Future of Big Earth Data Analytics

Christopher Lynnes[1]

Thomas Huang[2]

[1]NASA/Goddard Space Flight Center (Civil Servant)

[2]NASA/Jet Propulsion Laboratory (employed by California Institute of Technology)

Corresponding author: Christopher Lynnes (christopher.s.lynnes@nasa.gov)

Key Points:

- Improvements in sensor and platform technology are improving data resolution in multiple dimensions, resulting in enormous increases in data Volume.
- The proliferation of instruments and platforms will lead to increased Variety amongst Earth Observations, with a concomitant need to assess data quality (Veracity).
- Cloud computing and the rise of machine learning techniques and communities will increase the capabilities of end users to derive information from Earth Observation data. However, classical statistical analysis will remain important, particularly in the roles of elucidating how phenomena work and as a tool for interactive data exploration.
- Analytics architectures will strive to colocate the data and the analysis in order to bring massive computing to bear on large datasets.

Abstract

The state of the art of Big Earth Data Analytics can be expected to evolve rapidly in the coming years. The forces driving evolution come from both growth in the data and advancement in the field of data analytics. In the data area, advances in sensor instrumentation and platform miniaturization are increasing both data resolution and coverage, resulting in enormous growth in data Volume. Increases in temporal resolution in particular also generate demands for higher data Velocity. At the same time, the proliferation of instruments and the platforms on which they reside is increasing the Variety of datasets. The Variety increase in turn leads to questions about the Veracity of the data. In the algorithm area, powerful machine learning methods are coming to the fore, particularly Deep Neural Networks. These are powerful at detecting interesting features in the data, integrating many different measurements (i.e., data fusion), and classification problems. However, they are still challenging when seeking explanations of how natural or socio-economic phenomena work using Earth Observations. Thus, classical analysis techniques will remain relevant when the emphasis is on forming or testing explanations, as well as to support interactive data exploration.

1 Introduction

The state of the art of Big Earth Data Analytics can be expected to evolve rapidly in the coming years. The forces driving evolution come from both growth in the data and advancements in the field of data analytics. It is useful to examine the sources of the "Big" in Big Data in order to try to identify how the data landscape is likely to evolve. These sources include technology advances that are improving data resolution in virtually every dimension with a concomitant increase in data Volume. Meanwhile, technology advances have also led to significant growth in the sheer number of instruments and observing platforms, leading to a Variety increase. At the same time, the field of data analysis has exploded with the widespread availability of highly usable code for a variety of analysis problems in several coding languages. Both of these factors are attracting new practitioners to the field with fresh perspectives and requirements.

2 How Data Get Bigger

The volume of Earth Observation data in major data systems has grown exponentially over the last two decades or so. Fig. 1 shows the cumulative archive size of the NASA Earth Observing System Data and Information System (EOSDIS) over the period 2000, the dawn of the Earth Observing System, to 2017. Over that time, the annual growth in the EOSDIS data archive has averaged about 25% per year. Where does this kind of growth come from?

To begin with, the simple passage of time increases the data record for many flagship satellite instruments, though this is merely a linear increase with time. Advances in instrumentation, however, are pushing horizontal resolution to higher values, causing a power of two growth factor for imaging instruments. Advances in sensor development have contributed to resolution improvement. An example of this can be seen in a comparison between the Ice, Cloud, and land Elevation Satellite (ICESat) satellite (Schutz et al., 2006), which launched in 2003 and its successor ICEsat-2 (Moussavi et al., 2014), which launched in 2018. Improvements in detector sensitivity enabled the use of lower power in the 532 nm lasers that generate the photons. As a result, while the ICESat instrument GLAS generates 40 pulses per second, the ATLAS instrument on ICESat-2 generates 10,000 pulses per second, a 250X increase in altitude

readings. More sensitive and energy-efficient sensors in turn enable the miniaturization of the observing platforms, both spaceborne (e.g., Cubesat) and airborne (drones). The increased affordability of small but capable Earth Observation sources is likely to fuel significant growth in the volume of Earth Observation data generated not only by national space programs but now a wider community of commercial and other non-governmental organizations, some of them quite small. Sensor improvements help in other ways as well: the ability to more accurately measure additional wavelengths or other quantities (such as Lidar returns) will enable scientists to retrieve geophysical or biological quantities that were hitherto not reliably retrieved. Likewise, improvements in horizontal resolution allow us to resolve finer-scaled physical processes than before, which also contributes to new data products. The fine resolution has another indirect effect, in facilitating and incentivizing the downscaling of models that either incorporate the data (e.g., as initial or boundary conditions or in assimilations) or are validated against the data.

For geosynchronous satellites, such as the Geostationary Operational Environmental Satellites (GOES) that collect data for the National Oceanographic and Atmospheric Administration (NOAA), contributions to data growth come from an even greater variety of sources. In addition to a factor of 2 increase in horizontal resolution in the GOES-16 and later satellites (resulting in factor of 4 data increase), the spectral resolution is increased from 5 to 16 spectral bands in GOES-16, and temporal resolution is increased by a factor of 3, from 15 minutes between snapshots to 5 minutes. Beyond this, the instruments can operate in a variety of modes to increase spatial coverage or temporal resolution as needed, and the later generation instrument obtains 12 to 14 bits of information compared to the previous 10 bits.

