



# Global Data Flows Study

Prepared by the CEOS Space Data Coordination Group (SDCG)  
for the Global Forest Observations Initiative (GFOI)

Draft Version In Progress  
April 2016



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

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

---

Draft Version In Progress

6 April 2016

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Overview	3
1.2	Purpose	4
1.3	Context	5
1.4	Contents	5
<b>2</b>	<b>Global Forest Observation Scenarios</b>	<b>7</b>
2.1	Scenario components	8
2.2	Analysis Ready Data	9
2.3	Scenarios	10
<b>3</b>	<b>Scenario Architecture Description</b>	<b>14</b>
3.1	Data Transmission	14
3.2	Data Storage	15
3.3	Data Services	15
3.4	Data Processing	16
3.5	Forest Map Production	16
3.6	Costs	16
<b>4</b>	<b>Evaluation Criteria</b>	<b>18</b>
<b>5</b>	<b>Evaluation</b>	<b>19</b>
<b>6</b>	<b>Recommendations</b>	<b>21</b>
6.1	Recommendations for Space Agencies	21
6.2	Capacity Building Recommendations	21
6.3	Recommendations for Country and Donor Agencies	22

# 1 Introduction

## 1.1 Overview

This document evaluates alternate solutions for reducing barriers to effective use of satellite data in implementing measurement, reporting and verification (MRV) within national forest monitoring systems (NFMS) in support of REDD+. The launch of Landsat 8, Sentinel-1A, and Sentinel-2A, and resulting access to large volumes of data at no cost to the user changes the environment within which REDD+ countries can access and analyze data. However capacity is often lacking to handle and process large amounts of data, and the acceptance of higher level global products created by external organisations is often low within government organisations. These rapid changes in data availability provide both opportunities and challenges for significant advances in MRV methodologies including:

- 1) Opportunity: The increased number of satellites provide a dense time series to improve change detection and classification of global forest cover
- 2) Opportunity: Access to Optical and SAR core data flows provide alternate solutions for complex problems, landforms and persistently cloudy regions
- 3) Challenge: Higher data volume increases storage and delivery costs
- 4) Challenge: Additional choices adds complexity to data discovery and selection
- 5) Challenge: Multiple sensors make georegistration and cross calibration more complicated
- 6) Challenge: New methodologies are needed to benefit from increased data volumes and increased sensor complexity
- 7) Challenge: Acceptance and take-up of new methodologies by country's authorities and agencies

To realize these opportunities and address the challenges, the Space Data Coordination Group (SDCG) needs to consider a coordinated phased approach with 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) Provide tools for use in maintaining local databases through incremental updates as new data, that meet specified agency criteria, become available
- 2) Provide improved discovery and selection tools needed to assist countries in more effectively identifying required data before attempting downloads.
- 3) Work with space agencies and partners to reduce the burden of data preprocessing on forest management organizations
- 4) Work with space agencies and partners to implement new preprocessing, change detection and classification methodologies
- 5) Acknowledge that simple and more appropriate solutions may exist that do not require an investment in expensive infrastructure

A fundamental premise of this study is the shared objective to reduce barriers and increase efficiencies to the production of national forest product maps used in generating activity data for reporting greenhouse gas emissions and removals from REDD+ activities. This may involve new data intensive methodologies or more efficient selection, discovery and access tools. An underlying premise is the vested interest in the REDD+ donor community to ensure that Forest Product Maps meet quality and content guidelines as efficiently and cost effectively as possible.

## 1.2 Purpose

It is acknowledged that with vastly increased volumes of data from missions providing continuous global coverage to users at no cost (e.g. the Landsat and Sentinel series), and new methodologies requiring long and dense time series, the movement of such data around the world is becoming increasingly unsustainable. The current business model of uncoordinated delivery of minimally processed data on media is not an effective mechanism for maintaining a national archive - neither at facilities within a country, nor in the “cloud.” This is due in part, but not entirely, to limitations of national internet and high performance computing infrastructure. Even in countries with access to high-speed networks, the construction and maintenance of these large datasets is a time consuming and costly exercise.

More efficient mechanisms are needed for the discovery and maintenance of large national databases. Computing infrastructure must be scoped to meet national requirements and budgets. Among infrastructure that needs to be considered are cloud-based solutions, regional solutions, national solutions and solutions using global archives at CEOS space agencies and commercial partners.

