THE FEASIBILITY AND UTILITY OF IMPLEMENTING TEMPORAL DATA CUBES TO SUPPORT PROJECTION OR "FORECAST" MODELS AND LAND CHANGE TRENDS

A Report of the National Geospatial Advisory Committee
Landsat Advisory Group
April 2018

THE FEASIBILITY AND UTILITY OF IMPLEMENTING TEMPORAL DATA CUBES TO SUPPORT PROJECTION OR "FORECAST" MODELS AND LAND CHANGE TRENDS

Executive Summary

In August of 2016, the U.S. Geological Survey (USGS) requested that the Landsat Advisory Group (LAG), a subcommittee of the National Geospatial Advisory Committee, study the feasibility and utility of implementing temporal data cubes to support projection or 'forecast' models of land change trends. This study was a follow-on to two previous LAG study papers on "Product Improvement" and "Cloud computing" that had both been published in 2013. The study was proposed to help address whether a deeper market demand for forecasting land change would develop. Several questions were also posed based on the presumptive use of a data cube with Landsat derived information, as a measure, and time, as a dimension, which this report discusses.

Background

The joint National Aeronautics and Space Administration (NASA)/ United States Geological Survey (USGS) Landsat program provides the longest continuous and openly available space-based record of Earth's land in existence. Landsat missions have acquired moderate resolution multispectral data for over 40 years. The European Space Agency (ESA) has been gathering Earth observation data for a long time and initiated systematic archiving and analysis of data from other agencies' satellites in the early 1980s. It began its own Earth observations with Europe's first Earth Resources Satellite (ERS). The ESA Earth Observing Sentinel satellites over nearly the past four years have added to the amount, the complexity, and the relevance of readily accessible remotely sensed data. Having a facile, agile, and reliable way for all to interact, directly or indirectly, with these, already vast but also growing, collections has both national and international interest. These collections pose the "Big Data" technology challenge to previous data architectures and tools to manipulate or to interrogate priceless observations from a wide-range of sensors. Higher spatial, spectral, and temporal resolution of the collection compounds the challenge as well as the opportunities to better understand our Earth. Improved approaches to the management, preparation, distribution and analysis will relieve some of the data-to-information-to-knowledge progression stress. Algorithms for statistical analysis of increasingly larger samples (and perhaps significantly varying) "Big Data," used under different conditions to address different issues and perspectives, must be wisely selected and used to avoid erroneous statistical inference or inadequate conclusion.

The Federal Geographic Data Committee (FGDC) requested, for the 2016 program, that the Landsat Advisory Group provide advice on "the feasibility and utility of implementing temporal data cubes to support projection or 'forecast' models of land change trends" and noted that this work was "intended as a follow on topic to the LAG study papers on *Product Improvement* and *Cloud computing* published in 2013." Five questions were posed:

- In addition to Landsat, what other data sources (to include EO, SAR, and LIDAR) are optimally suited for leveraging (e.g., co-registered) to support data cube implementations for land change analysis and forecast modeling?

- What kinds of Landsat time-series products would have the broadest community use or most impactful contribution in specific areas?
- Which organizations with expertise in forecast modeling are best postured to evaluate and demonstrate the forecast potential from a Landsat-based temporal data cube?
- How far back in time into the Landsat archive should the staging of 'analysis ready data' be considered? E.g., early data collections such as multi-spectral scanner (MSS) data are less equipped (in terms of metadata) to support rigorous geometric and radiometric calibration compared to later collections.
- How could efficient synergy be realized among government and commercial roles for data cube development, and operations (processing, storage, distribution) to satisfy broad community needs?

The NGAC Paper, dated 11 December 2013, on "Product Improvement: Advise USGS on potential means of modifying the current products to make them more useful to commercial information providers and value-added analysts"[1] made the general recommendation that "USGS further improve Landsat products to both enhance the scientific value of the imagery, but also to provide additional value to the commercial and government organizations wishing to extract the maximum value from the imagery." Seven points expanded that summary recommendation for USGS:

- Clearly define what USGS will produce and avoid competition with commercial work.
- Refine geometric accuracy and radiometric measurements to enable better change detection.
- Improve L1G product geometric accuracy and co-registration.
- Define a standard surface reflectance product.
- Consolidate scientific research and publish best practices for a range of products.
- Provide certification/validation facilities for products not produced by USGS.
- Simplify access to the L1T product.

The second NGAC Paper of the same date, entitled "Cloud Computing: Potential New Approaches to Data Management and Distribution"[1] endorsed the use of cloud computing and suggested how USGS/ Earth Resources Observation and Science (EROS) should leverage that technology by:

- Supporting third-party cloud providers by providing bulk data download;
- Co-locating data and on-demand processing for only the desired information;
- Transmitting the required processing model to the cloud so massive data could be handled by multiple CPUs;
- Downloading subsets of L1T products;
- Giving attention to use of open software standards to avoid tying any services to proprietary software; and
- Streamlining security.

