



# Global Data Flow Study

for the  
Global Forest Observations Initiative

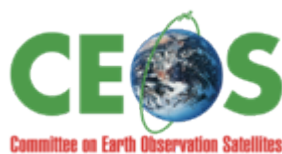
Version 1.0  
October 2016

Committee on Earth Observations (CEOS)  
Ad-hoc Space Data Coordination Group (SDCG)

**Global Data Flow Study for  
the Global Forest Observations Initiative (GFOI)**

---

Version 1.0  
October 2016



## Issues for Plenary Discussion and Decision

1. The Global Data Flows report from SDCG is provided for information and reference of CEOS. Agencies are asked to note the implications for agency and CEOS work in relation to Future Data Architectures.
2. In particular, SDCG notes the perceived importance of Analysis Ready Data to the study respondents and commends CEOS efforts to accelerate the definition and development of CEOS Analysis Ready Data. User-facing programmes such as GFOI will benefit significantly from this work.
3. SDCG supports the progress to date on the Future Data Architectures initiative and the new initiative on the interoperability of moderate resolution data products and stresses the strategic significance of these for the future success of government-sponsored EO programmes and the effectiveness of CEOS as the primary coordination body for those programmes.
4. SDCG commends the individual recommendations to all CEOS space agencies and to the various GFOI stakeholders. SDCG will pursue these recommendations through GFOI channels.

<b>Issues for Plenary Discussion and Decision</b>	<b>2</b>
<b>Executive Summary</b>	<b>4</b>
<b>1 Introduction</b>	<b>7</b>
1.1 Overview	7
1.2 Purpose	8
1.3 Context	9
1.4 Contents	9
<b>2 Global Forest Observation Scenarios</b>	<b>11</b>
2.1 Scenario Components	13
2.2 Analysis Ready Data (ARD)	14
2.3 Scenarios	16
<b>3 Scenario Data Architecture Description</b>	<b>20</b>
3.1 Transmission	20
3.2 Storage	20
3.3 Discovery and Download Tools	21
3.4 Processing	21
3.5 Forest Map Production	22
3.6 Costs	22
<b>4 Evaluation</b>	<b>23</b>
4.1 Criteria	23
4.2 Process and Cost Impact	24
<b>5 Conclusions and Recommendations</b>	<b>27</b>
5.1 Conclusions	27
5.2 Recommendations	29
<b>Appendix A Acronyms</b>	<b>33</b>

## Executive Summary

This study was conducted by the space data component of the Global Forest Observations Initiative (GFOI), and considers barriers to the effective use of satellite data in support of National Forest Monitoring systems (NFMS) and REDD+. A recent increase in satellite capacity, with the launch of significant new Landsat and Sentinel missions, amongst others, has resulted in access to large volumes of data, which presents significant challenges in data handling, as well as opportunities and challenges to advancing MRV methodologies.

### Opportunities

- a) The increased number of satellites provide a dense time series which enables new opportunities to improve change detection and classification of global forest cover.
- b) Access to multiple optical and SAR data flows provide alternate solutions for complex ecological systems, such as degraded or multiple use cases, and persistently cloudy regions that cannot be observed reliably with a single sensor.
- c) Generation of specific data products that assist with detection of land cover change and meeting MRV reporting requirements.

### Challenges

- a) Higher data volume increases storage and delivery cost.
- b) Increased choice adds complexity to data discovery and selection.
- c) Multiple sensors make geo-registration and cross calibration computationally more demanding.
- d) New methodologies are needed to benefit from increased data volumes and increased sensor complexity.
- e) Acceptance and capacity to uptake of new methodologies by country authorities and agencies.

### Scenarios

The study considers how challenges and opportunities present within the business-as-usual (BAU) scenario as well as two alternative scenarios. Factors considered include the level of satellite data products, where products flow from (space data providers) and to (users), and where the work of intermediate and final forest products production takes place.

Production Stage	Business As Usual (BAU)	Option 1	Option 2
Satellite product pre-processing level	Level 1 products, for optical typically Top-of-Atmosphere (TOA)	Analysis Ready Data (ARD, see Section 2.2), for optical, typically surface reflectance products	
Satellite products flow to	Unstructured satellite product flows based on user download preference, typical data processing and	In country agencies	High volume satellite products downloaded to processing hubs maintained by technology or capacity

Production Stage	Business As Usual (BAU)	Option 1	Option 2
	delivery mechanism to date		building partners ↓
Intermediate forest products produced at	In country agencies	In country agencies	In country agencies remotely produce more compact national forest products from satellite products ↓ Forest products are downloaded in country for further analysis
Final forest products produced at	In country agencies, where in country agencies control the selection of the source data, process, and the results Restricted distribution data (e.g. commercial) delivered to the in country agencies or to a capacity building partner		

Table 1 Global Data Flow scenarios

A number of components of NFMS systems were considered, including space data provider pre-processing and ground segments, data infrastructure, the roles of capacity building organisations and supporters, as well as sustainability.

The study considers data flow volumes and rates for the 70 GFOI countries, and it was found that for the average of these countries, it would take approximately 19 days to download 1 TB of data (approximately 500 Landsat scenes). And for these countries, the average volume of a national wall-to-wall coverage required by the GFOI Methods and Guidance (MGD) would be 2.8TB. These volumes only increase when considering the addition of Sentinel-2, where the number of bands is increased, and the spatial resolution for the optical bands is finer.

## Conclusions

The study reached several conclusions as a result of its investigations, summarized below.

- The increase in satellite data volumes resulting from new capacity being brought online by space data providers is outstripping the capacity of the national data handling infrastructure of GFOI countries.
- The BAU approach with unstructured data flows is considered unsustainable, and in general a move towards more centralized data handling and processing will be required in order to make satellite-data support NFMS systems sustainable.
- The increase in data volumes and number of data sources requires more efficient and effective data discovery and access tools.
- The cost and burden of pre-processing satellite data needs to be minimized to enable their uptake within NFMS systems. Creating Analysis Ready Data (ARD) products

directly by space data providers or using their tools are steps in that direction. The specification and mechanisms for delivery of ARD products will vary, but the data quality and interoperability of the ARD products needs to be validated by the space data providers.

### Recommendations

Based on its conclusions and to meet the challenges listed above, the study has made a number of recommendations for consideration by space data providers, as well as countries considering the design of their NFMS and capacity building partners and donors supporting these efforts. These recommendations are summarised as follows.

- CEOS space agencies and capacity building partners must move toward the support of higher level **analysis ready data** products including global pre-map products to reduce cost for the use of satellite products by countries.
- Improved **data discovery and access tools** are required to improve the uptake of satellite EO data for forest monitoring.
- **Interoperability** among the core data streams would be of significant value to forest monitoring users with an initial emphasis on Landsat-8 and Sentinel-2.
- CEOS and its member agencies should support the development of a **model pilot end-to-end NFMS based on**, and collaboration with, the GFOI components.
- CEOS space agencies, capacity building partners and donors, and country agencies must **invest in new methodologies and pilot activities that develop and demonstrate** the use of dense time series to improve future forest products.
- CEOS should consider strategies for partnerships with **donor bodies, UN agencies, and 'data giants'**.
- Where appropriate, CEOS agencies should **promote GFOI methods and guidance** within forest monitoring programmes sponsored by their national governments.

## 1 Introduction

### 1.1 Overview

This document considers the barriers to effective use of satellite data in implementing measurement, reporting and verification (MRV) within national forest monitoring systems (NFMS) in support of REDD+, and evaluates different solutions for reducing or removing these barriers and ensuring efficient global flows of the data.

The launch of Landsat-8, Sentinel-1A and -1B, and Sentinel-2A, and resulting access to large volumes of data at no cost to the user, changes the environment within which REDD+ countries access and analyze satellite imagery. However capacity is often lacking within countries to download, handle, and process large amounts of data. A number of initiatives have generated global scale forest information products (e.g. Global Forest Watch), however there is often hesitation by national governments to accept these products and national expertise is required to adapt these products to meet UNFCCC reporting requirements and IPCC guidelines. The rapid changes in data availability provide both opportunities for significant advances in MRV methodologies, as well as significant challenges.

#### Opportunities

- d) The increased number of satellites provide a dense time series which enables new opportunities to improve change detection and classification of global forest cover.
- e) Access to multiple optical and SAR data flows provide alternate solutions for complex ecological systems, such as degraded or multiple use cases, and persistently cloudy regions that cannot be observed reliably with a single sensor.
- f) Generation of specific data products that assist with detection of land cover change and meeting MRV reporting requirements.

#### Challenges

- f) Higher data volume increases storage and delivery cost.
- g) Increased choice adds complexity to data discovery and selection.
- h) Multiple sensors make geo-registration and cross calibration computationally more demanding.
- i) New methodologies are needed to benefit from increased data volumes and increased sensor complexity.
- j) Acceptance and capacity to uptake of new methodologies by country authorities and agencies.