The objective of this study is to explore practical scenarios building on multi-year collections of Landsat and Sentinel-2 data. Other data, such as Sentinel-1, CBERS-4, ASTER, SPOT, ALOS-1 will be discussed as they pertain to the study. Although the focus of the study is on freely available core data streams, unique characteristics of commercial and other restricted distribution data will be addressed. The study is scoped to address the pros and cons 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, including FAO, SilvaCarbon, and Australia’s Department of the Environment (DOE), are central to the success of any global data flows study. Capacity building partners bridge the gap between the REDD+ donor and national forest monitoring communities.

It is understood that many of the data flows issues are institutional and cannot be solved with better products or access. Forest map products are often created within short-term projects, which don’t have funding or a mandate to manage the long term archives necessary for continuous forest management. The lack of coordination among donors exacerbates the problem through inconsistent requirements, tightly defined requirements and short term funding resulting in products that are not comparable and resources that cannot be easily shared among projects. Solutions to these institutional 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 report is cast within the framework of the 3-Year Work Plan of the CEOS Space Data Coordination Group (SDCG) for the Global Forest Observations Initiative. Outcome 2 of the 3 Year 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.”

The study builds on activities at the space agencies and FAO. The Australian Geoscience Data Cube (AGDC) methodology under development by Geoscience Australia (GA), the Commonwealth Scientific and Industrial Research Organization (CSIRO) and the Australia National Computational Infrastructure (NCI), is central to development efforts implemented by the CEOS Systems Engineering Office (SEO) and at USGS. The SEO is working with the AGDC to implement the Data Cube locally, regionally and in the “cloud”, and USGS is implementing the Data Cube methodology in its Land Change, Monitoring and Prediction (LCMAP) system. ESA’s Thematic Exploitation Platform and Big Data projects are investigating similar methodologies. System for Earth Observation Data Access, Processing and Analysis for Land Monitoring (SEPAL), FAO’s scene-based operational data management tool for forest monitoring can leverage lessons learned from the Global Data Flow Study to advance its implementation. The CEOS Land Surface Imaging Virtual Constellation (LSI-VC) will use the outcome of the study as the foundation for more general land surface applications.

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 participate in reporting forest-related greenhouse gas emissions and national forest carbon stocks to UNFCCC under the REDD+ provisions. This study aims to shift the emphasis from data acquisition to data and information access.

### 1.4 Contents

This study compares the main parameters and costs associated with a ‘business as usual’ data architecture with alternate architectures and draws conclusions as to the pros and cons 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 operate under alternate future scenarios is provided. The need for efficient data selection and database maintenance mechanism will also be addressed. The study takes into account practical considerations identified from the pilot work underway with Kenya and Colombia, and identifies areas for further consideration by CEOS through the *ad hoc* team on future data architectures 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 data segment plans of core data stream providers like USGS and ESA/EC.

The study is framed as follows:

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

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

**Chapter 4** specifies the evaluation criteria for assessing benefits comparing business as usual scenarios to cloud-based analysis ready data models.

**Chapter 5** presents the evaluation of the alternate scenarios.

**Chapter 6** 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 (see Figure 1) are evaluated. The Business As Usual (BAU) scenario describes the unstructured global data flow of scene-based Top-of-Atmosphere (TOA) products. Alternate scenarios (Option 1 and Option 2) are proposed to provide cost savings, higher efficiencies and improved forest resource analysis and monitoring capabilities in the future. Option 1 describes the flow of Analysis Ready Data (ARD) products to country agencies for the local production of Forest Map products. Option 2 describes the flow of data to a computing cloud or data hub, where country agencies remotely produce national forest products, which are then downloaded to country agencies for further analysis. The country agency controls the selection of the source data, process and the results. Restricted distribution and access data will be delivered to the in country agency or to a capacity building partner under an agreement including the data provider and the capacity building partner.

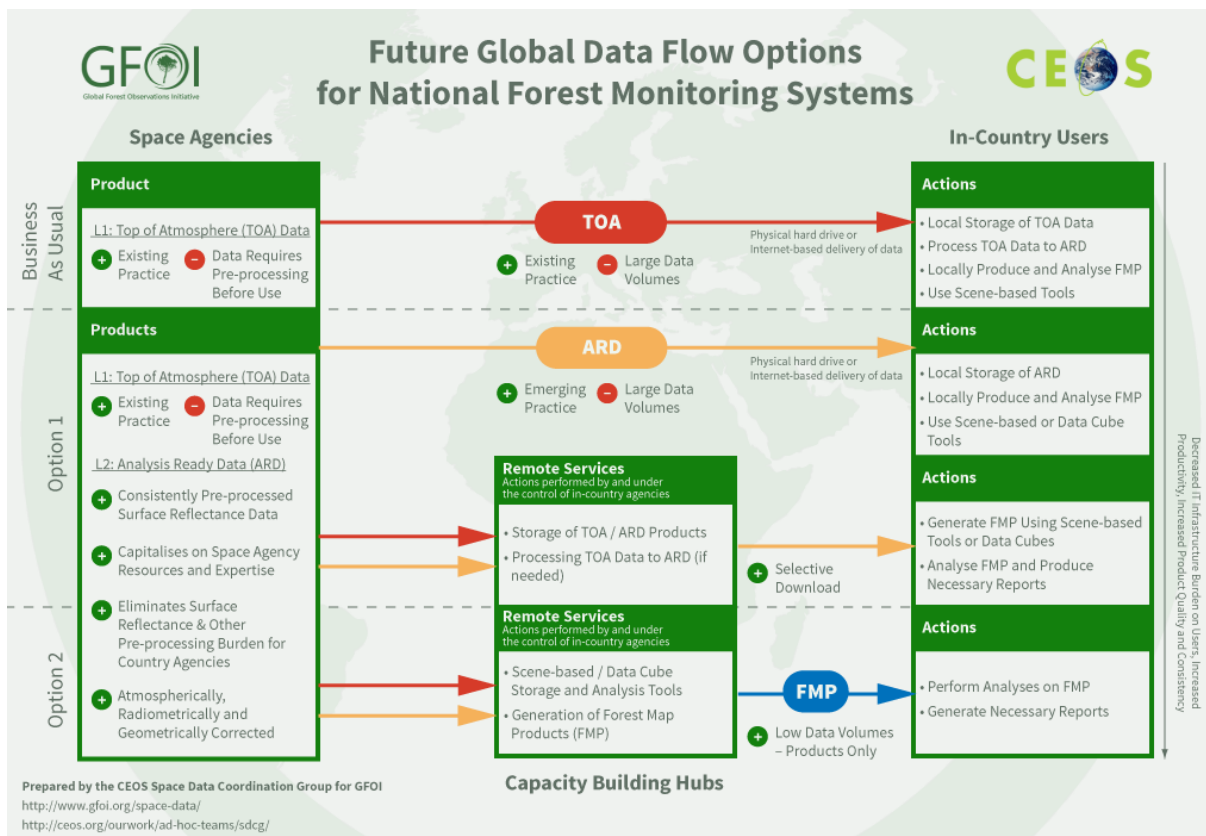


Figure 1. Three general data flow scenarios are evaluated. The difference between these scenarios depends on where the data processing is performed, where the data is stored and where the forest map product is created. Points of note: 1) Activities that take place in the CB hubs are performed by and 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 which is a great reduction in data volume; 4) the ARD plus is the shift of SR and other pre-processing burden from the countries which will translate to significant cost savings; and 5) TOA negative is the cost of data pre-processing (in staff and time)



TOA radiance and reflectance products are often called “level-1 data” and ARD products are often called “level-2 data”. TOA data are radiometrically calibrated and geometrically corrected. ARD is produced from TOA using atmospheric correction to yield a surface reflectance product and has explicit requirements for geolocational accuracy. Existing forest maps serve an important function as baseline forest products and as “bootstrap” products for new forest maps.

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 globally relevant. The criteria for success need to meet “typical” country requirements. For GFOI, these criteria are designed to meet donor specified REDD+ requirements, rather than the requirements of a specific country. However, individual countries may have other requirements that would benefit by more general solutions.

Expectations must be managed to achieve a sustainable solution that can grow as 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 growth. The scenarios presented provide some transitional alternatives for managing costs and risks.

National agency users control the data flow to meet their forest monitoring requirements. Capacity building partners, such as FAO, or SilvaCarbon, and technology partners such as the CEOS SEO, Google or Amazon Web Services, can facilitate the delivery and processing of data to meet the requirements established by the national agency and donors. Partner agencies may be other national agencies. Application to countries will require scaling of data volumes, user skills, local infrastructure and budget. Important component studies lie along the analysis path. At each step alternatives may exist that are relevant for specific national requirements.