Introduction

Explaining interest in the spatio-temporal data cube

The options for storing and accessing relevant data offer a range of functionality but are somewhat limited with massive data and specific requirements to satisfy particular business cases. In many cases, a data warehouse can adequately support information processing as a stable platform for consolidated

[1] Two of the 2013 NGAC Key Documents found at https://www.fgdc.gov/ngac/key-documents

and transactional data. However, of increasing interest, online analytical processing (OLAP) more adequately allows for multi-faceted consumption of data to meet varied needs. The data cube provides not only a storage structure but also the "staging" space for analysis of the information. The OLAP cube is a multi-dimensional database, which has drawn increasing attention over the past several years for earth observation collection. A marketing promotion for an Earthserver Project workshop on data cubes described the daylong workshop focus in the following way. "The data cube concept promises to tackle some of the challenges that come along with large volumes of environmental and geospatial data. Data cubes offer a more on-demand and analysis-ready access to n-dimensional data, which can be accessed along any axis, allowing for efficient trim or slice operations. The data cube concept makes large volumes of environmental and geospatial data more manageable and thus, increases the general uptake of Big Earth data."[2]

Examining a notional architecture

The Committee on Earth Observation Satellites (CEOS) of which the US is a member country began an Open Data cube initiative in 2016. Brian Killough (NASA) and Robert Woodcock (CSIRO of Australia) have been principal advocates for the initiative. When the initiative launched, use of the data cube, with the dimensions of space, time, and data type, was already a proven concept by Geoscience Australia and the Australian Space Agency and within development for their Landsat data archive. The objective was to have 20 countries operationally involved by 2022.[3] The pace, however, is exceeding the July 2017 plan. In July 2017, three countries (Australia, Colombia, and Switzerland) had operational capability. Four other were under development and twenty-one countries were under review. During a teleconferenced discussion with Dr. Killough and the Task Team 2 in mid-October 2017, he commented that 29 countries were already under review. In March 2018, during a briefing at the CEOS 7th Working Group for Capacity Building and Data Democracy Annual Meeting in Brazil, it was mentioned that at least 40 countries have entered into some level of discussion although the objective does remain 20. The speaker noted that Australia, Colombia, and Switzerland are still doing well. The United Kingdom, Uganda, Vietnam, Taiwan, Georgia, and Moldova are making progress. There are African regional data cubes in Ghana, Kenya, Senegal, Sierra Leone, and Tanzania. Therefore, the notion of the data cube is gaining interest and support. The global nature of the interest, however, adds to the complexity of how the US plans to expand its efforts with the Landsat collections.

Finding 1*:* Internationally the utility of the data cube for organizing Landsat data over time and location has growing acknowledgement to support time series analysis.[3] Colombia has found value in examining land change since 2000 and enabling understanding the trends for forest mapping and management. The main objectives of the Swiss Data Cube (SDC) are to support the Swiss government for environmental monitoring.[4] The Vietnam Data Cube is intended to create broad applications for socio-economic sustainable development goals for Vietnam as well as other countries in the region.[5]

Analysis Ready Data (ARD)[6] feed the formation of a data cube. Landsat 8 Operational Land Imager (OLI)/ Thermal Infrared Sensor (TIRS) Tier 1 and 2, Landsat 7 Enhanced Thematic Mapper Plus (ETM+) Tier 1,

---

[2] https://themes.jrc.ec.europa.eu/news/view/158675/earthserver-workshop-data-cubes-for-big-earth-data-19th-20th- october-2017-frascati-rm-italy

[3] Killough, Brian *Open Data Cube Background and Vision*, https://www.opendata cube.org/events July 7th, 2017

[4] http://www.swissdata cube.org/

[5] https://vnsc.org.vn/en/news-events/news/internal-news/introduction-of-satellite-data-sharing-system-vietnam-data-cube/

[6] https://landsat.usgs.gov/ard

and Landsat 4-5 Thematic Mapper (TM) Tier 1 comprise the contiguous US, Alaska, and Hawai'i ARD, which is available from the EROS Center, using EarthExplorer to download. Starting mid-March 2018, two new Landsat science products, Surface Temperature and Dynamic Surface Water Extent will begin to be integrated.

Dr. Robert Woodcock, who has worked for almost two decades in the field of visualization, spatial information systems and analytics and its application to Earth Science with a focus on ensuring research