To realize these opportunities and address the challenges, the Space Data Coordination Group (SDCG) needs to take a considered and phased approach with Global Forest Observation Initiative (GFOI) capacity building partners that meets countries immediate needs whilst also working toward long-term solutions. The following approaches are evaluated in this document:

- 1) Work with space agencies and partners to reduce the burden of data pre-processing on forest management organizations.
- 2) Work with space agencies and partners to implement new pre-processing, change detection, and classification methodologies.



- 3) Provide improved discovery and selection tools to assist countries in more effectively identifying required data before attempting downloads.
- 4) Provide tools for use in maintaining local databases through incremental updates as new data, that meet specified agency criteria, become available.
- 5) Acknowledge that simple and more appropriate solutions may exist that do not require an investment in expensive infrastructure.

A fundamental objective of this study is to help reduce barriers and increase efficiencies for the consistent and timely production of national forest remote sensing products, including products used in generating activity data for reporting greenhouse gas emissions and removals from REDD+ activities as defined within the land use, land-use change and forestry (LULUCF) sector by the United Nations Climate Change Secretariat. This may involve improved distribution of existing image-based products, new data intensive methodologies, and more efficient selection, discovery and access tools. An underlying premise is the vested interest in the REDD+ donor community to ensure that MRV requirements are met as efficiently and cost effectively as possible.

## 1.2 Purpose

It is acknowledged that with vastly increased volumes of data from satellites providing continuous global coverage, and the possible emergence of new methodologies requiring long and dense time series, the transfer of such data around the world is becoming increasingly unsustainable. Uncoordinated delivery of data is not an effective mechanism for addressing the demand or maintaining national archives - neither at facilities within a country, nor in the 'cloud'. Too frequently large volumes of data are delivered which are not needed or for which capacity does not exist to actively manage and use. Even in countries with access to high-speed networks and computational infrastructure, the construction and maintenance of these large datasets is a time consuming and costly exercise.

More efficient mechanisms are needed for constructing and sustaining large national databases, and computing infrastructure must be scoped to meet national requirements and budgets. Among infrastructure that needs to be considered are cloud computing based solutions, regional solutions, national solutions and solutions using global archives at data providers (e.g. CEOS, commercial partners) depending on the particular needs and capacities as national programmes develop.

This study aims to explore practical scenarios - building on multi-year collections of Landsat and Sentinel-2 data, including the complementary use of data from Sentinel-1, CBERS-4, ASTER, SPOT, ALOS for forest monitoring activities. Although the focus of the study is on freely available data, unique characteristics of commercial and other restricted distribution data is recognized. The study addresses the advantages and disadvantages of different architectures, technical elements, and implementation of data storage, handling, and processing tools to support the range of requirements of all GFOI countries. Capacity building partners and donors, including GOF-C-GOLD, FAO and SilvaCarbon, are central to the success of GFOI and are in many cases are the bridge to national partners.

It is understood that many of the data flow issues are institutional and cannot be solved only with better products or access. Forest products are often created within short-term projects, and not sustained or repeated systematically by existing forest or cartographic institutions. The lack of coordination among donors willing to support such products exacerbates the problem through inconsistent approaches, tightly defined requirements, lack of

understanding of country circumstances and internal dynamics, and short term funding. This lack of coordination often results in products that are not comparable, and blunts the desired cumulative effects of investing towards a 'national system'. Solutions to these institutional and coordination issues are beyond the scope of this document, but must be acknowledged.

While the focus of this document is on supporting GFOI and REDD+ countries, the conclusions have broader relevance for consideration by CEOS and its agencies in the context of next generation data architecture design, national data requirements, data volumes, bandwidth, processing capacity, 'analysis ready' data products, data cube storage architectures, national infrastructure, costs, and technical capacity.

### 1.3 Context

This study is undertaken within the framework of the 2016-2018 Year Work Plan of the CEOS SDCG for GFOI. Outcome 2 of the Work Plan is the identification and implementation of "efficient and effective global flows of data to accommodate in-country development of GFOI recommended Forest Map products."

This study was defined and executed by the CEOS SDCG for GFOI to respond to a number of important issues that the SDCG identified in the course of its journey to this point; issues that may equally apply to other CEOS initiatives and with significant and strategic implications for the way in which CEOS might approach these initiatives and for enabling work that may be required in support of their common resolution across all the CEOS initiatives.

This realisation led the SDCG to seek to better characterise the obstacles that countries face in managing and applying EO satellite data in this new era of data-rich land surface imaging from space, and to explore the pros and cons of different solutions to these obstacles and how CEOS might help coordinate these solutions.

The study incorporates feedback based on the practical experience of a diverse number of countries with interest in forest monitoring, including: Colombia, Kenya, Mozambique, Nepal, and Cameroon. Feedback was also received from key capacity building partners: World Bank, FAO, and SilvaCarbon. And from practitioners in industry that have broad in-country experience in the development of national MRV.

The primary purpose of the CEOS Global Baseline Data Acquisition Strategy for GFOI is to assure the acquisition of the minimum required satellite data for countries to fulfil their national reporting requirements for forest-related greenhouse gas emissions and national forest carbon stocks to UNFCCC and REDD+. Supporting users in the application of this data, building on the success of the Acquisition Strategy, is one of the main purposes of this study.

### 1.4 Contents

This study compares the main parameters and costs of a 'Business as Usual' (BAU) data architecture to alternate architectures and draws conclusions as to the advantages and disadvantages, and lessons for future data architectures of the main data suppliers and of global initiatives such as GFOI. A sense of how the global data flows may operate under alternate future scenarios is provided. The study takes into account practical considerations

identified from the pilot work the CEOS SEO is conducting on behalf of the SDCG, and identifies areas for further consideration by CEOS through the *ad hoc* team on Future Data Architectures (FDA) established at the 29<sup>th</sup> CEOS Plenary (November 2015). Recommendations specific to GFOI will help inform and update the SDCG 3-Year Work Plan, as well as the plans of core data stream providers like USGS and ESA/EC.

**Chapter 2** provides reference points defined as a set of global forest observation scenarios.

**Chapter 3** describes system architecture alternatives available for implementation.

**Chapter 4** specifies the evaluation criteria for assessing benefits comparing BAU scenarios to cloud-based analysis ready data models, and presents the evaluation of the alternate scenarios.

**Chapter 5** summarizes the analysis and presents recommendations for efficient and cost effective global data flows for forest monitoring.

## 2 Global Forest Observation Scenarios

Three current and proposed global data flow scenarios have been evaluated. These are summarised in Table 1, and the different linkages are shown in (see Figure 1). The BAU scenario describes the current mode of operation for many users, while the alternate scenarios (Option 1 and Option 2) are proposed to support cost savings and improved forest resource analysis and monitoring capabilities (e.g. via higher efficiencies).

Production Stage	Business As Usual (BAU)	Option 1	Option 2
Satellite product pre-processing level	Level 1 products, for optical typically Top-of-Atmosphere (TOA)	Analysis Ready Data (ARD, see Section 2.2), for optical, typically surface reflectance products	
Satellite products flow to	Unstructured satellite product flows based on user download preference, typical data processing and delivery mechanism to date	In country agencies	High volume satellite products downloaded to processing hubs maintained by technology or capacity building partners ↓
Intermediate forest products produced at	In country agencies	In country agencies	In country agencies remotely produce more compact national forest products from satellite products ↓ Forest products are downloaded in country for further analysis
Final forest products produced at	In country agencies, where in country agencies control the selection of the source data, process, and the results Restricted distribution data (e.g. commercial) delivered to the in country agencies or to a capacity building partner		

Table 2 Global Data Flow scenarios

The BAU option presumes the continuing delivery of Level 1 satellite products, for optical typically Top-of-Atmosphere (TOA), where Options 1 and 2 propose a move to Analysis Ready Data (ARD, see Section 2.2), for optical, typically surface reflectance products. The satellite products would be preprocessed to ARD by space data providers (e.g. CEOS

agencies, commercial providers). The main difference between Option 1 and 2 are the location of user data processing, data storage and data analysis (see Figure 1). Many other combinations and permutations are possible and likely, but those outlined in Table 1 have been selected to highlight the nature of the issues common to all and for ease of feedback from countries and capacity building stakeholders.