Many of the factors driving data volume growth (miniaturization, cost reduction, sensor improvement) have a secondary effect on one of the other Big Data V's, namely Variety. The diversity of inexpensive sensors and platforms will require some adaptation in order to use the data emanating them. Airborne instruments in particular can be difficult to work with at the data analysis phase, particularly when attempting to synthesize results across large areas. Even the increasingly diverse community of data providers is likely to introduce a variety of data processing methods which may result in a similarly diverse set of data products, even when the same basic quantity is being measured. Biases among different instruments and processing algorithms will become increasingly important to account for.

Another contributor to the Variety problem comes from the increase in data resolution, spatial, temporal and spectral. All of these are likely to contribute to being able to sample the Earth's geophysical properties over a wider domain, and thus under a wider variety of observing conditions. Variations in, say, land cover, that would have averaged out over a 90 m pixel may now be observable at a 3 meter resolution. This changes the statistical properties of many measurements, especially land-based ones, where spatial variation is generally stronger at smaller length scales than at oceanic or atmospheric length scales. Still, even oceanic instruments such as the Surface Water Ocean Topography (SWOT) are expected to illuminate phenomena that have not hitherto been observed due to its ability to resolve finer scale changes in surface water height than any previous satellite-borne instrument (Durand et al., 2010).

With the increase in data Variety comes concerns about the Veracity of the data. In the past, most Earth Observation (EO) data were produced by large science teams with instruments and algorithms that had been thoroughly reviewed by the community leading up to launch. Many such facility or agency instruments also had dedicated teams for calibration and validation of the data products. However, the proliferation of smaller data collectors and producers make the establishment of data quality both more critical and more difficult at the same time. Robust

machine learning algorithms can help to scale out quality assessment by establishing biases using in situ and other validation data (Lary et al., 2018).

Our Big Data challenges are not only due to the advancement of high-resolution instruments. The ease of access to a vast array of data opens up the ability to derive complex scientific inferences. With recent advancements in affordable low-power teraflop-scale computing technology, along with machine learning and the ability to collocate data analysis capabilities next to the sensor in the age of Internet of Things (IoT), it is now possible for researchers and decision makers to access, analyze, and derive inferences across collections of multivariate measurements. Edge Computing is a computing paradigm through connected smart devices. Rather than only investing in creating solutions for moving massive data to the Cloud, smart instruments will offer higher quality data and analysis results before transmission to the ground. With low-cost, high-performance smart instruments, it is possible to have science data processing directly at the sensor and deliver the high-quality final products to the ground. With automatic detection capability on the sensor, it is possible for the sensor to detect physical phenomenon and quickly direct other subsystems to acquire other relevant measurements. While our Big Data challenge will only get bigger, our connected world and low-power computing technologies will deliver higher quality information at rapid pace.

3 The Evolution of Analytics Algorithms

As we have seen above, the growth of data Volume and Variety bring both tremendous opportunities as well as new challenges. However, the growth of data analytics capabilities available to the user community is growing as well. In particular, the explosion of machine learning has been remarkable. There are now multiple well-written off-the-shelf frameworks to support machine learning in a variety of languages, though the largest explosion has been in the Python and R areas. Most of the interest appears to be in the area of Deep Neural Networks (DNNs), which have become well-used with the arrival of accessible computing power, including processing units particularly suited for DNNs such as Graphics Processing Units, or even purpose built neural processing units (Van der Made and Mankar, 2017). In addition, several variations on the theme of DNNs have begun to be deployed to tackle EO problems. For example, one of the problems with scaling land cover and land use studies to global scale is the difficulty with applying a model trained in one area to work in other areas. Transfer Learning shows promise for solving this conundrum.

Another significant challenge to DNNs is the availability of deep learning training data., though there is a variety of approaches to address this. One approach is to leverage crowdsourced labels, particularly in the area of land cover classification (Johnson et al., 2017). Another approach is to use physical retrievals, either real or synthetic (Rodriguez-Fernandez et al., 2017). However, a more common approach that is coming to the fore is to use Generative Adversarial Networks (GAN) to create realistic synthetic data. In this semi-supervised method, a small amount of real, labeled data is supplemented by fake data generated by the generative neural network. The data are classified by a discriminator neural network that attempts to classify the data while detecting fake data (as one of the classes). Meanwhile, the generative network generates ever more realistic data as it attempts to keep its data from being recognized as fake. The method is broadly applicable and can be expected to be deployed ever more frequently.