## 2.1 Scenario components

In 2016, data acquisitions of the core optical data sets will provide repeat global coverage. Data limitations will increasingly shift to country specific challenges. Data access and the establishment of infrastructures need to evolve in order to accommodate emerging methodologies that utilize the large volumes of newly available satellite data including SAR and high-resolution. Managing the flow of the vast amount of new satellite data will become increasingly complicated - made even more complex by the need to make multiple data flows compatible across satellite sources. The infrastructure needed to move, store and analyse these data becomes increasingly difficult in the BAU model. In fact, these models will not be sustainable in the future and are only marginally effective now. Emerging monitoring methodologies increasingly depend on access to dense multiyear temporal cross-calibrated and geo-registered satellite databases.

For countries with limited access, storage or computational infrastructure, the delivery of ARD may represent significant savings. In addition, the use of intermediate storage options (e.g. cloud computing, data hubs) will reduce the dependency on internet bandwidth and also reduce long-term costs, as shown in Section 3.6.

## 2.2 Analysis Ready Data

The definition of “Analysis Ready Data” depends on national infrastructure and requirements. The intent of ARD is to shift the burden of preparing space data for analysis from data users to data providers. Capacity building and technology partners can assist in the shift as space agencies transition to the delivery of ARD and as national infrastructure evolves.

The transition to ARD has two components. 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 data cubes to optimize time series analysis. Scene-based structures represent the sampling mechanism of the satellite. Data cubes are tiled data structures where each geographic tile represents the same location on the ground. Data cubes provide indices and access tools to more readily support the extraction of pixels through time. Whether scene-based or data cubes are preferred depends on national agency requirements.

Table 1. Analysis Ready Data

<b>Optical</b>	Radiometry	Absolute calibration Cross calibration Band difference adjustment Viewing angle correction Solar illumination angle correction Atmospheric correction
	Geometry	Systematic correction Orthorectification Projection Image to image registration
	Pixel level metadata	No data Clouds Shadows Water
<b>Radar</b>	Radiometry	Absolute calibration Radiometric correction for topography Normalization of cross track (near-far range) incidence angle variation Rain attenuation
	Geometry	Systematic correction Orthorectification Projection Image to image registration
	Pixel level metadata	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	

Surface reflectance data registered to within a third of a pixel are the foundational ARD product for optical sensors. Different paths exist to reach this requirement. Most space

agencies produce TOA, also called “at sensor”, radiance or reflectance data products, while some produce higher level products:

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

The current trend is toward the distribution of tiled ARD Data Cubes by the space agencies. No agency currently distributes their ARD as Data Cubes. The data cubes for the Kenyan and Colombian pilot studies are created using open source ingest algorithms developed by the NASA SEO, GA and CSIRO teams. Data Cube architecture and applications are currently in research and development and may not be appropriate for GFOI operational applications.

GFOI’s MGD and R&D teams in partnership with capacity building partners and countries can define higher-level ARD products specifically designed to meet the goals of GFOI. Example derived products 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 ARD input for the production of new forest products. All scenarios are based on the assumption that national agencies have the access and processing infrastructure required to produce the final national forest products.

## 2.3 Scenarios

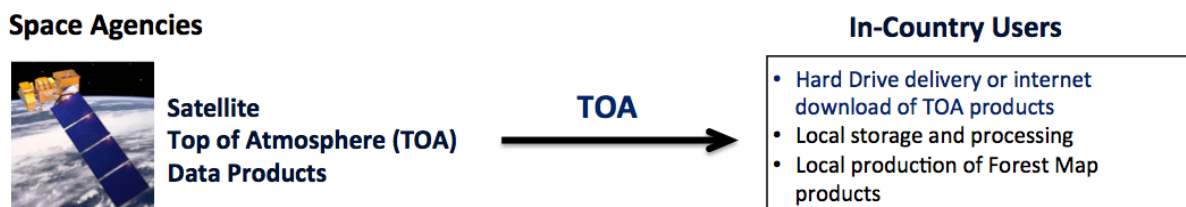
Three scenarios are evaluated. The first is the Business As Usual (BAU) scenario representing how data have typically been processed and delivered to countries. Within the scope of national forest monitoring systems (NFMS) the two alternate scenarios have a common thread: they both rely on the automation of the data processing and reformatting of the satellite data to create ARD. Countries can use either scene-based or tiled data (e.g. Data Cubes) to reduce computational load in analysis tools. The alternate scenarios (Option 1 and Option 2) consider differences in the location of data processing, data storage and data analysis.