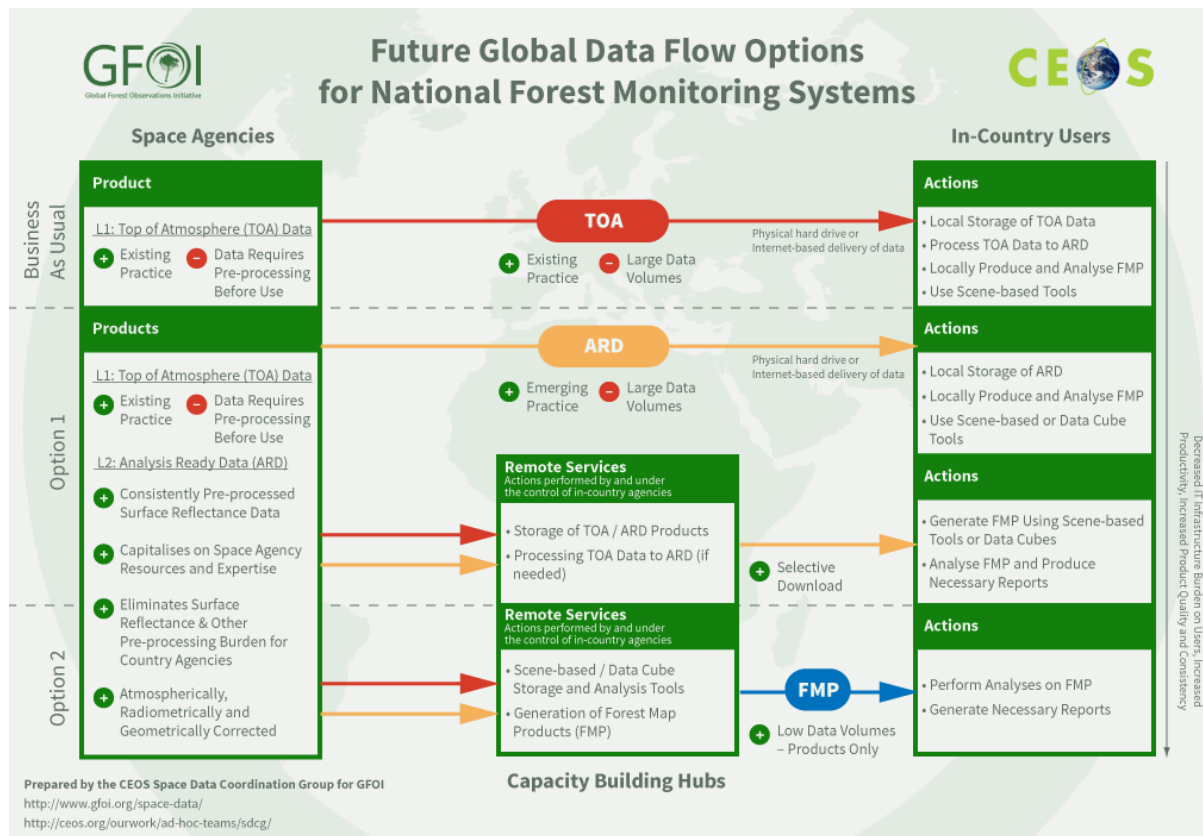


Figure 1 Three general data flow scenarios are evaluated. The difference between these scenarios depends on where the data processing is performed, where the data is stored and where the forest map product is created. Points of note: 1) Activities that take place in the Capacity Building hubs are performed by and under the control of the country agencies; 2) ARD products do not reduce data volume; 3) Data volumes are only improved through tighter specification of data needs and in the case of Option 2 creating the Forest Map Products at the data hub – resulting in a significant reduction in data volume; 4) the ARD removes the pre-processing burden from countries which will translate to significant cost savings; and 5) TOA negative is the cost of data pre-processing (in staff and time)

Each scenario includes functional components that are implemented using system architectures consisting of storage and computational infrastructure linked by transmission infrastructure. The functional components are outlined in Section 2.1 and the architectural alternatives are outlined in Section 3.

The scenarios are designed with components that are relevant for most countries. The criteria for success need to meet ‘typical’ country requirements. For GFOI, these criteria are designed to meet donor specified REDD+ requirements, augmented by countries that would benefit from more general solutions.

Expectations must be managed to achieve a sustainable solution that can be adapted by a country as its capacity increases. Significant risks are associated with maintaining infrastructure and expertise. A focus on the minimum data needed to satisfy the specific deliverables required by national and international funding sources is vital until stable

funding and staffing exist for upgrades. The scenarios presented provide some transitional alternatives for managing costs and risks.

National agency users control the data flow to meet their forest monitoring needs and requirements. Capacity building partners, such as FAO and SilvaCarbon, and technology partners such as the CEOS SEO, Google or Amazon Web Services (AWS), can support the delivery and processing of data to meet the requirements established by the national agency and donors. Capacity building hubs may include providers of high level products such as land use land cover or forest cover datasets, and may include collaboration with other national agencies. Application to countries will require scaling of data volumes, user skills, local infrastructure and budget.

## 2.1 Scenario Components

The evolution of data flows should consider the following components in support of a national MRV system.

**Data supply from the space segment** as of 2016, ongoing data acquisitions provide repeat global coverage with increasing capacity (and resulting data volumes), and increasing capabilities (and resulting complexity in data discovery and application, e.g. multiple optical and SAR data sources). In future, supply volume can be expected to grow, and diversify as new capabilities come online. (e.g. multiple SAR bands, hyperspectral imagery, LiDAR, data from UAVs).

**Space data provider ground segments** are increasing their capacity in response to the changes in data supply, though the management of these data flows is increasingly complex. Data volumes are presenting new challenges for ground segment capacity, including transmission to users via internet distribution. And the increased diversity of data types available is making archival data discovery and access, as well as pre-processing more complex. These changes mean the BAU scenario for space data provider ground segments does not scale, and future alternatives will need to be implemented.

The presumption is that space data providers (e.g. space agencies, commercial) will remain the stewards of the data, providing a full 'gold master' reference copy of the data, and serve as the primary source for data. However, commercial partners such as Google and AWS are increasingly becoming high volume secondary distributors of data to the user community. It is also presumed that as pre-processing data becomes more of a burden to users, space data provider ground segments will move 'closer to the end user'. In practical terms, this means processing of data from TOA to ARD (e.g. surface reflectance), and also possibly further classification of the data (e.g. into quality tiers), and assembly of the data (e.g. into time series formats).

These additional roles for space data provider ground segments have important infrastructure implications which need to be considered.

**National and user data infrastructure** solutions also need to evolve to enable end users to leverage these new data streams. A large variety of end user data management solutions are expected, ranging from single computers, nationally and regionally hosted data centres and processing hubs, to cloud computing infrastructure solutions. It is also presumed that end user infrastructure will evolve along with national programmes, and that different infrastructure solutions may suite different implementation phases. For example, national research and development may take place within a small computing laboratory

environment, initial pilot deployment may take place in a cloud computing environment to enable scaling up from sub-national to national processing, and then once infrastructure requirements are accurately scoped, dedicated hardware might be setup to support national needs on an operational basis. Many other scenarios could also be envisioned.

**Capacity building agencies, partners, and donors** are a vital component of national MRV systems for many countries – in particular those who lack an existing monitoring system or national capacity. Donors also play an important role in providing resources to support implementation, and often rely on and work through capacity building partners (e.g. FAO, SilvaCarbon). These partners are often critical in providing the expertise and standards needed for the implementation of MRV solutions.

**Sustainability** while not a physically tangible component, sustainability is an important cross-cutting outcome to consider when looking at how the various components work together. Multiple configurations of space and ground segment, end user infrastructure and short term capacity building and donor arrangements are possible – but without proper consideration, these components will not necessarily result in a stable and sustainable national solution.

## 2.2 Analysis Ready Data (ARD)

The ‘analysis readiness’ of a particular dataset depends on both the nature of the analysis to be performed, as well as on the processing expertise of the end user and the capacity of available infrastructure and supporting datasets. For the purposes of this study, the intent of defining Analysis Ready Data (ARD) is to consider how best to shift the pre-processing burden away from users and onto a more systematic solution using space data provider ground segments. It is acknowledged that some users will want to do this pre-processing work themselves, and that by its nature a ‘standard’ ARD product will not service all users.

Further work on the coordination and definition of ARD is ongoing within the CEOS LSI-VC, which is working on a more general approach. And in many ways, this is a natural extension of processes that have been ongoing for years within the community, for example as Landsat has moved towards the definition and generation of their standard ‘L1T’ TOA radiance and reflectance products.

Optical TOA products are often referred to as, “Level-1 data”. After further processing, TOA products are transformed into “Level-2 data”, using atmospheric correction to yield a surface reflectance (SR) product and has explicit requirements for geolocational accuracy. Likewise ARD products for radar are radiometrically and geometrically corrected (Table 1) to support interoperability with optical products and other geographically referenced data. For the purposes of this study, ARD will refer to these higher level products.

The transformation of data from TOA to ARD has two main steps. The first is the creation of scene-based radiometrically and geometrically corrected products that include terrain, atmospheric and viewing angle corrections (Table 1).