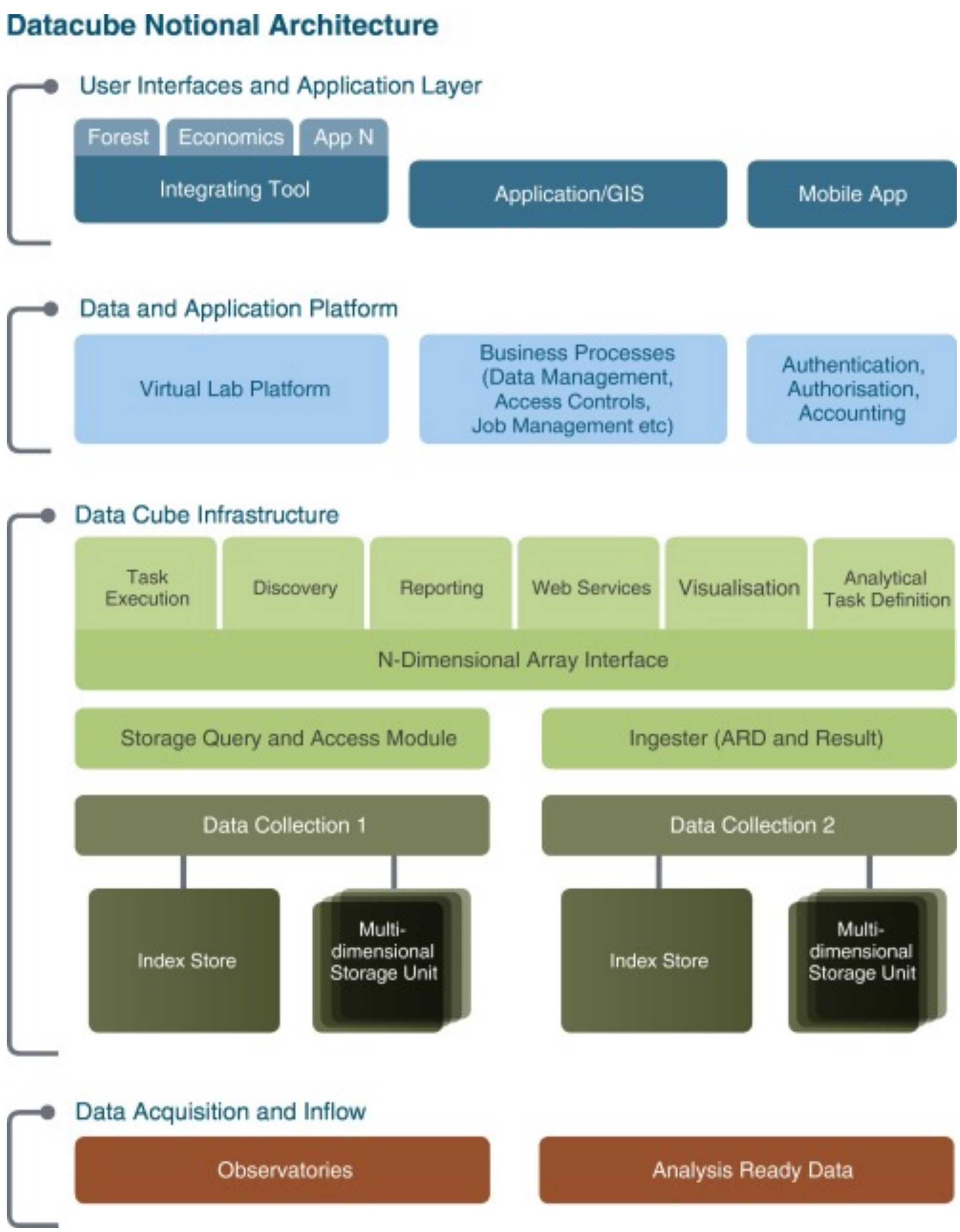

Figure 1. The architectural concepts of the Australian Geoscience Data Cube

innovation leads to business innovation, has been reinforcing the above-mentioned work with the Open Data cube initiative using his extensive experience. He, with some colleagues, prepared the diagram seen in Figure 1,[7] which describes a notional architecture employing the data cube. The four layers from bottom to top as follows:

Data Acquisition and Inflow - Observations are collected and pre-processed to an 'analysis ready' level by various custodians;
Data cube Infrastructure - analysis ready data are indexed into the AGDCv2 including ingestion into multi-dimensional datasets, with a suite of tools for task execution, discovery, visualization and so on;
Data and Application Platform - Platforms and environments that allow routine generation of products, and, exploration of new products in a 'virtual laboratory' environment; and
UI and Application Layer - A diverse set of applications is enabled by the underlying infrastructure.

Finding 2*:* The recommendations from the aforementioned LAG papers can be aligned with this notional architecture "to both enhance the scientific value of the imagery, but also to provide additional value to the commercial and government organizations wishing to extract the maximum value from the imagery" and to offer "potential new approaches to data management and distribution."

EROS Center and data cubes:

In November 2016, USGS/EROS provided the LAG team with a briefing on the Land Change Monitoring, Assessment, and Projection (LCMAP) initiative "to harness the Landsat record in order to provide state-of-the-art land change capabilities needed by scientists, resource managers, and decision makers." As explained during the presentation, to manage the resultant land-change products required addressing the issue "that the Landsat archive, currently organized as path rows, is not sufficiently efficient for time series studies. Moving to a grid-based data cube approach with API's that condition and serve data per user specification will reduce data preparation time." The data structure to be used was identified as an OLAP cube. The data content itself is the Analysis Ready Data (ARD) in the diagram above. The tiling scheme is modeled upon the Web Enabled Landsat Data (WELD) and will use the Albers Equal Area Conic projection and the Word Geodetic System 84 datum. ARD are standardized well-characterized radiometric and geometric products. Dr. Tom Loveland characterized the ARD as Landsat data processed to a level that enables direct use in applications.

- It will support geospatial, multi-spectral, and multi-temporal manipulations for the purposes of data reduction, analysis, and interpretation.
- It offers consistent radiometric processing scaled both to top-of-atmosphere (TOA) reflectance and surface reflectance.
- It is designed for consistent geometry including spatial coverage and cartographic projection – e.g., pixels align through time, <12m RMSE.
- It provides metadata on data provenance, geographic extent, and data quality.

[7] Adam Lewis, Simon Oliver, Leo Lymburner, Ben Evans, Lesley Wyborn, Norman Mueller, Gregory Raevksi, Jeremy Hooke, Rob Woodcock, Joshua Sixsmith, Wenjun Wu, Peter Tan, Fuqin Li, Brian Killough, Stuart Minchin, Dale Roberts, Damien Ayers, Biswajit Bala, Lan-Wei Wang *The Australian Geoscience Data Cube — Foundations and lessons learned* https://www.sciencedirect.com/science/article/pii/S0034425717301086

In simple words, ARD are intended to provide some pre-processed products that alleviate some work burden on the part of the users. Therein lies both its benefit for most consumers and concern for some other Landsat users that will be addressed later.

Questions Posed by USGS

In addition to Landsat, what other data sources (to include EO, SAR, and LIDAR) are optimally suited for leveraging (e.g., co-registered) to support data cube implementations for land change analysis and forecast modeling?

Among the efforts considered by the LCMAP team already has been to increase time series density by adding Sentinel-2.[8] In the CEOS initiative, Colombia and Switzerland are studying how to incorporate both Sentinel-1 (SAR) and Sentinel-2 (multi-spectral). The Vietnam prototype includes both Sentinel-2 and ALOS data. Progress was discussed on 6 March 2018, when the Vietnam National Space Center organized a workshop "Introduction of satellite data sharing system Vietnam Data Cube" in Hanoi.

One should not assume that all other data sources could, would, or should be housed by USGS/EROS. The data cube design must allow additional dimensions or layers to the cube. It will often be necessary for another government, academic or commercial organization to incorporate their own, sometimes proprietary, dataset to improve the results or to prepare tailored analysis. Thus, in Figure 1, one could consider analysis ready data to be multiple data sets that have been readied by some pre-processing to enter into the data cube structuring. Here as seen in Figure 2[9], layers of different data source products and extensions of more locations or times can be adaptively incorporated to address either some specific or generic issue. The graphic may obscure the reality that prospective "layering" demands consideration of some standardizing structure and functional guidelines.

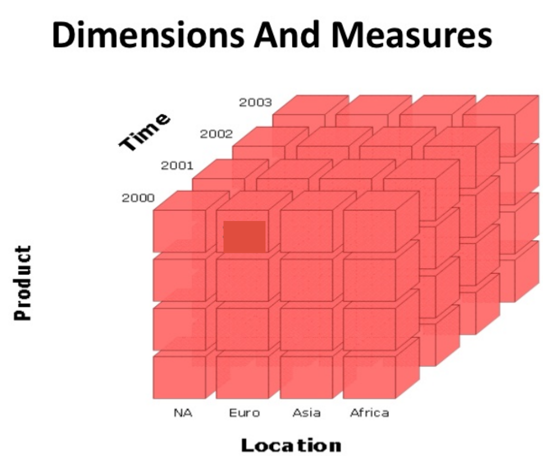

Figure 2. Graphic of Conceptual Data Cube

The notion that a variety of possible sources of data would accompany the ARD within the framework of the USGS LCMAP initiative was characterized in the graphic of Figure 3 provided by Dr. Loveland.

[8] Dwyer, John, "USGS Analysis Ready Data" presentation to the Landsat Science Team on January 14, 2016 and recent release indicating interest: https://landsat.usgs.gov/february-17-2018-us-landsat-ard-special-issue-call-manuscripts

[9] Adapted from https://www.slideshare.net/algum/data-cubes-7923771/5

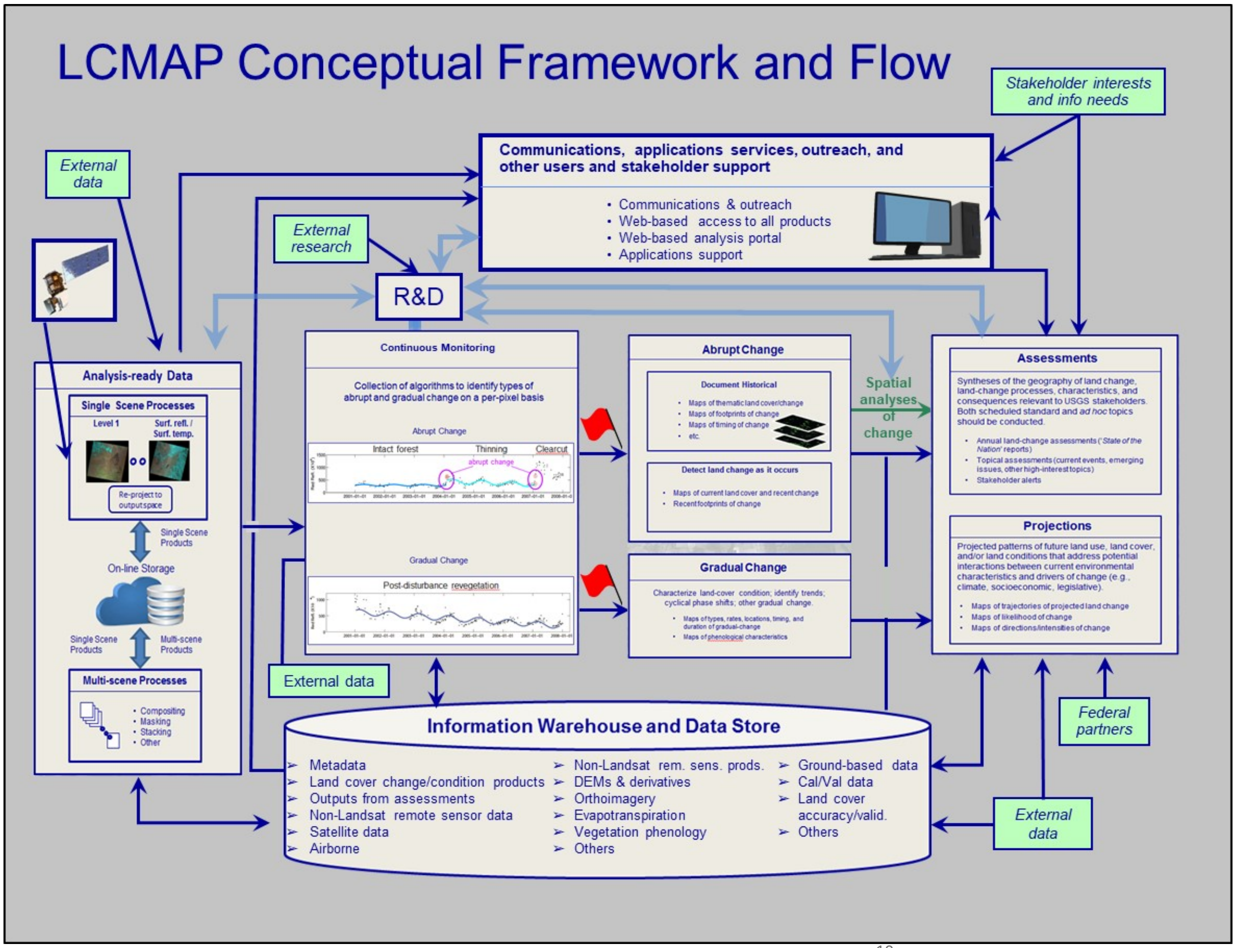

Figure3. LCMAP Conceptual Framework and Flow[10]

What kinds of Landsat time-series products would have the broadest community use or most impactful contribution in specific areas?

The Analysis-ready data (ARD) preparation, in general, rests upon foundational technology that can benefit nearly all users of Landsat data, not just a few specific applications. For example, ensuring that all data are consistently calibrated and carry appropriate quality-assurance metadata is of benefit to everyone, regardless of whether they are using data in one of the existing UTM grids or a new country-specific grid. The U.S. Landsat ARD tiling system is a modified version of the WELD structure. Three tile grid extents are defined for CONUS, Alaska, and Hawaii. The grid origins are defined in relation to the WGS datum but adjusted to align with the National Land Cover Database. They are country specific. In addition, the development of a US-specific ARD-based data cube in an Albers Equal Area Conic mapping projection is well-aligned with the mission of the USGS serving its US customers, as is preprocessing other geographically- coincident datasets to be available in that same projection. That approach both enables and facilitates the development of a range of US-specific higher-level data products and services. However, as data cubes become ubiquitous, what works well for the US may be quite awkward

[10] Loveland, Thomas *An LCMAP Overview: Land Change Monitoring, Assessment, and Projection, a Discussion with the Landsat Advisory Group and AmericaView Members*, November 16, 2016

for other countries. When the Open Geospatial Consortium (OGC®) first began its Discrete Global Grid System (DGGS)[11] working group, it sought to establish a specification to address collating spatial data from multiple places and sources and overcoming the challenges of working with different reference or grid systems. ARD present the Big Data challenge.

As explained by USGS/EROS, ARD are the foundation of LCMAP providing standardized well-characterized radiometric and geometric products (Level-1 Collection 1), the atmospheric correction and geo-physical surface reflectance and surface brightness retrievals (Level-2), and hierarchical metadata to include pixel-level attributes. Also available would be "open source" code to establish the processing and metadata standards, to accommodate scalable architectures, to deploy into public or private clouds. The latter are all points consistent with the recommendations from the Cloud Computing paper. Task team members endorsed this openness but recommend also the distribution of verification procedures that the methods and workflows have been replicated properly for any non-USGS production that incorporated other sources and initiates tailored analysis. Those procedures might mirror what USGS itself use. At this time, the ARD's "open source" code is accessible through https://github.com/USGS-EROS. The download of a tile still involves 5000x5000 pixels per collection event and any partitioning down to some smaller geographic footprint for a more local area occurs in the chosen environment of the user. Improvements to the lengthy and space demanding download and processing tasks are needed.

Finding 3: Non-USGS processing of data using the open-source code and algorithms available from USGS could necessitate that USGS also release procedures documentation and some verification test datasets.

The task team responsible for this report cautions that the USGS should ensure that its various efforts relating to Landsat data processing distribution are well aligned with each other and clearly articulated to the user community. In particular, the relationship between the Collection 1 reprocessing effort, existing Surface Reflectance processing and distribution efforts, and the Analysis-Ready Data effort, may need to be clarified. Fundamental improvements to processes, like sensor calibration, should be applied equally to processing and delivery of both UTM and data cube data. Similarly, both TOA and Surface Reflectance data are of value in all product forms and should be made available in a consistent manner. Keeping all these efforts aligned may minimize duplication of effort, but more importantly, it will avoid user confusion, which could otherwise lead to erroneous use of data by end users.[12]

Recently brought to the attention of the Task Team has been the voice of those who worry about the impact of "normalizing" the reflectance product across all the collections. From their perspective, they agree that pre-processing the Landsat data into this "normalized" state so that time-series analysis of multiple collections over a large area brings great efficiencies by reducing processing burden on a many or even most of the users. What the concerned group questions is what error is introduced in that pre-processing that might affect analysis of smaller footprints and more restricted time sequences. Importantly, they are not claiming that significant errors might result. Rather they are concerned that whatever analysis may have been completed before moving ahead with ARD has not been quantified for them. They endorse that the Level 1T products will remain available and will want to do more study on

[11] http://www.opengeospatial.org/projects/groups/dggsswg
[12] Steven J. Covington, Principal Systems Engineer for the USGS Land Remote Sensing Program, commented Current thinking has Collection 2 encoded with Cloud Optimized GeoTIFF (COG) to enable efficient extraction of user-defined areas smaller than the planned storage granule (a WRS Scene)

the algorithms that have been used to create ARD so they can reliably assess the error, if any or negligible, introduced into the data. A recommendation would be that USGS EROS Center release any study analysis completed on the error impact of the preprocessing or initiate such a study. (It is recognized that if there is concern that cannot be resolved, one can reverse the process that produced the TOA product and have the historical radiance product.)

Which organizations with expertise in forecast modeling are best postured to evaluate and demonstrate the forecast potential from a Landsat-based temporal data cube?

Much has been written about problems of forecasting with any Big Data, including all the imagery and geospatial collections - with the foremost challenge being the lack of personnel skilled for this task. The tiling scheme chosen for ARD and applied to the Landsat images over the US should assure alignment of tiles so that "drilling" though several images over the same geography provides the same footprint for subsequent time series analysis that could lead to forecasting future conditions based upon past information. It is trusted that rigorous testing has been done to assure the layered footprints over time are positioned within some defined degree of positional accuracy.

One objectives of the ARD effort could be to improve use of "big geodata." Within some of the research and analysis work with large quantities of geospatial data has been discussion of the frustrating insufficiency of traditional statistical techniques or of the challenging selection of the most appropriate statistical technique to obtain reliable and consistent forecasts from large quantities of data. In the initial releases of the Landsat ARD and the temporal data cube, it would be wise to consider the use of academic research centers to assess how much the new structure actually facilitates analysis and to encourage universities to revise classroom modules that prepare the future analysts and information managers. Will ARD enable better forecasts with Big Data using a variety of novel techniques? Not only can academic organizations be excellent partners with the government using these vast stores of data but also several private companies will be eager to use the ARD and build versions of the data cube tailored to support processing that delivers the answers needed by their customers.

How far back in time into the Landsat archive should the staging of 'analysis ready data' be considered? E.g., early data collections such as multi-spectral scanner (MSS) data are less equipped (in terms of metadata) to support rigorous geometric and radiometric calibration compared to later collections.

The decision to include the MSS data has been strongly recommended within USGS at the EROS Center. Addressing the question may be a moot point, given its value in the long term of continuous Earth imaging and observation and its inclusion being strongly recommended by some members of the previous Landsat Science Team. However, this task team strongly recommends that prioritizing development work should be carefully scrutinized within USGS. Is global ARD without MSS of greater value, to a growing international community of users, than US ARD with MSS? In addition, following some of the concern about forecasts from massive data stores, the issue of signal to noise (not easily mitigated by the seriously diminished amount of metadata for MSS data) should also be evaluated.

How could efficient synergy be realized among government and commercial roles for data cube development, and operations (processing, storage, distribution) to satisfy broad community needs?

Caution was urged by team members about how much of the production workload should be assumed by USGS. The analysis of the ARD, as ingested into the data cube infrastructure, should not be solely dependent on the computing infrastructure of the USGS, which is unlikely to have ready access to some of the latest technology advancements, given the budgeting processes. Many of the organizations very interested in the promise of LCMAP might need attention more focused on specific areas that the EROS Center had not planned to address immediately. Those specific areas might have larger or smaller footprints or they might be outside the US. It is not clear that the government is prepared for such flexible response for building such specific datacubes, nor has any good justification been provided for why the government should assume that role of production. Members of the Task Team highly recommended more consideration of the private public partnership concept in the end-to-end process from Landsat level 1 products to ARD to user-tailored data cube. The Task Team agreed that USGS, as the Landsat source experts, should be responsible for Landsat ARD quality and consistency, although they would likely benefit from commercial support for the processing and distribution infrastructure.

***Finding 4**:* The commercial sector is ready to provide data cube tailoring assistance, given its increasing experience with global geospatial data. It is also prepared to provision infrastructure to assist in the production of ARD.

As the needed tools and techniques mature, the team similarly recommends that USGS should not undertake to scale this country-specific effort globally themselves. There is no one peerless global projection coordinate system. Given specific needs, any spatial multi-dimensioned data cube can be quite parochial, and each country or region that wants a data cube would likely select their own tiling grid to minimize distortion in their region and maximize interoperability with other existing regional datasets. The USGS should focus on opening up its tools and the necessary input datasets so that third parties in the private sector can offer a service of building these data cubes for global customers in accordance with USGS best practices. Such a scenario, might also involve other countries producing their own ARD, and if from Landsat, that could require the USGS to release image data (perhaps Level 0), DEM, GCP data, and all other necessary inputs in addition to the code that USGS uses to create the US ARD product. In this way, the USGS could focus on developing expertise and on building operational systems for the US, without straying into building operational systems for the world. The concern about USGS producing either ARD or data cubes for the global customer relates back to the earlier description of both a mapping projection and a grid system that do not apply well globally. That raised the question about the priorities of the USGS production plans and how and why the private sector can step forward.

The CEOS initiative is not without questions for similar challenges. Even if global stakeholders agree that an Open Data cube vision has promise, will they make their contributions to mitigate the risk that the concept cannot be scaled with limited resources? Given the adoption of the concept and the development of national data cubes under the CEOS initiative, having excellent transformation algorithms for the projections would allow necessary flexibility. The tiling scheme, however, could be far more challenging, if and when adjacent countries build national data cubes and select differing schemes. The role of CEOS in establishing or instantiating standards and specifications, like those in the DGGS mentioned above, should not be underestimated.

Previously this paper mentioned standards with respect to the open software standards needed to avail any requester of the software, who might require the algorithms used by USGS in preparing the ARD at any point in the anticipated improvements over time.

Finding 5: The data cube implementation involves a broad scope of standards issues.

- In February 2018, after an informal discussion of the topic, an OGC group prepared an OGC discussion paper: "In response to a recent discussion (via the OGC email lists) regarding perceptions about data cubes and DGGS, it was suggested that we begin a more formal discussion on this topic within the OGC Technical Committee. This information document aims to initiate a discussion of the broader definition of a data cube and the complementary role that DGGS technologies play."[13] The following future actions were identified. "…the Authors recommend some targeted actions with which we should proceed in collaboration with the community of the DGGS specification and domain groups. These actions mainly focus on investigating the efficiency on querying and exploring large multi-dimensional arrays while using the DGGS technologies on Datacubes. These activities will be exercised under specific ongoing big data research international projects."

- Also in February, the Open Geospatial Consortium announced that it was seeking public comment on Web Coverage Service (WCS) 2.1 Candidate Standard. The qualifier statement for the announcement read "*Updated WCS 2.1 Standard will simplify access to spatio-temporal 'big data cubes'".*[14] The release also offers more explanation. "By supporting the more general data cube model of CIS 1.1, the WCS 2.1 standard will simplify access to spatio-temporal 'big data cubes', with an operation spectrum ranging from simple sub-setting in space and time up to complex spatio-temporal analytics through Web Coverage Processing Service (WCPS). WCPS offers a protocol-independent language for the extraction, processing, and analysis of multi-dimensional coverages representing sensor, image, or statistics data, such as might be enveloped within a data cube.
- In 2017, Dr. Peter Baumann, Professor of Computer Science, Jacobs University Bremen, published a positively provocative paper within the community of interest, named the *Data Cube Manifesto*[15], in which he commented, "Recently, the term data cube is receiving increasing attention as it has the potential of greatly simplifying "Big Earth Data" services for users by providing massive spatio-temporal data in an analysis-ready way. However, there is considerable confusion about the data and service model of such data cubes." That statement was followed by his six principles of data cube service concluding with the sixth being "Data cubes shall support a language allowing clients to submit simple as well as composite extraction, processing, filtering, and fusion tasks in an ad-hoc fashion…The OGC data cube standards, CIS and WCS/WCPS, are embraced by open-source and proprietary implementers, coming with compliance tests enabling interoperability down to the level of single pixels. Availability of data cube standards and tools is heralding a new era of service quality and, ultimately, better data insights."

The Task Team recommends that OGC be encouraged to continue work on the standards that support the agile and reliable and consistent use of a data cube approach. This would help address this section's question about the efficient synergy between public and private sector use to meet customer/client requirements.

Another quite relevant point that has emerged during the months of discussion on this LAG task assignment has been the question of local or cloud storage and/or processing. Assumptions about the desire for nations to want all their data downloaded to their own servers rather than preferring the value-added solutions provided by the cloud service providers are not necessarily reinforced by the

[13] Purss, M., Peterson, P., Strobl, P., and Sabeur, Z. *Discussion Paper: A DGGS Perspective on Datacubes 18-006*, 14 February 2018 (Permission to use: The OGC Working Group that developed the paper approved release for use by the NGAC membership. 26 March 2018.)

[14] http://www.opengeospatial.org/pressroom/pressreleases/2738

[15] Baumann, Peter, The Datacube Manifesto http://earthserver.eu/tech/datacube-manifesto Research supported by EC contract 654367

emerging evidence.[16] In some countries, the tools to work with the massive data are either not available or the skill levels are currently inadequate. The private sector working closely with the national government's imagery could dramatically simplify data use when solutions rather than data is the desired outcome. Cloud computing, in addition or in lieu of cloud storage, may be the tailored approach. Even when data are what may be needed, the response could be a tailored data cube provision where the national data, like ARD, are layered with other data sources and refined to a particular footprint, consistent with the preceding discussion in this study.

One question raised was how the private sector might collaborate to help with tiling the additional sources to match that of ARD as the layers of the data cube are incorporated. During the CEOS briefing in Brazil in March 2018[17], the topic of cooperation with the private sector under some grant agreements was included as a needed facilitator of the global effort. Partnerships with Google, Amazon, and others were seen as enabling the "scalable solution."

Major Recommendations

This report makes some specific recommendations, specifically with respect to the U.S. Landsat Analysis Ready Data (ARD) and its potential for being incorporated in a variety of datacubes, as a direct-use dataset in monitoring and assessing landscape change.

1. Task team members endorse the openness of the EROS Center commitment to provide the source data and to publish, as "open source," the software and algorithms used to produce ARD. The Team recommends the USGS should publish verification procedures that the methods and workflows have been replicated properly for any non-USGS processing. These procedures would likely reflect the very processes that USGS has used in preparing ARD. The verification task itself would not be the responsibility of the EROS Center but rather of any other entity using the software and algorithms.

2. Studies may already exist that characterize how "normalizing" reflectance across sensors and years might affect values. The Task Team recommends that USGS EROS Center release any error/difference study and analysis between the reflectance values of traditional scene pixels and the ARD unit pixels, which may have already been completed, to determine any radiometric changes resulting from preprocessing to create the ARD. Offering access to those studies could be beneficial to some researchers. If such an analysis has not been completed, the Task Team recommends that one be initiated.

3. The Task Team expects processing techniques, algorithms, and associated tools to improve over time. Reprocessing the entire data set, vice limiting new approaches to only data acquired *after* the development of improvements, would meet the "analysis-ready" objective of reducing the data processing load of data uses. The Team believes that complete revision of the entire ARD could follow a MODIS approach. The team was advised by USGS that such processing of so much data could take up to ten months so a reasonable schedule for updates will need to be established. The Task Team recommends that when improved processing approaches are ready, the reprocessing

[16] The CEOS Data Cube, Three-Year Work Plan 2016-2018 http://ceos.org/document_management/Ad_Hoc_Teams/SDCG_for_GFOI/Meetings/SDCG-10/Cube%203-Year%20Work%20Plan%20-%20v1.0.pdf.

[17] Holloway, Kim "Open Data Cube Initiative" Agenda item #8, CEOS 7th Working Group for Capacity Building and Data Democracy Annual Meeting, INPE Jose dos Campos, Brazil, 6-8th March 2018

should apply to the entire data set in use and that users should not be required themselves to apply compatibility adjustments to any ARD received prior to the change.