Minimizing cost at national agencies requires the transition to ARD products, controlling the amount of data delivered, and improving access tools for database update and maintenance. Production of ARD products by national resource agencies is expensive. No more data should be acquired than can be or needs to be used 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.

The key to success is building on shared methodologies, such as those defined within the GFOI Methods and Guidance Document (MGD) and a database maintenance organization 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 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 core data flow with Landsat 8 data just becoming available. At that time the broad usage of SAR data was very limited for forest applications.



Prior to Landsat 8, most core data products could be delivered on hard 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 the community is in transition to SR products.

The defining limitation of BAU scenarios 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.

An alternative BAU scenario is the delivery of a forest map by a research institute, NGO or commercial partner – with or without country participation. Although the maps are frequently of high quality, the provenance and source of the data needed for comparison to new forest maps is often lacking.

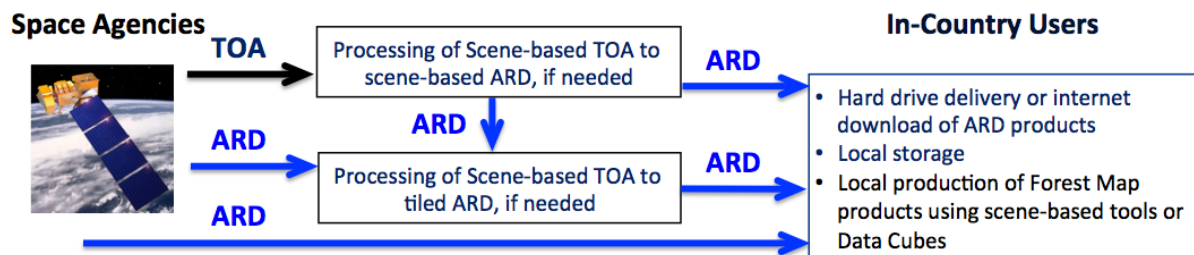
The BAU scenario requires countries to manage large and complex archives. In the past large volumes of data have been delivered to countries with little evaluation of the requirements or the capacity to maintain the archive. Archives of space data are rarely static. The archive needs to be reprocessed 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 may 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 disk drive (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 Amazon Web Services may provide access to infrastructure as needed. 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.

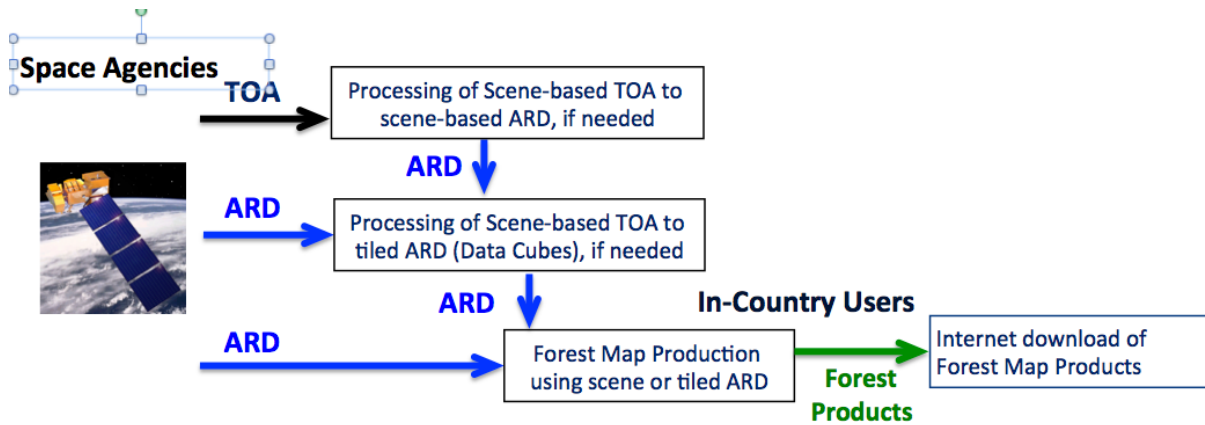


Option 1 does not reduce the data access requirements. However, the switch to ARD products will reduce costs at agencies. The requirement to pre-process the data will be removed and the skills needed to perform these tasks may not be needed at the country or partner agency. Tiled ARD products will provide additional efficiencies since all data for a geographic location can be immediately identified supporting comparison through time and for 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 (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; National agency access to the hub is a requirement. National agency staff implement the forest application analyses and download the forest map results once complete.