One of the challenges that DNNs may help to solve is the variety of instruments, particularly in the optical sensing category where there are a myriad of commercial and government sponsored satellite borne sensors, complemented by drone-borne sensors. Ideally,

we would like to be able to combine the stability and well-known quality characteristics of high-quality government sponsored satellites like Landsat with the vast array of measurements coming from small CubeSat based commercial sources. However, combining these data correctly is difficult in the face of the variations in geolocation and radiometric quality, as well as the designed response characteristics. Neural network based data fusion, however, should be able to tackle the biases and uncertainties involved in the combination of observations. Due to the meteoric rise of DNNs and their versatility in tackling problems, other supervised machine learning techniques will still continue to be relevant, though possibly in niche roles. In the area of land use and land cover classification, decision trees and random forests are still common, for example.

One important limitation of many machine learning methods is that the derived model is often a "black box", providing little insight in why a given answer was produced by the model. For many practical applications, this is not an issue: the user is looking for the best answer to be obtained given the available data. Some estimate of expected precision and recall is usually available for machine learning exercises, allowing a certain amount of confidence to be placed in the statistical properties of the end results. However, it can be hazardous to place too much faith in any individual prediction. As a result, recent research sometimes uses machine learning as a guide for sampling to produce estimates of aggregate properties, such as the work by Song et al. (2017) to estimate area under soybean cultivation with satellite imagery, thus marrying classical statistics with machine learning.

The "black box" issue is more troublesome in the area of scientific research, where many (though not all) studies are aimed at understanding how the world works. The mathematical and statistical models that result from machine learning rarely provide direct insight into phenomena and their inter-relationships. Nonetheless, machine learning techniques can still find use in a number of roles for scientific research. One fruitful area is mining datasets for interesting events such as storms, or even estimating the storm strength (e.g., Pradhan, 2018), i.e. as adjunct methods for inferred measurements or data fusion. However, what is desired is a physics-, biology-, or socioeconomic-based model to explain the observations, because the fundamental goal is to further the understanding of the observed phenomena. One way to apply machine learning in this case is to predict quantities using EO proxies, such as night lights or roofing type to predict economic activity (Jean et al. 2016). The success of the machine learning algorithm can then shed some light on how well the proxy quantities correlate with the target quantity. The value of EO data in this case is the ability to investigate such phenomena relationships at larger spatial scale, and with higher temporal resolution. Alternately, sensitivity studies using known perturbations to the input data may at least help illuminate the inner workings of the model (Sundararajan et al., 2017), even if they do not lead straightforwardly to insights of the workings of the phenomena themselves.

The EO domain presents some unique challenges in the Big Data analysis field. One of these is the challenge of bridging the difference in scale between high spatial resolution / low coverage data (e.g., drone-collected) and lower spatial resolution / global coverage . The gap between local and global datasets is exacerbated by the irregular spatial and temporal sampling of the small, numerous and diverse platforms and instruments collecting the former. This will place a premium on fusion and assimilation methods that can incorporate both scales of measurements. Machine learning techniques are being used more frequently to aid the downscaling process (e.g., Liu et al., 2018).

Another factor in moving to finer-scale phenomena is that the diversity of local observing conditions becomes more important. For example, when looking at forests with EO data, individual trees, including their species and habit, become important at 1 m scale. Whereas in the past, local studies were commonly undertaken by researchers with some local knowledge, it is now possible to obtain local-scale resolution over continental areas. Furthermore, the advent of cloud computing can provide the necessary scaling up of computing power to conduct broad continental scale studies using high-resolution data on trees, such as the study by Brandt et al. (2018). However, local contextual knowledge is not too easy to scale to much larger regions. For instance, local cultivation practices may affect land use estimation.

Although machine-learning-based analytics is growing quickly, classical statistical approaches also benefit from the increased availability of both data and computing. As data resolution improves, it becomes possible to study the statistics of fine-scale (either spatial or temporal) phenomena. At the same time, running classical statistical analyses on Big Data is likely to require Big Computing as well: often such analysis is embarrassingly parallel, but only if the data are organized along the access of the statistical computation. These statistical analyses are particularly prevalent in supporting interactive exploration of data. The advent of cloud computing, in conjunction with optimizations to data storage and organization in order to accelerate the computing, promise potential reductions in analysis time that can be measured in orders of magnitudes. This raises the possibility of supporting exploration of entire datasets in a fully interactive mode.

4 Analytics Architectures

The increases in data volume together with the growing appetite for data of analytics algorithms will place a premium on data system architectures that can feed the algorithms with data at high rate. In most cases, this relies on distributing the data through the cloud in such a way that coincident (or proximal) computing power can produce a result independently of other data-compute combinations, i.e., the mapping phase of a map-and-reduce algorithm. The system then collects the results and performs any necessary reduction operation. A variety of scalable filesystems and highly distributed databases have been shown to provide this capability. However, it is still important to consider both the data and the analysis to be performed so that the mapping of the data can accomplish enough of the processing to provide the desired speed-up.

One of the interesting effects of data improvement is that new Big Data requirements may appear in indirect forms. Increasing the time resolution obtainable from geosynchronous data, for instance, drives an increased demand for decreasing the latency of the data delivery, in other words the Velocity aspect of Big Data. In general, latency expectations tend to be slightly less than, but on the same order of magnitude as, the temporal resolution. