost rather than persisting them.

Table 4-1 summarizes the state of several data architecture options and their ability to accomplish key user requirements and user applications. This matrix is meant to compare traditional approaches to data storage and distribution (scene based methods) and newer innovative approaches (pixel-based methods). It is assumed that pixel-based methods can take advantage of subsetting (e.g. only acquire the spatial and temporal data needed) and compression. A summary of the assessment is found below the table.

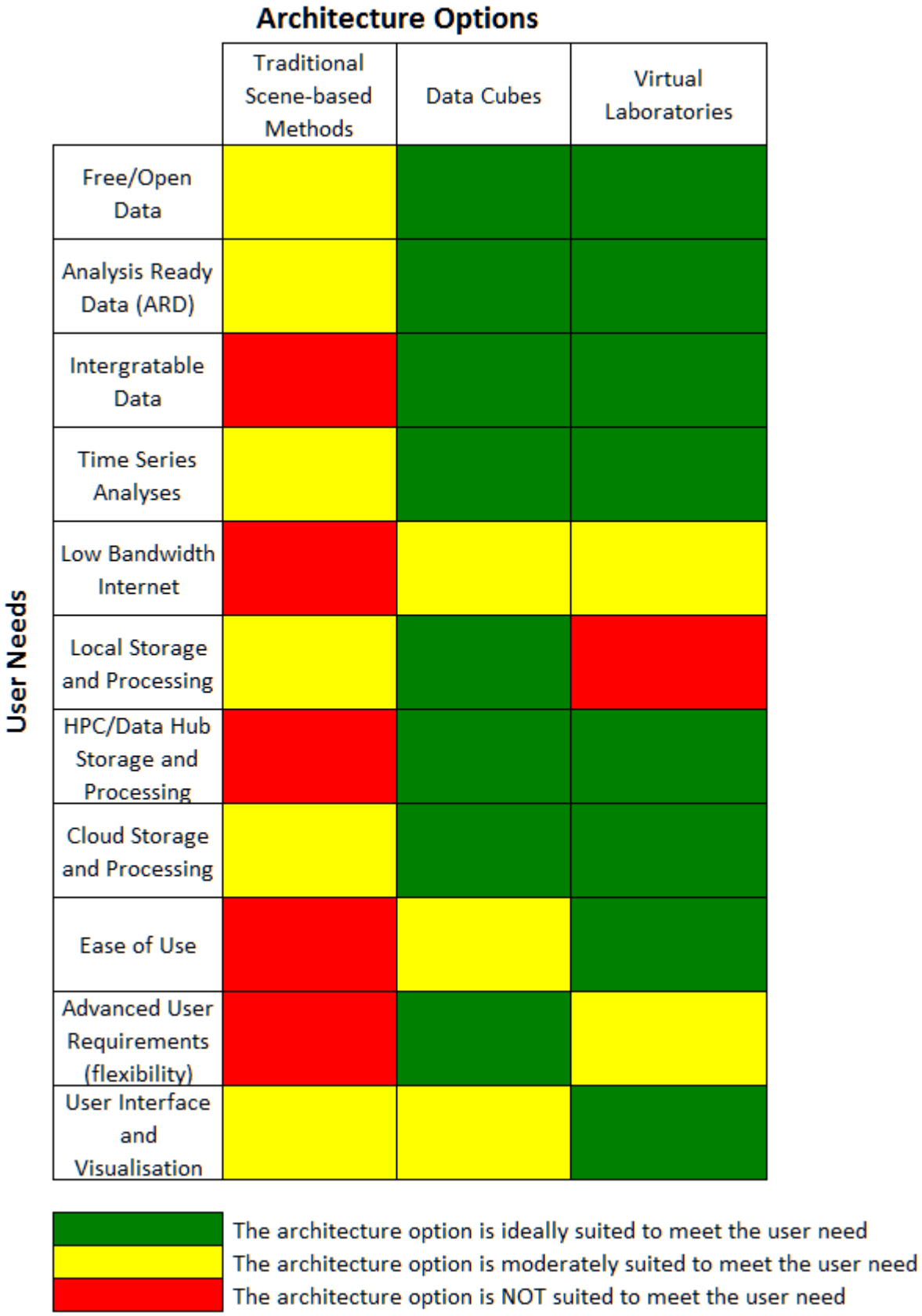

Table 4-1: Assessment of User Requirements and User Applications vs. Data Architecture Options

Bringing the user to the data: Earth Observation Virtual Laboratories

EO Virtual Laboratories (VL), accessible through web browsers, virtual machines, mobile applications and other web-based or machine-to-machine interfaces, are one means to address the objectives above. Such platforms are virtual environments in which the users - individually or collaboratively - have access to the required analysis ready data sources and processing tools, as opposed to downloading and handling the data 'at home'. A key quality of such platforms is that they are shaped by and scalable according to the needs and ambitions of users, they co-locate data collections and computational capacity and they are composed of a range of flexibly interconnected services, often federated, allowing substantial tailoring and re-use. These VL platforms are typically implemented in the "Cloud" connecting hundreds to several thousand computer nodes across a network of data centres or in regional hubs with High Performance Data capabilities. Such EO VL platforms are intended to bring together the following main functionalities:

- data for both EO and non-EO applications
- powerful computing resources
- large-scale storing and archiving capabilities
- collaborative tools for processing, data mining, data analysis
- concurrent design and test bench functions with reference data
- high-bandwidth web-based access
- application shops and market place functionalities
- communication tools (social network) and documentation
- accounting tools to manage resource utilization and flexibility in free or pay-for-use business models
- security and privacy enforcement

The term Virtual Laboratory, isn't universal nor exclusive and there are many other valid names for such systems. The phrase is used here to simply convey the set of characteristics that are common to such environments, regardless of what name they are referred to. With this definition the AGDC when combined with suitable web services interface, the CEOS SEO Data Cube, and the European Thematic Exploitation Platforms (Fig 4-1) are all valid examples with many, if not all, of these characteristics.

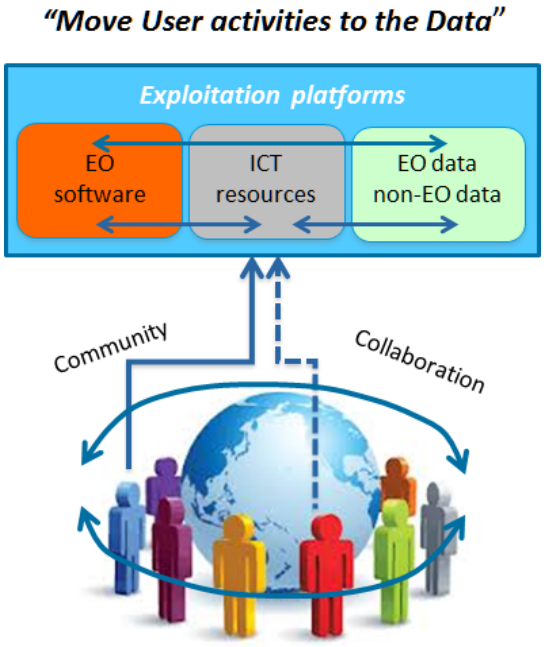

FIG.4-1 - EO Community Platform Concept. Image courtesy of ESA.

Conceptually, this approach is paving the way for an "Information as a Service" scenario (Fig.4-2). EO and non-EO data are flexibly and intelligently linked and combined by means of modern ICT services ("Infrastructure as a Service", "Software as a Service", "Data as a Service", etc.) thus increasingly integrating the EO sector into the overall digital economy. The inherent challenge in such an approach is the orchestration of heterogeneous systems, data sets, processing tools, and distribution platforms leading to the creation of innovative and high-quality information services for a broad range of users.