The space data (TOA or SR) are pulled from space agencies to the data hub. If the data are delivered as TOA data they need to be converted to Surface Reflectance (SR). The preference is that the TOA-SR step is performed by space agencies, but if that is not possible, then open source or commercial processing tools can be installed on the cloud or data hub to create ARD products. At this point, further application analyses are possible with scene-based tools. 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 MGD algorithms, by national agency staff. Information products are downloaded by the national agency for further analysis, reporting and decision-making.

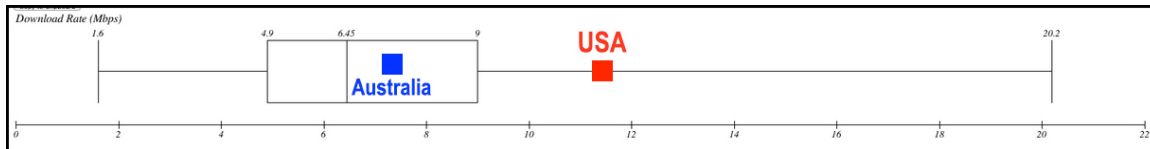
Option 2 may represent a long-term solution for national agencies that cannot, or do not, see the need to invest in national infrastructure. An alternative is to implement operational solutions at partner agencies, such as FAO, regional partners, or NGOs. Future data cube methodologies may best be implemented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

### 3 Scenario Architecture Description

Each of the scenarios defined in Section 2 differ in their requirements for data transmission, data storage, data processing, technical capacity, and cost. The variation in technical capacity (e.g., storage, computing, internet) of GFOI countries must be considered in the assessment of any solution scenario. In addition, 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.

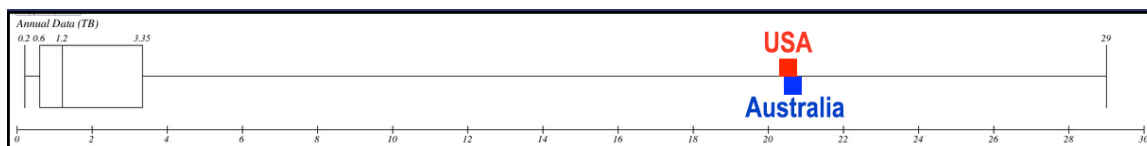
#### 3.1 Data Transmission

It is assumed that “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. As the figure 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.



### 3.2 Data 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 mission. The TOA file sizes are on average 500 MB (Landsat 7), 1.8 TB (Landsat 8) and 600 MB (Sentinel-2A subscene). 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 missions are utilized. The box-and-whisker figure below shows the range of required storage for the 70 GFOI countries in the study. The figure shows that 50% of the countries have a required 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 missions and prior years can be scaled and added to these results to assess total data storage requirements for a given country.



### 3.3 Data Services

Tools are needed to allow agencies to search and discover data to meet forest mapping requirements and to manage databases once assembled. 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 Systems Engineering Office (SEO), the Space Data Coordination Group (SDCG) for GFOI and the 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.



### 3.4 Data Processing

The scenarios presented in Section 2 depend on the creation of an analysis ready data (ARD) product. At some point in the data flow, processing will be needed. This processing can be done at the Space Agency (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 Agencies (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.

### 3.5 Forest Map Production

Methodologies needed 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 intentions of the GFOI Methods and Guidance Document (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. Until now, partner agencies have only utilized accepted tools (e.g. OpenForis Toolset, SEPAL) or tools desired by individual countries to perform analyses and produce forest maps. As technologies advance and data volumes increase, it is believed that more countries will desire to utilize a Data Cube infrastructure for enhanced analyses and time series studies. The CEOS SEO is leading a prototype project in Kenya to test the implementation of a data cube system in parallel to their existing scene-based analyses. Lessons learned from this project will be used to expand the capabilities of the data cube system and ultimately evaluate its potential to be the baseline approach for future forest analyses.