The second is the reorganization of the data into tiled structures optimized for time series analysis. Scene-based structures represent the sampling mechanism of the satellite. In tiled data structures, each geographic tile, represents the same block of data on the ground. The indexing capability supported by these tiled structures can be leveraged by access and processing tools. Tiled data structures can be shallow with only a few dates included, or deep with many dates included depending on the application, and may be constructed to

provide simple annual or seasonal cloud free datasets all the way to phenological models that use every available date. The suitability of scene-based, simple tiled, or data cube structures depends on national agency monitoring objectives and requirements. Dense time series data structures are often referred to as Data Cubes. These Data Cubes can be organized using physical tiles of data or can be organized as indexed scenes.

<b>Optical</b>	<b>Radiometry</b>	<b>Absolute calibration</b>
		Cross calibration
		Band difference adjustment
		Viewing angle correction
		Solar illumination angle correction
	<b>Geometry</b>	Atmospheric correction
		Systematic correction
		Orthorectification
		Projection
	<b>Pixel level metadata</b>	Image to image registration
No data		
Clouds Shadows Water		
<b>Radar</b>	<b>Radiometry</b>	Absolute calibration
		Radiometric correction for topography
		Normalization of cross track (near-far range) incidence angle variation
		Rain attenuation
		Systematic correction
	<b>Geometry</b>	Orthorectification
		Projection
		Image to image registration
	<b>Pixel level metadata</b>	No Data
		Water
Layover		
Shadow		
Land		
<b>DEM</b>	Elevation data are needed to correct Optical and Radar data and for forest monitoring	
<b>Global Forest Products</b>	Global forest maps are useful as “bootstrap” classifications and validation	

Table 3 Analysis Ready Data

Georegistered SR data are the foundational ARD product for optical sensors and different paths exist to derive these. Most space agencies produce TOA, also called “at sensor”, radiance or reflectance data products, while some produce higher level products. For example:

- USGS produces on-demand ARD SR products, also called top of canopy reflectance;
- ESA provides a toolbox to convert TOA products to ARD SR products from Sentinel-2 Multispectral Instrument data (systematic provision of SR products are currently under investigation within the Copernicus ground segment) using Thematic Exploitation Platforms (TEP); and
- JAXA provides ortho-rectified ALOS SAR products.



There is a trend toward the use and future distribution of tiled time series of ARD by space data providers. Open source ingest algorithms developed by the NASA SEO, GA, and CSIRO teams are under development. Tiled and data cube structures are equally relevant to both optical and radar products, but optical applications are more mature. Integrated optical and radar structures also provide future opportunities.

GFOI’s Methods and Guidance (MGD) and R&D teams, in partnership with capacity building partners and countries, may define higher-level ARD products specifically designed to meet the goals of GFOI. Example derived products may include cloud-free mosaics, greenness, wetness, bareness indices and products derived from algorithms such as the continuous change detection and classification of land cover. Existing forest maps also serve as *de facto* high level ARD input for the production of new forest products where local knowledge and data are used to refine regional or global products.

## 2.3 Scenarios

The three options outlined in Table 1 have been evaluated. Minimizing cost to national agencies requires the transition to ARD products, minimising the amount of data that needs to be transmitted, and improving access tools for dataset update and maintenance. Production of ARD products by national agencies is usually expensive so only ‘enough data’ should be acquired to create the forest map products. Tools need to be available to permit the maintenance of local archives, which includes mechanisms for tracking data and processing provenance. Partnerships among national agencies and projects may add initial cost but may decrease overall costs and create a cadre of national experts.

The key to success is building on shared methodologies, such as those defined within the GFOI MGD and a database maintenance organization when supporting sustainable and comparable forest mapping practices. The analysis recognizes that real world implementations will find a balance between these options based on data availability, skills, infrastructure, and cost.

### 2.3.1 Business as Usual (BAU) Scenario

The BAU scenario (shown below) sets the baseline for comparison. This scenario represents the state of satellite data flows in 2014-2015, when Landsat was the primary data stream with Landsat 8 data just becoming available. At that time broad usage of SAR data was very limited for forest applications.

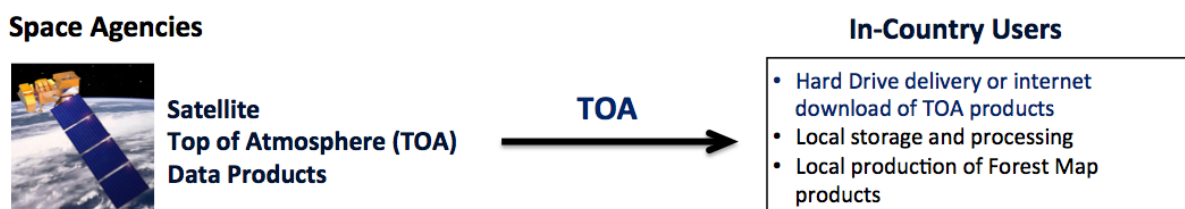


Figure 2 Business as Usual Scenario

Prior to Landsat 8, most data products could be delivered on physical media or over the internet with the assumption that countries had access to sufficient storage and processing infrastructure. Space satellite data were delivered to the country and then ingested,

processed and analysed to create the forest products using methodologies implemented in each NFMS. Most analysis tools used scene-based TOA radiance products. In large part this is the current model although in general the community is in transition to SR products.

The defining limitation of BAU scenario is the lack of an ‘organized’ data flow, methodologies and maintenance structure leading toward comparable forest products that can evolve to meet future requirements at known and controllable costs and quality. The skills to acquire and process the satellite data may exist in a different agency than the agency responsible for the forest management, further complicating data flows. Many current assessments are treated as independent events. New data are acquired and new analyses are performed with little inheritance from past data and practices to new forest maps. In some cases, this may include the delivery of a forest map by a research institute, non-governmental organisation (NGO), or commercial partner.

The BAU scenario requires countries to manage large and complex archives. In the past, large volumes of data have been delivered to countries with little evaluation of the requirements or the capacity to maintain the archive. Archives of space data are rarely static, and need to be reprocessed and updated with new data as calibration improves and new data become available. Managing the size of the archive is the most important step toward creating a data flow that can be implemented and maintained. The shift to ARD products and improved access methods will help countries keep archives current and comparable through time.

### 2.3.2 Option 1: Delivery of analysis ready data to a country agency

The current trend among space agencies is to shift toward the distribution of ARD. Option 1 (shown below) represents a national agency solution where Space agencies provide access to the satellite data, and capacity building partners (e.g. sustained by countries, supported by donors) serve as intermediaries, as space agencies transition to the distribution of ARD. These data will be delivered to the country agency via the internet or physical media (see Section 2.2).

These data will initially be scene-based, but tile-based solutions (e.g. Data Cubes) are expected to follow. Technology partners such as the CEOS SEO, Google or AWS may furnish access to infrastructure; however stable funding solutions need to be identified to move beyond the prototype stage. National information products will be created in the country NFMS, by national agency staff and used for further analysis. New data will be added to the national image database by national agency staff using automated algorithms.

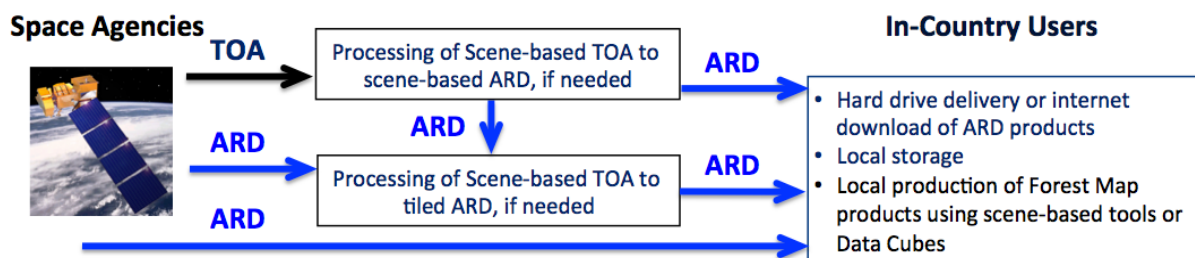


Figure 3 Scenario 1: Delivery of analysis ready data to a country agency

Option 1 does not reduce the data access requirements. However, the switch to ARD products will reduce costs to country agencies. The requirement to preprocess the data will

be removed (from the country agencies) and the skills needed to perform these tasks may not be needed at the country or partner agency. Tiled ARD products will provide additional efficiencies since all data for a geographic location can be immediately identified, supporting comparison through time and the creation of fundamental interpretable products such as cloud free mosaics.

Option 1 may represent a long-term solution for national agencies with sufficient processing, storage and internet infrastructure, as well as the expertise to support national requirements and a need for a national image database to meet broader requirements. Option 1 removes the burden of producing ARD, while providing maximum flexibility for the country users - including the growth potential to use future data cube application functionality.