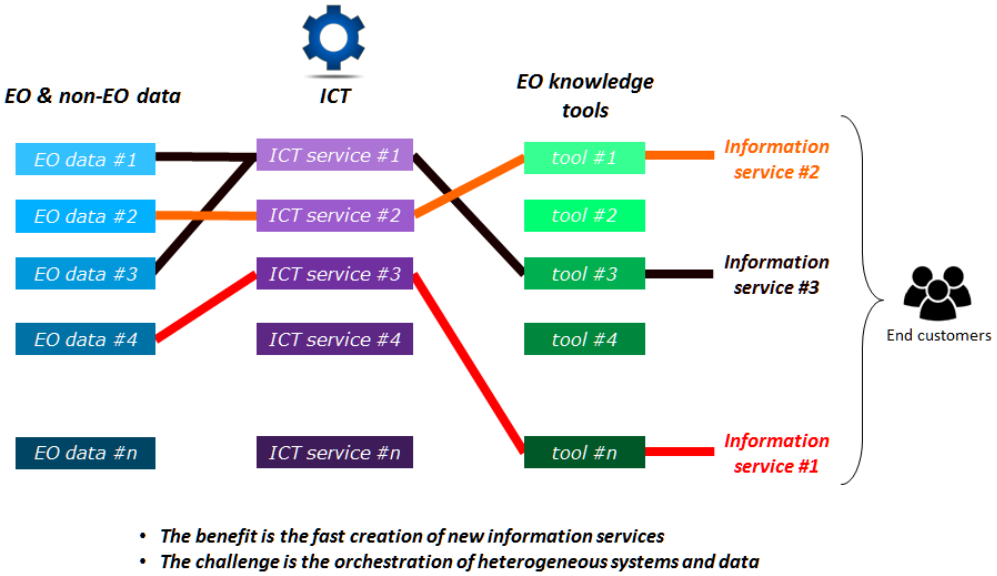

Fig.4-2 - Model for the generation of "Information as a Service". Image courtesy of ESA.

In such environments, evolution is driven by user communities and their needs. Ideally, user communities will have access to a fully scalable IT-infrastructure enabling them to develop new business models and to introduce new applications and services. Crowdsourcing Platforms, which in some cases are already providing significant "Citizen Science" output, may provide a number of relevant pointers in this context.

Europe is moving in this direction through the implementation of the "EO Innovation Europe" concept in coordination between the European Commission, ESA and other European Space Agencies, and industry. The "EO Innovation Europe" concept has the objective to enable large scale exploitation of the comprehensive European EO data assets for stimulating innovation and to maximize their impact. Similar architectures exist in the US, Japan, and outside space agencies in the broader CEOS community with the CEOS SEO Data Cube project developing such platforms for Kenya and Colombia and the Australian Geoscience Data Cube combining multiple agency collections observing Australia from the US, Japan and European missions into a single platform for broader use. Taken together these activities, whilst still in development, illustrate much of the future of EO data architectures with improved accessibility and greater distribution of computation empowering diverse end-user applications.

Architectural change

Architecturally these EO VLs are a network of interoperable interconnected platforms built around core-enabling elements, open to multi-source funding initiatives or implementation by independent organisations and relying on common standards for integration. The commoditization and separation of core-enabling elements beyond a single agency boundary supports scalability and empowers end-users and industry to value-add and allows Space Agencies to utilise large scale commercial infrastructures (e.g. Cloud computing) when it is more cost effective to do so.

Figure 4-3 illustrates the EO Innovation Europe approach where diverse institutional and commercial platforms already exist or are currently being implemented. Based on functional analyses and identification of best practices, EO- Innovation Europe has been structured around three elements: an enabling element (acting as a back office), a stimulating element and an outreach element (acting as a front office).

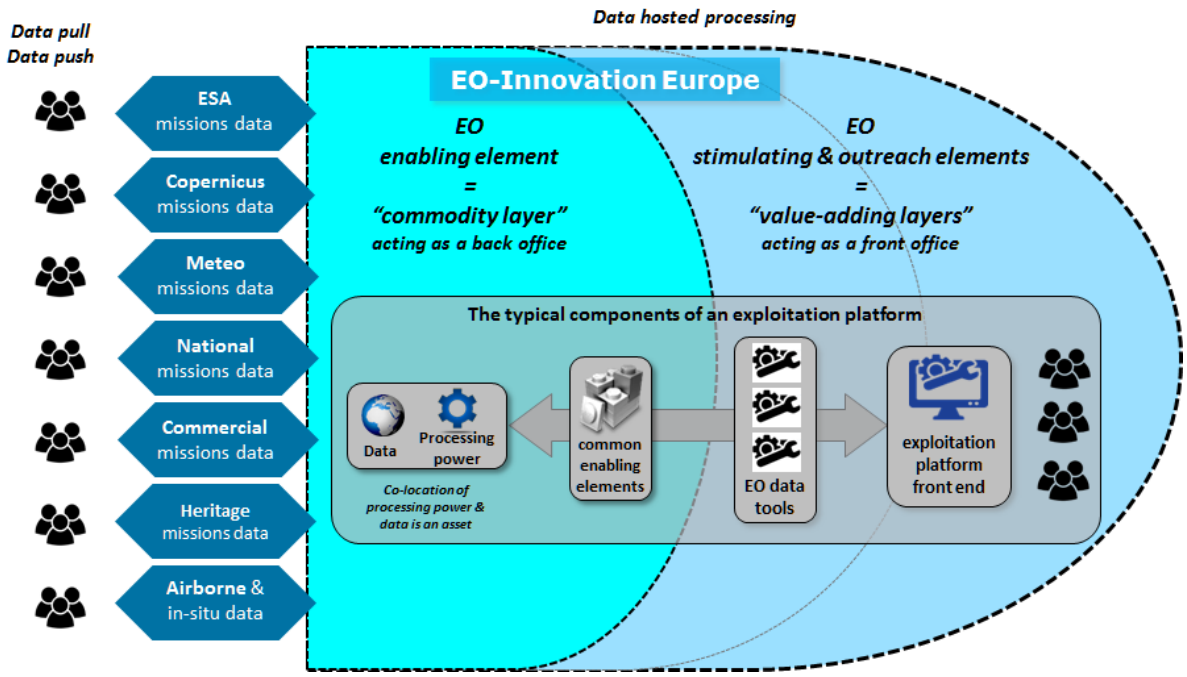

Fig.4-3 – EO Innovation Europe. Image courtesy of ESA.

Within the enabling element (back office), a "mutualisation" (i.e. sharing) of efforts and funding between public institutions should prevent an unnecessary duplication of investments for enabling infrastructures and will stimulate the existence of many exploitation platforms or value-adding add-ons funded by different public and private entities in the outreach element (front office). Compared to existing EO architectures rather than supply an end to end product or base data, the Space Agency provides the enabling element and suitable data preparation and interfaces to allow dynamic use by third-parties for research or industry exploitation as required. The ability for third-parties to quickly and flexibly create new value chains and provide innovative services over the generic enabling elements is expected to significantly enlarge the user base.

The change to a more distributed architecture, both in terms of components and participation, is accompanied by several architectural principles many of which are already evident in the trends discussed in Section 2:

Discovery and Access

1. Machine Level Discovery and Access: All data are available for search and access with machine-callable APIs
2. Cross-agency Discovery: Cross-agency data discovery is seamless at a pixel based access level (pixel and attribute level sub-setting)
3. Dataset Selection Guidance: Guidance is available on data selection based on fitness for purpose.
4. Metadata Naming Conventions: Key metadata follow standard naming conventions for Variables, Platforms, Instruments, Spatial Resolution, Temporal Resolution
5. Virtual Collections: Virtual collections can be organized / oriented around a science problem or Theme (eg. hurricanes, agriculture, algal blooms, fires) containing multiple sensor collections rather than single sensor based collection (eg. Landsat,

Sentinel 2). The NASA ESDSWG Virtual Collections Working Group has explored ideas about such "Virtual Collections"
6. Analysis Ready Data: Preparation and distribution of trusted, calibrated, well documented data from multiple sensors in an analysis ready form for land, inland water, coastal and ocean applications.
7. Changes to metadata related to discovery and quality to enable pixel-based retrieval and fine grained query of very large data collections (e.g. % of cloud cover for my region and sensors of interest, rather than % of cloud cover in a scene file).
8. Web based seamless visualisation and browsing of entire collections