### 3.6 Costs

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 order to make a reasonable comparison, the costs will assume an average annual data volume of 1 TB, and a past data volume of 5 TB (5 times the annual volume). This is reasonable assumption, based on Landsat 7 data, which was the only baseline dataset available until mid-2015.

It is important to make the distinction of how these costs are spread among various entities. For example, the BAU costs are entirely covered by capacity building partners and therefore operationally unsustainable. The increased costs to implement Data Cubes should not be directly compared to BAU costs, since these costs are covered by a variety of groups, such as space agencies, individual countries (local deployment or cloud-based deployment), other partners (e.g. Google partnership with FAO, CEOS SEO partnership with Amazon) or stakeholders (e.g., regional data hubs, GFOI, FAO). In the end, the future cost of delivering large amounts of data cannot be covered by space agencies but must be shared by the users, partners and stakeholders.

**Business as Usual (BAU):** The estimated cost to deliver 1 TB of data annually from the Space Agencies (USGS and EC/ESA) to all 127 GFOI countries is >\$500,000 U.S. These costs are dominated by the requirements of smaller developing nations (~50%), as one would assume they do not have adequate internet and can only support hard drive delivery of data. Such delivery requires annual space agency costs for management and preparation (e.g., one dedicated person, per space agency, per year), hard drives, shipping and possibly travel for capacity building and training. It is assumed that no cost is needed to distribute data from space agency servers to larger and developed countries (~50%) over the internet, as this is already part of the basic program plan for these agencies. While such costs are not sustainable, the requirement to supply functional data to all of the GFOI countries remains.

**Country-based Data Management (Option 1):** Countries will need a reasonable local computer to run Data Cube algorithms (if desired), store large volumes of data and perform forest application analyses. The baseline system (e.g., Titan Workstation) requires a Linux operating system with a 6-core CPU, 3.7 GHz processor, 64 GB RAM, and 12 TB RAID storage. The cost estimate for this system is \$5,000 U.S. In addition, it is assumed the country can provide the necessary system administration and accommodate expansion for the future.

**Cloud Computing or Data Hubs (Option 2):** Costs for regional data hubs and/or cloud-computing hubs can be estimated from recent Amazon Web Services (AWS) costs incurred by the CEOS SEO to support the Kenyan Data Cube project. A reasonable assumption is a total data storage of 10 TB and sufficient processing (EC2, shared 8 cores, 64 GB RAM) for creating ARD, creating data cubes, and running MGD algorithms. This cost estimate is \$10,000 U.S. per year per country and applies primarily to the cloud computing option. In the case of a regional data hub (e.g. SERVIR) it may be possible for countries to obtain free data services, as many of these hubs are funded by external philanthropic organizations. Regardless of the data hub choice, there are still some costs associated with space agency management and preparation of data to populate the hubs, but those costs are significantly less than the BAU scenario.

## 4 Evaluation Criteria

The functional requirement is to position countries to establish operational measurement, reporting and verification (MRV) processes within national forest monitoring systems (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 agencies must be able to use the recommendations to constrain costs and to ensure 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 assessed in Section 5 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
6. Costs

Functional criteria discussed in Section 6 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)

## 5 Evaluation

The evaluation is organized by responsible party and scenario. Capacity building and technology partners serve as intermediaries between space agencies and countries. Capacity building partners will also advance technology and technology partners will provide capacity building. The function option references scenario components. The technical and cost references the component evaluation criteria described in Section 3.

Responsible Group	Functional Option	GDF Scenario	Technical Impact	Cost Impact		
				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	Typical 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	Not operational product for all space agencies. Requires reasonable country internet access, storage and processing.	Medium	Medium	Medium
	Process and deliver data cube ARD (level 2) over internet to country or partner agency	Option 1	Currently not operational product at any space agency. Requires substantial space agency commitment. Optimized for time series analysis. Requires reasonable country internet access, storage and new processing methodologies.	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
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



Global Data Flows Study for the Global Forest Observations Initiative (GFOI)

Responsible Group	Functional Option	GDF Scenario	Technical Impact	Cost Impact		
				Space agency	Partner	Country
	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	Country is responsible for download of TOA, processing to SR, local storage, and creation of forest products.	Low	Low	High

The total cost is reflected as a function of all costs. The country costs reflect the end goal. The minimization of country costs may require high donor and space agency commitments. These 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 product that also benefits the national interests of the space agency. Investment in high partners 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.