### 2.3.3 Option 2: Delivery of data to a cloud or data hub for country agency access and analysis

Option 2 (shown below) represents a cloud computing or data hub solution where the data exists at a partner agency (e.g. FAO or other regional partner) hub or in the commercial cloud. Space agency data are downloaded to the hub or cloud where processing can then take place, if needed. The hub may exist on a commercial cloud or at a partner agency, such as FAO or other regional partner. The data hub infrastructure is maintained by the partner, with national agency access assured. National agency staff implement the forest application analyses and download the forest map results once complete.

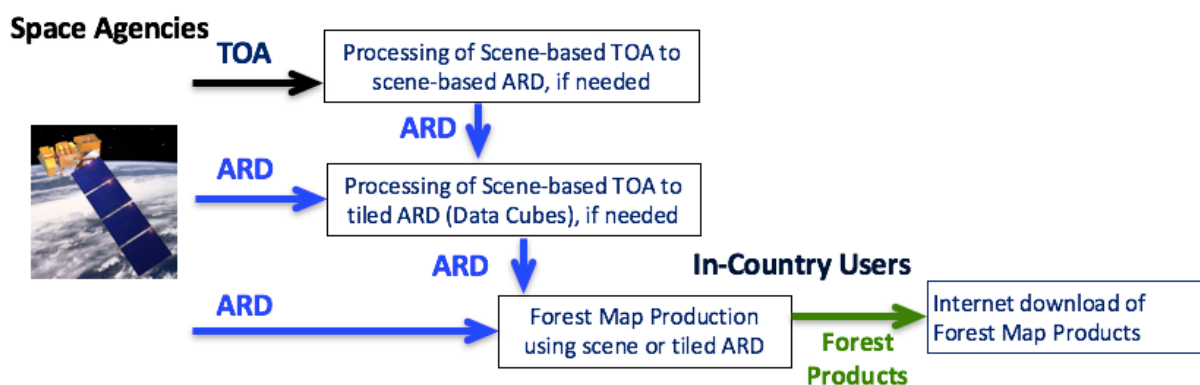


Figure 4 Scenario 2 Data delivered to hub for storage, processing with analysis results created and download by country

The space data (TOA or SR) are downloaded from space agencies to the data hub. If the data are delivered as TOA data, they need to be converted to SR. Open source or commercial processing tools can be installed on the cloud or data hub to create ARD products. If desired, the SR data can be further reformatted to the specification of the national grid for Data Cube analysis. National information products will be created at the data hub using algorithms described in the MGD, by national agency staff. Information products are downloaded by the national agency for further analysis, reporting and decision-making.

Global or regional land use land cover or forest cover data provide countries with first estimations of national forest cover that (with the use of local knowledge) can yield reliable and accurate national forest cover estimates. The GFOI MGD provides guidance on how to use these global datasets and calibrate them for national circumstances, and there are examples where GFOI, has been supporting this approach, for example through the

SilvaCarbon workshops. The University of Maryland published tree cover maps, and cumulative tree cover gains and annual losses globally and global datasets adapted to a country level are useful for estimating activity data in the context of REDD+.

Option 2 may represent a long-term solution for national agencies that cannot, or do not, see the need to invest in specific national infrastructure. Space agencies may provide access to a data hub solution, such as the Forestry Thematic Exploitation Platform proposed by ESA. An alternative is to implement operational solutions at partner agencies, such as FAO, regional partners, or NGOs. Future data cube methodologies may best be implemented as cloud applications through regional/national or commercial partners.

### 3 Scenario Data Architecture Description

Each of the scenarios defined in Section 2 differ in their requirements for data transmission, storage, processing, technical capacity, and cost. The variation in technical capacity of GFOI countries must be considered in the assessment of any solution scenario, noting also that these technical capacities may change over time. The solution scenarios should be flexible to accommodate a wide variety of country resources and allow those scenarios to change as countries gain capacity. This section describes these requirements in detail such that an evaluation can be conducted to understand the trade-offs and assess the preferred options for the future.

#### 3.1 Transmission

“Data transmission” is the transfer of data to a given country over the internet. A study was conducted by the CEOS SEO to evaluate internet download rates for 70 GFOI countries that are part of various REDD+ groups. Internet speeds were sourced from Akamai’s State of the Internet website (<https://www.stateoftheinternet.com>). As of early 2015, the report suggests that internet download speeds above 10 Mbps are considered “fast” and the global average speed was 5 Mbps. The fastest region of the world is Southeast Asia, with average speeds just over 20 Mbps. The box-and-whisker figure below shows the range of data transmission rates for the 70 GFOI countries in the study (Figure 4). As Figure 4 shows, 50% of the countries have download speeds between 4.9 and 9 Mbps. For comparison, the average rates of Australia (7.4) and USA (11.5) are shown. The lowest rates for GFOI countries are Benin (1.6) and Sudan (2.3). The highest rates for GFOI countries are Vietnam (17.9) and Thailand (20.2). If one were to consider an average annual data load of 1TB of data (around 500 Landsat 8 scenes), the transmission time at the average global download rate of 5.0 Mbps would be ~19 days. This estimate represents the upper bound. Given careful selection and filtering of data, the data requirements can be fulfilled practically. Download rate will impact the selection of global data flows that are optimal for each country.

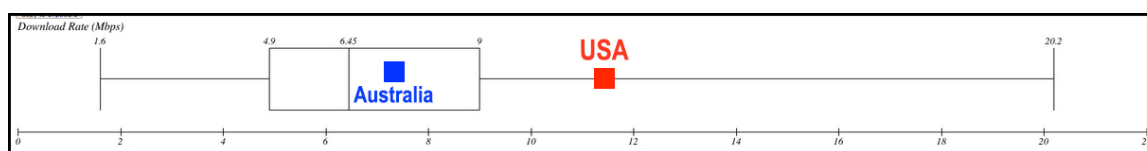


Figure 5 Distribution of internet download speeds for the 70 GFOI countries. The box bounds 50 per cent of the countries and the centre line is the median download speed. The brackets are the upper and lower values.

#### 3.2 Storage

The SEO study noted in Section 3.1 also calculated the annual data volume for the 70 GFOI countries assuming each country received complete coverage from the Landsat-7, Landsat-8 and Sentinel-2A missions. The CEOS COVE coverage analyzer tool (<http://www.ceos-cove.org/>) was used to calculate the number of scenes for each satellite. The TOA file sizes are on average 500 MB (Landsat 7), 1.8 TB (Landsat 8) and 600 MB (Sentinel-2A sub-scene). Many of the GFOI countries are rather small (e.g., Bhutan, Jamaica, Vanuatu), but several are quite large (e.g., DRC, Argentina, Brazil) and will generate a large amount of annual data if all 3 satellites are utilized. The box-and-whisker figure below shows the range of required storage for the 70 GFOI countries in the study (Figure 5). Figure 5 shows that 50% of the

countries have a maximum annual data volume of 0.6 to 3.35 TB. The mean volume is 2.8 TB and the median volume is 1.2 TB. For comparison, the data volume for USA (20.5 TB) and Australia (20.8 TB) is shown. It should be noted that data from other satellites and prior years can be scaled and added to these results to assess total data storage requirements for a given country. The data volume needed to satisfy GFOI and REDD+ requirements may be considerably less depending on national requirements and scenario followed.

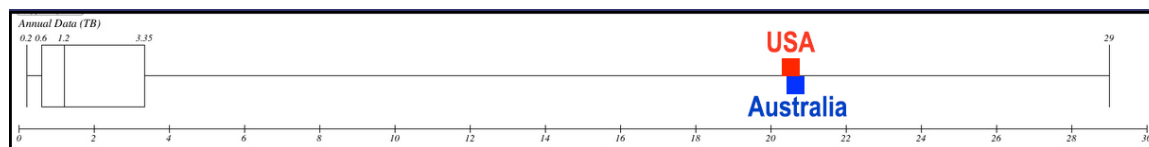


Figure 6 Distribution of the Landsat and Sentinel-2 annual data volumes for the 70 GFOI countries. The box bounds 50 per cent of the countries and the centre line is the median country data volume. (The brackets are the upper and lower values.)

### 3.3 Discovery and Download Tools

Tools are needed to allow agencies to search and discover data to meet forest mapping requirements and to manage databases once assembled. If the cartographic data provision is not centralized and can provide access to the data required, to minimize costs, only data to produce the forest map product should be acquired and maintained. Setting database scope to match requirements results in fewer images to download and update. The goal is to create an appropriate database that may contain a richer set of source data including SAR and high resolution data to meet specific landform, climate and accuracy criteria.

CEOS, through the SEO, the SDCG for GFOI and its Working Group on Information Systems and Services (WGISS) are continually working on new search and discovery tools and enhancements to existing tools. The goal is to provide countries with a CEOS multi-mission portal where they can easily search and discover past (archive) data as well as understand future acquisition plans. Regardless of the data flow scenario chosen for a country (i.e., BAU, Option 1, Option 2), any country will be able to identify required datasets, obtain those datasets in the format desired and perform forest application analyses now and in the future.