Usage

1. Intelligent Tool Catalogs: Intelligent tool catalogs automatically suggest data analytics / visualization tools to work with the data.
2. Live Data Citation: Publications are linked to data and tools that allow interactions with the data.
3. Mobile Data and Processing: Data and processing move transparently as necessary to achieve optimal performance.
4. Quantitative Quality: All data have quantitative measures for data quality.
5. Reproducibility: Scientists can reproduce other scientists' research results with high precision.
6. High-Quality Documentation: Concise, Comprehensive and Consistent documentation exists for all data variables.
7. Capacity Building: A rich set of capacity-building and translation mechanisms exists to facilitate leveraging data for use by people with limited literacy in science and advanced technology, and/or English.
8. Data Analysis at Scale: Users are able to analyze the entire data record for any data variable over any arbitrarily defined area.
9. Dataset Upgrading: High-value datasets are upgraded as necessary to fully support in the rich capabilities available in the data systems.

Integration

1. Data: EO data can be easily compared, merged, fused and/or assimilated with (depending on the application) data from other agencies, nations and other entities. This not only requires clarity and quality of geospatial and temporal characteristics but comparable calibration on geophysical observations (e.g. in using two different optical satellites with similar sensor detection wavelengths).
2. Tools and Services: Tools and services within the community are easy to use in concert and co-hosted with large EO data collections
3. Sharing: The EO community are able to share all scientific resources (data, tools, results, workflows, contextual knowledge)
4. Standardisation of programming interfaces and vocabularies across sensor types to better support data integration, discovery, and analysis.

Infrastructure

1. Hosted processing infrastructure with EO toolboxes (visualisation, analytics) and data available for Industry and research use
2. Virtualisation of hardware and software services to support “pay for what you use” scalability
3. Use of Open Source software licensing as a mechanism to support innovation by third parties building out from agency and research supplied tools
4. Consolidation of common architectural components across collections (same delivery mechanism/experience for all sensor types)
5. Automation and acceleration of the data preparation life cycle (calibration, atmospheric and terrain correction) to support near real time analysis

5. Conclusions

CSIRO, as CEOS Chair for 2016, established an Ad-hoc Team on Future Data Access & Analysis Architectures (FDA) to survey the challenges and opportunities around EO data architectures given the operating environment in which government sponsored EO programmes are working.

Progress to date has established that CEOS should intensify strategic efforts in relation to FDA if EO is to realise its full potential in support of society, including making best use of all available CEOS agency data so that:

- dense time series analysis and applications are made feasible for all users, particularly in the 'land domain' but not restricted to it given the interactions between domains. ;
- CEOS efforts in relation to the grand challenges of climate, food security, the Sustainable Development Goals (SDGs), and disaster mitigation can be supported through application of EO data and products by users across sectors and nations.

The first conclusion of the team's 2016 efforts is to recognise that this activity is very much needed and indeed overdue within CEOS. All CEOS agencies recognise the need to do more to remove obstacles to data access, analysis, uptake, application and realisation of benefits to society. There is significant activity across CEOS agencies in this area with great diversity in approaches and capacities. This means there is a lot to build on; it also means that moving forward in a coordinated way will be more difficult.

The team would like to highlight key trends that are compelling action by space agencies in the area of FDA (presented in more detail in section 2 of the report):

- The move to fully 'on-line' data systems, plus the increased size and complexity of the data being produced (referred to in the report as volume, velocity and variety of data) is overwhelming traditional approaches to data architectures;

- The Big Data players and their advanced platforms, populated with CEOS agency data (amongst others), are changing expectations as to how easy it could and should be to access, analyse and apply EO satellite data;

- These new capabilities are providing a welcome broadening of the user base for EO satellite data, to more sectors and more users, many or whom are non-expert, and/or not from large technical institutions. These efforts are demonstrating that up-front effort to remove the obstacles to EO data handling and use are paying dividends to the mainstreaming of EO data and to its societal impact. CEOS agencies must take note as to the implications for ease of data handling for CEOS agency mission data;

- CEOS has been placing more emphasis on supporting uptake and application of data, including for grand themes like the SDGs, climate, and food security. These initiatives, such as GFOI, have reported difficulties in user uptake due to complexity of data access and handling, and the changes in user expectations from exposure to advanced platforms of commercial big data companies. Users are seeing solutions to traditional obstacles to EO data access and application and expect space agencies to adopt them.