4. The Task Team does agree that MSS should be incorporated into ARD to optimize use of the entire forty-five years of collection history. However, the Team recommends that prioritizing development work should be carefully scrutinized with consideration given to whether globally extending ARD may be more important than spending available time incorporating the MSS collection. In general, it is recommended that USGS assess all needs and wants and establish criteria to prioritize Landsat work, including enhancements to the ARD initiative.

5. The Task Team recommends that USGS should not undertake to scale the US ARD coverage effort globally by themselves, as the private sector is better prepared with needed tools, mature techniques, and, particularly scalable infrastructure.

This report also makes recommendations about geospatial datacubes, as they become more globally employed to manage and exchange information for a variety of applications.

1. The Task Team recommends that USGS representation, as a Strategic Member, to the Open Geospatial Consortium should advocate for and participate in more discussion about datacube standards within the OGC Technical Committee.
2. The Task Team recommends that preparing datacubes for specific uses should not be an objective of the government, which should be cautious about proceeding even with production of some generic forms of a datacube. The tailored data cubes should not be a federal government production responsibility.

Additional recommendations are made with reference to this report.

1. The Task Team recommends that a subsequent request be made to a future LAG Team to evaluate progress on the findings and recommendations of this paper and to update as needed.

2. The USGS has only fledgling experience with ARD, having first released it to the community of users at the end of October 2017. At this point, there has not been extensive experience on the part of ARD users and certainly not much evidence of the resulting datacubes. It would be helpful for USGS to survey those who request the ARD on some routine basis, gathering information for a subsequent report. Among the factors to be surveyed would be if users are transitioning to ARD or still requesting the previous distribution formats. The Pecora Conference in mid-November 2017 provided an initial opportunity for groups of Landsat users to discuss their early reactions to the release of ARD. Since that time, use has increased but not all users are fully comfortable knowing how to use the data to its best advantage. Similarly, on-going information exchanges between the public and private sectors may provide more insight into defining the interdependencies to make datacubes the most effective way to advance use of imagery and expansion of GIS technology.

Acronym List for this Paper

ALOS	Advanced Land Observing Satellite Japanese Earth-observation satellite, developed by JAXA (Japan Aerospace Exploration Agency)
ARD	Analysis Ready Data
CEOS	Committee on Earth Observation Satellites
COG	Cloud Optimized GeoTIFF
CSIRO	Commonwealth Scientific and Industrial Research Organisation is an independent agency of the Australian Federal Government responsible for scientific research in Australia.
DGGS	Discrete Global Grid System
EO	Electro-optical systems operate in the optical portion of the electromagnetic spectrum.
EROS	Earth Resources Observation and Science, a USGS Center near Sioux Fall, SD
ERS	Earth Resources Satellite, the first two remote sensing satellites launched by ESA
ESA	European Space Agency
ETM+	Enhanced Thematic Mapper Plus, a sensor onboard the Landsat 7 satellite
FGDC	Federal Geographic Data Committee
GeoTIFF	Georeferenced Tagged Image File Format, a public domain metadata standard which allows georeferencing information to be embedded within a TIFF file
L1	Level-1 Landsat products with the best available processing level for each particular scene
L1G	Level-1 Landsat radiometrically calibrated with systematic geometric corrections using spacecraft ephemeris
L1T	Level-1 Landsat radiometrically calibrated and orthorectified using ground control points and digital elevation model data to correct for relief displacement
LAG	Landsat Advisory Group
LCMAP	Land Change Monitoring, Assessment, and Projection, a USGS initiative implemented at EROS
LIDAR (Lidar, LiDAR)	Light Detection and Ranging, a remote sensing and surveying method that measures distance to a target by illuminating the target with pulsed laser light and measuring the reflected pulses with a sensor
MSS	Multi-spectral scanner, line scanning devices observing the Earth perpendicular to the orbital track on the first five Landsats
NASA	National Aeronautics and Space Administration
OGC®	Open Geospatial Consortium
OLAP	Online analytical processing, use of data organized multi-dimensionally to allow comparisons from different perspectives
OLI	Operational Land Imager, a push broom scanner on Landsat 8 that uses a four-mirror telescope with fixed mirrors
SAR	Synthetic-aperture radar, a technique for producing fine resolution images from an intrinsically resolution-limited radar system
SDC	Swiss Data Cube

TIRS	Thermal Infrared Sensor, a system on Landsat that measures land surface temperature in two thermal bands
TOA	Top-of-atmosphere
USGS	U.S. Geological Survey
UTM	Universal Transverse Mercator, a coordinate system which divides the Earth into 60 zones, each 6° of longitude in width
WCPS	Web Coverage Processing Service, a protocol-independent language for the extraction, processing, and analysis of multi-dimensional coverages representing sensor, image, or statistics data
WELD	Web Enabled Landsat Data
WRS	The Worldwide Reference System, a global notation system for Landsat data

Acknowledgements

This paper was approved by the NGAC Landsat Advisory Group (LAG) on March 20, 2018 and adopted by the NGAC as a whole on April 3, 2018. The LAG team developing this paper included Roberta Lenczowski, Roberta E. Lenczowski Consulting (Team Lead); Frank Avila, National Geospatial-Intelligence Agency; Peter Becker, ESRI; Steven Brumby, Descartes Labs; Rebecca Moore, Google Inc.; and Tony Willardson, Western States Water Council. Matthew Hancher (Google, Inc.) and Sara Larsen (Western States Water Council) also contributed to this paper.