Many Space agencies, including USGS, NASA, ESA, INPE, CONAE, are building multi-mission global or region archives. Most CEOS space agencies and commercial data providers provide access to their Earth Observing Data. Commercial partners, such as Google Earth Engine and AWS, maintain global archives of satellite data and tools under an evolving suite of business models. These alternative solutions provide flexibility and redundancy for implementing data flows.

### 3.4 Processing

The scenarios presented in Section 2 depend on the creation of an ARD product. At some point in the data flow, processing will be needed. This processing can be done at the space data provider (Option 1), on a data hub (Option 2), or in a country with the necessary resources (BAU). In many cases, countries prefer to manage their own processing (e.g., Australia) in order to apply their own processing algorithms (e.g. local atmospheric correction, local DEM). It is believed that most of the GFOI countries will desire ARD and prefer to receive these data from space data providers (Option 1) or via a data hub (Option

2). In the latter case (regional or cloud computing hub), the country will need a reasonable internet download speed (>5 Mbps) to adequately connect to the data hub resources and download resulting analysis products. Data hubs, in many cases, will be intermediate solutions until space data provider and country agency functionality matures. However, data hubs can serve an important role within the Capacity Building community and as a for-fee service where the cost of buying services is cheaper than building local infrastructure.

### 3.5 Forest Map Production

Methodologies recommended by GFOI to perform forest application analyses and produce forest map products are coordinated by the Capacity Building teams at the partner agencies (e.g. FAO, SilvaCarbon, Australia Department of the Environment) and are aligned with the GFOI MGD. These analyses can be conducted using traditional scene-based approaches or newer Data Cube approaches, depending on the desires and technical capabilities of a country. Partner agencies have utilized tools designed to support REDD+ (e.g. OpenForis Toolset, SEPAL) or tools desired by individual countries to perform analyses and produce forest maps. As technologies advance and more data become available, enhanced analyses and time series studies using data cubes will be candidates for forest mapping. These new methodologies will support continuous change detection and classification of forest cover, but will be infrastructure intensive.

### 3.6 Costs

Relative, approximate costs can be estimated for each of the solution scenarios: BAU, country-based data management (Option 1) and cloud computing or data hubs (Option 2). In the evaluation in Section 5, relative costs are assigned to space, partner and country agencies for alternative solutions within the scenarios.

**BAU:** These costs are dominated by inefficient and uncoordinated delivery of data, often on media, and by country agency processing costs. The estimated cost to deliver 1 TB of data annually from the Space Agencies (USGS and EC/ESA) to all 127 GFOI countries is >\$US500,000. Space agency capacity building partners have in the past covered these costs. For example, SilvaCarbon covered costs of delivering Landsat products on media. Country agency costs include costs of specialized expertise for pre-processing satellite data and the cost of operational data maintenance and processing, in addition to the costs of producing the Forest Map Products. This option provides maximum flexibility given efficient data access, but has the highest country costs.

**Delivery of ARD to Countries (Option 1):** Option one shifts the cost of pre-processing satellite data to Space agencies. Space agencies can amortize these costs across the entire space data user community, capitalizing on efficiencies of scale. The cost of using the remotely sensed data to produce Forest Map and other land cover products remains with the country agencies. However, their costs of pre-processing satellite data are minimized, while the flexibility for creating value added products remains with the country agency. For space agency data that are not available as ARD, partner and country agency costs will remain and the processing may take place at hubs either in the cloud or at partner organizations.

**Delivery of data to a cloud or data hub for country agency access and analysis (Option 2):** Option 2 shifts costs to partner organizations maintaining data storage and processing hubs. These costs need to be covered through donor or country funding. Therefore, Option 2 may

not be a long term solution. However, Option 2 may provide long term solutions for some countries. Costs for data processing and storage hubs, for a total data storage of 10 TB and sufficient processing for creating ARD, creating data cubes, and running algorithms documented in the MGD, are of the order of \$US10,000 per year per country. These costs can be highly variable depending on the specific solution - adding volatility to long term reliance on data processing and storage hubs.

## 4 Evaluation

### 4.1 Criteria

The functional requirement is to position countries to establish operational MRV processes within NFMS for the next 15 years. As both NFMS mature and space agency products mature, the systems will evolve. The solution now will be different from the solution that is possible and needed fifteen years from now. Donor organizations and countries who provide funding for the production of REDD+ forest map products can use the recommendations to minimize production costs, while maintaining quality.

The qualitative functional criteria that summarize the overall scenario and the quantitative component criteria that estimate costs and performance of the components are outlined below. Each of the scenarios were evaluated using these criteria. When necessary, scenario variants are discussed to capture analysis alternatives, examples include the incorporation of commercial or regional data.

Component criteria discussed in Section 3 and technical impacts assessed in Section 4.2 are

- 1) Data transmission (access) – Speed and Reliability;
- 2) Data storage – Reliability, Performance, Sustainability;
- 3) Data services – Discovery and Selection, Management and Update;
- 4) Data processing – Adaptability and Performance;
- 5) Forest map production – Quality Assurance, Data Provenance, Reproducibility and Comparability of results; and,
- 6) Costs.

Functional options discussed in Section 4.2 include:

- 1) Does the option lead to an operational solution with opportunities for growth that can be sustained by the country and space agencies?
- 2) Does the solution readily expand to include other data sources such as SAR, high resolution, commercial, or other data sources that may have restricted access?
- 3) Does the solution support collaborative relationships with other partners and capacity building needs?
- 4) Is the option implementable? (maturity of system, set up costs, maintenance costs)



## 4.2 Process and Cost Impact

The evaluation is organized by responsible partner and data flow scenario. Capacity building and technology partners serve as intermediaries between space agencies and countries. The emphasis of capacity building partners is methodological and resource management, while technology partners is infrastructure and mechanisms. Capacity building partners and countries consume technologies to produce the needed products. Technology partners work with space agencies and capacity building partners to implement methodologies. The functional option references scenario components. The technical impacts reference the component evaluation criteria described in Section 3. The relative cost impacts are guides to where funding needs to be invested.

Responsible Group	Functional Option	GDF Scenario	Technical Impact	Cost		
				Space Agency	Partner	Country
Space Agency (Data Provider)	Process and deliver TOA (level 1) data over internet to country or partner agency	BAU Option 1/2	Core space agency function. May require partner agency assistance to create ARD. Requires reasonable country internet access, storage and processing.	Low	High	High
	Process and deliver scene-based ARD (level 2) over internet to country or partner agency	Option 1/2	Requires significant coordination among space agencies. Requires reasonable country internet access, storage and processing. Not operational product for all space agencies.	Medium	Medium	Medium
	Process and deliver data cube ARD (level 2) over internet to country or partner agency	Option 1	Requires substantial space agency investment. Optimized for time series analysis. Requires reasonable country internet access, storage and new processing methodologies. Only prototype products available.	High	Medium	Medium
Capacity Building Partner	Process, as needed, and store scene-based and data cube ARD on a temporary basis	Option 1/2	Requires hosting Data Cube software for on-demand Data Cube formation	Low	High	Low
	Deliver scene based or data cube ARD to countries over the internet or on media	Option 1	Requires reasonable partner and country storage and transmission infrastructure. Requires funding for media and shipping. Countries need storage for data.	Low	Medium	Medium
	Temporarily host services for the production of forest products by country agencies	Option 2	Requires hosting Data Cube software for on-demand Data Cube formation and open source tools for Forest Product creation	Low	High	Medium
	Produce high level global or regional forest products that can be used as precursor products	Option 2	Requires capacity building partner with the expertise and stable funding needed to provide products	Low	High	Low

## Global Data Flow Study for the Global Forest Observations Initiative (GFOI)

Responsible Group	Functional Option	GDF Scenario	Technical Impact	Cost		
				Space agency	Partner	Country
Technology Partner	Create and deliver Data Cube open source software to GitHub for country and partner use	Option 1/2	Sustained maintenance of GitHub and open source tools by CEOS	High	Medium	Low
	Create, store and maintain data cubes for technology demonstration	BAU Option 1/2	Partner investment in creation, storage and maintenance of Data Cubes for countries	Low	High	Low
	Deliver mini-cubes to countries to support CB partners for technology demonstration	Option 1/2	Partner investment in mini-cube creation and delivery and sustained investment in capacity building/training by partners	Low	High	Low
	Store global or regional collections of TOA or ARD on mirror sites	Option 1	Maintained and funded by external partners (e.g. Amazon, Google, SERVIR) or sibling space agency. Authoritative data source remains the originating space agency	Medium	Low	Low
GFOI Country (User)	Download scene-based or Data Cube ARD from space agency and produce forest products	Option 1	Requires internet capacity to download ARD. Country processes ARD locally to produce forest products.	Low	Medium	Medium
	Download ARD from storage and processing hub supported by CB or technology partner and produce forest products	Option 1	Requires investment in storage and processing hub. Requires internet capacity to download scene ARD. Country processes ARD locally to produce forest products.	Low	Medium	Medium
	Use free/open tools and storage in the cloud or at a partner's hub to produce forest products	Option 2	Requires capacity building and training for countries to utilize hub. Requires investment in hub.	Low	High	Low
	Download TOA data in scene-based format and use local storage and processing	BAU	Requires substantial infrastructure to download, store and process data. Requires substantial expertise to process data, and create forest products.	Low	Low	High
	Download precursor regional or global forest or land use land cover map	Option 2	Requires minimal infrastructure for download and processing. Requires local knowledge to assess and improve precursor product.	Low	Low	Medium

*Table 4 Matrix associating costs and benefits to responsible groups and functions*

Low (green) estimates have least impact on organizations. High (red) estimates have most impact, but often also have the highest return on investment. It is expected that space agency investment in tiled products leading to data cubes, will yield significant future reduction in country costs as well as significant uptake and application of space agency data. High partner cost estimates are often designed to yield cost benefits to countries. These will be a combination of short term benefits - such as providing access to infrastructure, and long term benefits - such as partner produced global and regional products that will reduce country costs.