- Evolution in the application space is creating more demand for capability to bring together different datasets to answer complex questions, over large areas, over long periods, and in combination with other (e.g. socio-economic) non-space data; in the terrestrial domain in particular this is far from easy to do.

CEOS has, over the years, invested considerable effort into cooperation in interoperability of data discovery and access. Effective future cooperation should extend this to encompass common work on user-data interaction, facilitation of data integration/interoperability, and compatible service interfaces for analysis. CEOS efforts should reflect the trends around increasing use of 'in-place' Analysis Ready Data to replace data discovery and download as a response to the 'volume, velocity, variety' challenge.

Individual agency strategies are quite diverse and include:

- bringing the user to the data, in contrast to trying to transmit large amounts of 'raw data' with the associated communications, storage and data management problems becoming a huge 'barrier to entry' for users;

- APIs/Virtual Laboratories, enabled by standards (such as Thematic Exploitation Platforms) that make it easier for the work of some to be integrated and linked with the work of others, to accelerate progress;

- pre-processing data to a point where it is a measurement comparable in space and time with measurements from both other satellite instruments and other sectors, helping isolate applications (and users) from non-relevant (to their application) changes in the space segment (i.e. sensor agnostic);

- novel service models (including new opportunities to integrate commercial data 'on the fly' to augment products primarily using public sector data);

- moving the burden of data preparation processing (calibration, location, atmospheric correction etc) for the extraction of application information from the users to the space agencies (such as Analysis Ready Data); and

- flexible approaches to computing (including HPC and Cloud) that showcase the ability to process multiple observation datasets, at full resolution, at continent and global scales.

Future Data Architectures can be assumed to consist of multiple approaches to cover all circumstances and uses. The Ad-hoc Team has not sought to judge individual agency approaches but instead to identify common ground for effective cooperation that will:

- deliver benefits back to Agency activities;

- lay the foundation for Agencies to work together, through CEOS, to offer a more integrated 'way in' to satellite Earth observation data analytics for users and global intiatives.

A number of activities involving core technologies and examples are already underway as pilots initiated by GFOI, the CEOS SEO and LSI-VC, in relation to Analysis Ready Data and the CEOS Data Cube. In addition, there is ongoing fundamental work within WGISS around metadata standards, data provenance and preservation, and ongoing technology exploration.

The technology and infrastructure behind FDA is changing rapidly and the challenge for CEOS will be to establish a programme of work to take advantage of these opportunities – with an emphasis on approaches that de-risk and simplify EO data for users, allowing users to make use of ALL available and relevant CEOS agency data, and that will support CEOS ambitions in relation to its chosen grand challenges. Future work should help users benefit from all relevant CEOS data at all stages – complementing past efforts which have emphasised unity of 'discovery', without fully addressing the significant challenges in the *analysis and exploitation* of disparate or incompatible data after discovery.

Recommendations

One year did not permit sufficient capacity or time to be brought to bear on what has been found to be one of the biggest strategic issues for CEOS Agencies and for CEOS as their forum for international coordination. The issues are complex, and more time is required to establish consensus among CEOS agencies as to the most productive way forward for CEOS as a coordinating body - reconciling the various agency approaches and priorities whilst finding common ground for co-operation.

As a consequence, the team recommends that CEOS adopt a two-stream approach to continue its engagement with this topic:

- a further year of work to continue exploration of key areas, and for facilitated strategic discussions;

- parallel efforts to progress established CEOS pilot projects to ensure strategic discussions are supported by real-world evidence.

This report therefore focuses only on short-term recommendations.

1. USGS and CSIRO have both agreed to continue to provide leadership for the Team, which will seek a 1-year extension to its operation as a CEOS Ad-hoc Team at the November 2016 Brisbane CEOS Plenary. The AHT recommends that a third co-chair be identified from ESA/EC to contribute to the formulation of a recommended way forward for CEOS in 2017, since the free and open global data programmes of both the US and Europe will necessarily be at the heart of any strategy ultimately adopted by CEOS. It is also critical that other Agencies actively engage also, as the way forward must provide opportunities for all Agencies to contribute and benefit. The

AHT recommends that CEOS Plenary direct the Team to conclude its work in time for debate and decision at the USGS-hosted Plenary in October 2017.

2. In parallel with the conclusion of the AHT study and report in 2017, CEOS should progress, accelerate, and integrate the pilot activities already underway within its subsidiary groups, including and in particular:

   - LSI-VC work in relation to definitions for CEOS Analysis Ready Data (ARD) for Land (CARD4L) and guidelines for its use; and
   - SEO/SDCG work in relation to demonstrations of a CEOS Data Cube and its benefits for both data providers and data users, and as a way of engaging with donor institutions on practical capacity development projects.

   The AHT recommends that 2017 efforts be designed around the following objectives:

   - A small-scale demonstration of the production and application of CEOS ARD and its value to both data providers and data users, drawing on key free and open global programmes; this should help all CEOS agencies develop an understanding as to the nature and scale of the activity of data production and integration – and importantly should incorporate a feedback loop from engaged user organisations and funding agencies as to lessons learned and the benefits for them; and should emphasise what space agencies get back by way of data uptake as the motivation for expansion of the concept of CEOS ARD;

   - Continued development of the CEOS Data Cube as a widely-supported, open source example of the benefits of HPC approaches to support ease of use of EO data. Significant work is already underway, documented in an informal 3-year Work Plan prepared by the SEO, and includes a pilot demonstration with the Government of Colombia for GFOI (and other) purposes. This pilot could be employed as the practical application focus for the ARD production trial – combining multiple optical and radar datasets using the CEOS ARD definitions, in a framework that would guarantee user feedback and practical lessons learned. The SEO and SDCG have both indicated a willingness to support such an integration of existing CEOS activities relevant to FDA. The work would demonstrate the benefits of dense time series data for GFOI purposes, amongst others.

   This approach would integrate different aspects of CEOS FDA work in an application focused activity that would guarantee user feedback and develop CEOS experience around the production of, and benefits from, ARD. The AHT could provide the necessary integration effort for these efforts in 2017 to inform the design of the CEOS FDA strategy whilst contributing activities would continue under leadership of LSI-VC, SEO, and SDCG, with expert input from groups such as WGISS and WGCV. Such a demonstration would likely require 12-24 months. It could effectively build on the existing foundations underway within GFOI/SDCG and could adapt to the final FDA Report recommendations in late 2017 as needed.

   The SEO, as the technical lead for this initiative, should be assured support from a core group of Agencies willing to contribute to evolution of the technical design and development of new capability.

3. CEOS should continue support for ongoing WGISS work in relation to: discovery search engine optimization (search relevancy, keyword search, persistent identifiers); access common standards for interoperability of product formats (metadata/data) and application program interface (API) for analytics and data access services; exploration of emerging big data services including cloud computing. WGISS should continue to provide guidance notes and best practices that agencies can take on board when planning future investments.

4. In the course of the 2016 work of the AHT further suggestions for pilot activities have emerged, including notably a proposal from ESA for a *CEOS Thematic Exploitation Platform for Disasters*. The leadership of the AHT should establish the available capacity and leadership for this and any other relevant proposals emerging from the work, and bring to CEOS (at the key meetings of SIT or Plenary) suitable proposals for debate and further development as appropriate. Such proposals might be evolutions of existing work and within existing groups or might merit establishment of new initiatives and the means to manage them. Any new initiatives should be developed in accordance with the CEOS Process Paper. USGS, as incoming CEOS Chair has already proposed to undertake 2017 activities in relation to interoperability of moderate resolution optical data products and these can contribute to the above and emerging proposals related to the FDA work.

5. Future CEOS decisions on a strategy around FDA issues will be both strategic and sensitive since the broader context involves activities and competitiveness of national/regional industries and companies in relation to EO data uptake and applications. Yet the consequences are so central to the future success and health of government-sponsored EO programmes, and to the effectiveness of CEOS as a coordination body with significant user-facing global initiatives, that CEOS must identify a workable cooperation path of common interest to its agencies. The AHT recommends that extensive CEOS Principal input be assured in the development of the final report on AHT matters and that CEOS Chair and SIT Chair cooperate in 2017 to ensure that the way forward is formulated in light of inputs from both working-level CEOS contributors (like the AHT) and those from CEOS Principals and Leadership.