## 6 Recommendations

The recommendations build on how well the alternative data flows satisfy the fundamental premises of:

- 1) reducing barriers in the production of national forest products by countries; and
- 2) meeting the strict quality and content guidelines set by the donors as efficiently and cost effectively as possible.

Opportunities and challenges exist for space agencies, capacity building partners, technology partners, country agencies, and donors.

### 6.1 Recommendations for Space Agencies

CEOS will provide coordination between space agencies to improve the space data products and access to the products.

- Establish definitions of ARD products and provide guidelines for their use.
- Evolve ARD products to meet current requirements for scene-based and future requirements for Data Cubes.
- Support the development of methodologies to use Data Cubes.
- Increase coordination among space agencies for cross calibration of products.
- Increase coordination among space agencies to achieve consistency and compatibility of optical and radar ARD products.
- Improve access to data directly from space agencies and through external partners.
- Increase access to older data sets for time series analysis and baseline classifications.
- Improve discovery tools appropriate for multi-sensor searches and for discovery and access to ARD products.

### 6.2 Capacity Building Recommendations

The options provide opportunities for capacity building and to support the implementation of REDD+ MRV. Technology partners can assist in the implementation of access and discovery solution. Capacity Building partners can implement these technology solutions to meet the needs of country agencies

- Technology partners
  - Coordinate 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
  - Improve multi sensor access and selection tools
  - Implement prototype data cube access and analysis methodologies
- Capacity building partners

- Coordinate with technology partners to adapt and implement prototype tools in operational environments
- Coordinate with other capacity building partners and countries to optimize practical and integrated solutions
- Investigate and evaluate multiple data flows to meet country and donor requirements
- Provide country access to storage and processing needed to generate ARD products
- Facilitate future transition to data cube solutions

### 6.3 Recommendations for Country and Donor Agencies

Countries have the responsibility to implement MRV solutions to meet national forest management and donor requirements. The recommendations to space agencies and to partner organizations are intended to facilitate national solutions. The intent is to provide a suite of solutions that countries can tailor to meet national requirements. These requirements and furthermore the solutions will be highly variable. Capacity building partners are often critical for providing a bridge between technical solutions, donor requirements and national forest management needs.

- Determine national infrastructure requirements
  - Identify minimum specific data requirements needed to satisfy donor requirements. Multiuse applications of the data may provide opportunity to share costs.
  - If an in-country storage and processing infrastructure is not required, the forest product can be procured from an external partner. Minimal storage and internet infrastructure is required
  - If there is no requirement for in-country storage and processing infrastructure, but national forest resource assessment expertise is available, an option 2 solution where the processing is performed by national staff via the internet may be satisfactory. Reasonable storage, processing and internet infrastructure is required.
  - If there is a requirement for in-country storage and processing infrastructure, an option 1 solution, where ARD data are delivered to countries may be preferred. Substantial storage, processing and internet access is required.
  - Country agencies with excellent storage, processing and internet access can assess future data cube methodologies for forest monitoring – option 1 with data cubes.
- Evaluate opportunities for growth
  - Capacity building partners can assist countries assess national and donor requirements and technology partner solutions.
  - Capacity building and technology partners can provide guidance to evolve processing systems to meet current and future requirements.
  - Country authority and agency acceptance depends on the demonstration of clear national benefits now and into the future.
  - Option 1 solutions provide the most opportunities for flexibility and multiuse solutions, but require the highest in country costs in infrastructure and staff.

- Option 2 solutions may provide a growth path to more complex data cube functionality with minimum initial national infrastructure costs. Through the use of open source solutions functionality can be moved to national agencies when stable funding and infrastructure is available.
- Evaluate sustainability of solutions
  - Sustainable solutions are key to acceptance by donors and country authorities.
  - Option 1 solutions provide national partners and multiuse scenarios with opportunities to create critical mass of expertise and to share costs.
  - Option 1 solutions are under the control of the national agency and provide opportunities to reduce cost uncertainties.
  - Option 2 solutions provide opportunities for capacity building and can control costs through incremental implementation of capability.
  - Option 1 solutions can be converted to Option 1 solutions once alternatives are evaluated and infrastructure is secured.
  - Option 2 solutions may be the most sustainable and may provide earlier access to new technologies at the cost of less flexibility for national adaptation of solutions.
  - Option 2 solutions may provide continuity of products and expertise. However commercial cloud and donor-funded cloud solutions will have financial uncertainties.