The minimization of country costs may require high donor and space agency investments. High space agency costs assumed by space agencies can be justified by amortization across a very large data volume and the production of improved data products that also benefit the national interests of the space agency. Investment in high partner costs should result in a better and more efficiently produced forest product that will meet the REDD+ donor requirements for many countries. The goal is to minimize country agency costs, while also minimizing per instance partner and space agency costs.

## 5 Conclusions and Recommendations

### 5.1 Conclusions

GFOI has moved CEOS coordination activities with its member space agencies closer to the end users of EO satellite data than ever before as users seek to maximise data uptake in support of societal needs. SDCG has noted that the same obstacles to country uptake and application of satellite data were being raised consistently and repeatedly, and were likely to continue to do so.

The study incorporates feedback based on the practical experience of a diverse number of countries with interest in forest monitoring, including: Colombia, Kenya, Mozambique, Nepal, and Cameroon. Feedback was also received from key capacity building partners The World Bank, FAO, and SilvaCarbon. And from practitioners in industry that have broad in-country experience in the development of national MRV systems.

SDCG has concluded that, with the advent of continuity of supply from the Landsat, Sentinel-1, Sentinel-2 (and other) data streams, the discussion among space agencies and its users has moved from ensuring data availability, to access and exploitation in data-rich environments. New and powerful applications will be possible using the dense time-series of observations that can be produced from these satellites, while providing a more selection choices for applications that do not require dense time-series. The analysis of dense-time series requires continued advances in interoperability of data from the different sources, and measures to reduce the complexity and cost of data management and movement, plus an evolution in local and cloud infrastructures to support the analysis.

SDCG sought to undertake a broader assessment of the underlying challenges and opportunities for CEOS, its member agencies, and space data providers in ensuring the uptake and application of data by typical user organisations, and draws attention to the following as the context for the study:

- i. GFOI user feedback indicates that the move to online data distribution systems, combined with data volume increases associated with higher spatial and spectral resolutions, is **effectively excluding** a large number of countries as potential users of satellite EO data to support their NFMS. Even without constrained internet infrastructure (e.g. bandwidth and reliability limitations), the cost and complexity of managing datasets are proving to be major obstacles to the realisation of societal benefit.
- ii. Parallel operation of US and European core land surface imaging data streams offers the potential of new forest monitoring applications and users, making best use of **dense time-series observations**. To realise the full potential of these data streams, new data management approaches are required.
- iii. In order to support achieving this full potential, the burden of satellite EO data pre-processing should be handled by space data providers building on their technical capacities, and to ensure data consistency. A common thread emerging is the emphasis on the supply of analysis ready data, which represents a major step towards removal of many of the obstacles identified by SDCG and to realisation of new applications and

- users from dense time-series datasets. **For maximum benefit, CEOS must coordinate analysis ready data development among its agencies.**
- iv. The advanced capabilities and user platforms provided by **'big data' players** (e.g. Google, AWS) have the promise to greatly improve the EO satellite data user experience. This includes the potential to lower the bar for access, and increase the ease with which data can be exploited and applied.
  - v. The CEOS balance of effort has swung to place **more emphasis on supporting uptake and application of data**, and support to GFOI and forest monitoring via SDCG is one example of this. This shift has significant implications for the nature of the partnerships that CEOS must engage in to ensure effective data flows. For example, effective engagement of countries through capacity building partners, such as UN agencies, investment banks and development aid arms of the governments that host CEOS agencies.
  - vi. A Future Data Architectures study is underway within CEOS which will report in parallel with this Global Data Flows study by SDCG. It is **exploring new technologies**, such as the CEOS Data Cube, which have significant potential to address some of the challenges identified by SDCG. In cooperation with the SEO SDCG has initiated pilot projects with GFOI countries to explore the value of these technologies. SDCG wishes to highlight the need for a broader-based effort, with multiple contributors and a clear strategy, including but broader than GFOI. For example, USGS has noted that *'Investment in ARD and the CEOS Data Cube can help CEOS realise a future which is information-focused and sensor-agnostic and where the user can create new and powerful applications that make maximum use, with minimum effort, of the more frequent coverage available from sensors provided by a range of CEOS agencies.'*

Conclusions specific to GFOI and SDCG activities, through country agencies and capacity building partners, are as follows:

- 1) Difficulties of data processing and management from user agencies has likely delayed the uptake of CEOS agency data for forest monitoring. The issue has been exacerbated by increased data volumes from Landsat-8 and Sentinel-2.
- 2) Although GFOI continues to explore collaboration among space agencies, donor agencies, and UN agencies for the effective uptake of EO satellite data for national forest monitoring, there is not yet a clear model for effective future partnerships for CEOS. In particular, strategies for effective engagement of UN agencies (e.g. FAO) and donors are needed. Better communication and coordination between the space and development aid arms of government is needed. The SDCG cooperation with SilvaCarbon to date, and possible upcoming collaboration with SERVIR, illustrate the benefits of greater collaboration.
- 3) The combination of the need for sovereign capability, and protection of the role of responsible teams within country agencies, may mean that not all countries will select the 'simplest and cheapest' scenario for use of EO satellite data, and they may wish to continue to retain the expert knowledge required for all processing steps. Future data architectures, including the Data Cube, must recognise the need for sovereign ownership of systems, data, and information – in particular for NFMS in support of formal reporting to groups like the UNFCCC.
- 4) The use of cloud computing infrastructure for the storage and processing of data and information is impacting every sector of society, including EO. Employing this infrastructure at the volumes and computational power involved for EO data may

present high costs which are not necessarily traditionally provided for in the budgets and staffing structures of the large data providers within CEOS. For some countries, centralised processing on cloud computing infrastructure with low bandwidth requirements for product access may be the only practical solution, and merit further investigation as part of the broader CEOS strategy for FDA. Sustained support and data supply to these national systems will be needed and the ongoing cost is a significant consideration. Sustained funding for national forest monitoring is a logical role for development aid agencies and investments banks and CEOS must emphasis functional partnerships with these bodies. SDCG recognises that cloud computing infrastructure has the potential to overcome some of the difficulties in sustaining technical programmes in country, and urges CEOS and its member agencies to consider demonstrations of this capability.

- 5) This study has looked at alternative scenarios for movement of data, and concluded that movement of terabytes of data annually for global forest monitoring is not affordable, sustainable or necessary for a large number of prospective user countries. SDCG foresees a future where EO satellite data does not move far from the hubs of the large data providers, but instead users can employ advanced interfaces to generate their information products of choice for download.

## 5.2 Recommendations

The Global Data Flows study undertaken by SDCG has confirmed concerns that space agency satellite EO programmes will increasingly fail to realise their full potential unless steps are taken to remove obstacles to data uptake and application by typical users. A range of opinion and feedback suggested that CEOS and space agencies should plan for a diversity of future scenarios to be able to service the needs and meet the capacities of all GFOI countries. These might extend from continued supply of national coverage of low level data products, through to supply of ARD, and beyond to the provision of cloud computing infrastructure storage and processing such that poor bandwidth or capacity countries might succeed in generating nationally-owned forest data products.

Some of the resulting recommendations from SDCG are specific to GFOI, whilst others are broader, applying to CEOS and its space agencies, capacity building actors, and further recommendations relate to user agencies within countries, as follows:

- 1) **Widely-accepted ARD descriptions and specifications** from CEOS are important to ensuring maximum benefit and interoperability for all available CEOS agency core land surface imaging data streams. CEOS, LSI-VC, and Working Groups (e.g. WGCV) should direct coordination among space agencies to achieve consistency and compatibility of ARD products. How these specifications are implemented will vary among the CEOS agencies.
- 2) GFOI users, including countries, and capacity building and support agencies, have indicated that basic functions like **improved data access directly** from space agencies and through external partners, and **improved data discovery tools** for multi-sensor search and ARD products would improve the uptake of satellite EO data for forest monitoring. CEOS should consider incorporating this feedback in the definition of Future Data Architectures.
- 3) CEOS and its member agencies should transparently **promote community uptake of ARD**, in particular to emerging user platforms from 'data giants' like Google and

Amazon. The objective should be to foster acceptance and interoperability so that CEOS agency data is easily ingested and applied in these emerging environments.

- 4) **Interoperability between the GFOI Core Data Streams of Landsat-8 and Sentinel-2** would be of significant value to GFOI countries and users. An interoperable or blended product would simplify the application of these data streams which are critical for national forest monitoring. In addition, these interoperable products have the potential to open up or strengthen existing time series applications. We recommend that CEOS explore prioritising the development of these products.
- 5) SDCG, the SEO, LSI-VC, and several individual CEOS agencies have progressed **pilot activities to exercise some of the fundamental and enabling elements of future data architectures** which have the potential to address some of the concerns raised by this study. Given this potential, these activities merit broad CEOS agency engagement and support and should be prioritised in any strategy on future data architectures:
  - a. SDCG recommends that CEOS and its member agencies include accommodation for **tiled data and time series stacks, compatible with the CEOS Data Cube** in the development pathway of ARD products, and provide guidelines for their use.
  - b. **Support for pilot activities initiated by SDCG** to demonstrate the potential of these technologies to both data providers and data users. Support should be provided in line with the CEOS Data Cube Work Plan, initiated by the SEO, and contributions to the Work Plan should be made as appropriate.
  - c. **Other technologies and approaches may be highlighted by the FDA report** and SDCG is ready to provide feedback as to which might appear most promising in support of the challenges facing GFOI and CEOS.
- 6) CEOS and its member agencies should **support the development of a model pilot end-to-end NFMS based on, and collaboration with, the GFOI components** (methods and guidance, space data, capacity building, and R&D). This pilot was proposed to GFOI by SDCG to accelerate country interest in the practical benefits of GFOI, and the first country that has been engaged is Colombia. One of the goals, being pursued directly by the SEO, is to create a Colombian Data Cube system that will follow GFOI Methods and Guidance Documentation to ensure IPCC-compliant reporting for Colombian REDD+ activities.
- 7) Many of the issues raised in this study have **relevance to the broader FDA study** in development by CEOS. SDCG recommends that the FDA *ad hoc* Team review this report, and the two groups should discuss and **identify areas of common interest and complementarity**.
- 8) Current and potential capacity building partners and technology providers (e.g. FAO, SilvaCarbon, Google, AWS), must support transition to ARD products. It is acknowledged that many ARD definitions exist and are appropriate from the fundamental atmospheric, radiometric, and geometric corrected products to global baseline classifications that can be used as national pre-maps. The overall goal is to reduce barriers to the use of space data.
- 9) In the course of developing future data architectures, CEOS and its member agencies should **give full consideration and support to pilot systems SDCG is currently developing to demonstrate the possibilities of alternate scenarios for the flow of space data** in support of forest monitoring. This includes SEO engagement with capacity

building partners such as The World Bank, and SERVIR, regarding applications of the CEOS Data Cube.

- 10) SDCG welcomed the discussions at SIT-31 which **acknowledged the progressing of CEOS activities towards user engagement**, and the need for strategic consideration of the partnerships that this move requires. This engagement is taking place through initiatives such as GFOI, GEOGLAM, and GEODARMA. SDCG recommends that the following issues be reflected in the follow-up discussions and actions:
  - a. **CEOS should develop a strategy for partnerships with donor bodies**, for example international investment banks, or the aid development branches of the governments of CEOS agencies. These partnerships are essential to establish the necessary interfaces and to sustain technical solutions offered to user countries. SDCG has seen that cloud computing infrastructure can potentially overcome some of the difficulties in sustaining technical programmes in country, and recommends that CEOS consider championing the combination of satellite EO data and cloud computing infrastructure in support of aid agency objectives.
  - b. SIT-31 noted that **UN agencies are often identified as important partners for CEOS**, in areas where there is a clear mandate and natural role for CEOS and satellite EO. However, the complexity of the UN organisational structure, and distributed of responsibility across several UN agencies has limited effectiveness. CEOS must either find a way of co-existence and collaborate with the key UN agencies, or consciously conclude that space agency goals might be more effectively served through alternative partnerships (e.g. aid agencies, investment banks).
  - c. The possibility of **novel partnerships with the 'data giants'** (many of whom already offer alternate sources of CEOS agency data) to help sustain CEOS pilot projects should be investigated. For GFOI, these potential infrastructure partners could: work with capacity building partners to prototype applications for access and production of forest maps; coordinate with space agencies to develop prototype tools and access mechanisms; work with space agencies to improve access and selection tools; and implement prototype access and analysis methodologies.
- 11) More broadly, **CEOS should continue to develop discussions around the partnerships and geometries that will be required for success in relation to the growing emphasis on user engagement**. There will be variations among initiatives but there may be common features around essential institutional arrangements (donors, UN, 'big data' players in addition to CEOS, GEO, space agencies, and country agencies), and technical solutions that could be highlighted as recipes for success for CEOS initiatives. SDCG sees significant opportunities resulting from the emergence of ARD, especially in combination with novel data architectures such as the CEOS Data Cube. A broad strategy is needed for space agencies to take maximum advantage of these opportunities. SDCG remains willing to serve in prototyping these developments in support of GFOI user countries, and significant related work is already underway.
- 12) **Realising the gains offered by a move towards ARD requires significant communications, promotional, sales, and capacity building efforts by CEOS**. ARD offers the potential for a significant reduction in complexity for many users interested in forest monitoring using satellite data, yet persuading users that they need not repeat the basic correction and processing of the data supplied by CEOS agencies may be required. The systematic availability of consistently-processed and interoperable satellite data can contribute to the streamlining of the domain of technical aid directed at forest



monitoring programmes, which currently addresses satellite data requirements in a largely unstructured manner.

- 13) **CEOS agency governments can all help with the promotion of GFOI MGD by insisting on its adoption in all of their forest-related aid programmes.** GFOI is predicated on having consistent and comparable national monitoring systems which are guaranteed to be IPCC-compliant, and adoption of the GFOI MGD is therefore a crucial measure of success.
- 14) Countries are best placed to determine the critical factors behind selection of the most promising scenario for their own satellite data flows in support of forest monitoring. SDCG remains available to assist in support of national space data assessments and to update countries on the latest developments within CEOS that may assist.

## Appendix A Acronyms

### *Agencies and Organisations*

Agency/Organisation	Name
CEOS	Committee on Earth Observation Satellites
CONAE	Comisión Nacional de Actividades Espaciales (Argentina)
CSIRO	Commonwealth Scientific and Industrial Research Organisation
ESA	European Space Agency
FAO	Food and Agriculture Organisation of the United Nations
GA	Geoscience Australia
GEO	Group on Earth Observations
GFOI	Global Forest Observations Initiative
GOFC-GOLD	Global Observation of Forest and Land Cover Dynamics
INPE	Instituto Nacional de Pesquisas Espaciais (Brazil)
IPCC	Intergovernmental Panel on Climate Change
JAXA	Japan Aerospace Exploration Agency
NASA	National Aeronautics and Space Administration
SDCG	Space Data Coordination Group (for GFOI)
SEO	(CEOS) Systems Engineering Office
UN	United Nations
USGS	US Geological Survey
WGCV	(CEOS) Working Group on Calibration and Validation

<b>WGISS</b>	(CEOS) Working Group on Information Systems and Services
--------------	--

*Acronyms*

<b>Acronym</b>	<b>Expansion</b>
<b>AGDC</b>	Australian Geoscience Data Cube
<b>ARD</b>	Analysis Ready Data
<b>AWS</b>	Amazon Web Services
<b>CB</b>	Capacity Building
<b>DEM</b>	Digital Elevation Model
<b>EO</b>	Earth Observation
<b>FDA</b>	Future Data Architectures
<b>GEODARMA</b>	GEO Data Access for Risk Management
<b>LCMAP</b>	Land Change Monitoring, Assessment, and Projection
<b>LSI-VC</b>	(CEOS) Land Surface Imaging Virtual Constellation
<b>MGD</b>	(GFOI) Methods and Guidance
<b>MRV</b>	Measurement, Reporting and Verification
<b>NFMS</b>	National Forest Monitoring Systems
<b>NGO</b>	Non-Governmental Organisation
<b>REDD</b>	Reduced Emissions from Deforestation and Forest Degradation
<b>SR</b>	Surface Reflectance
<b>TEP</b>	Thematic Exploitation Platforms
<b>TB</b>	Terabyte
<b>TOA</b>	Top of Atmosphere

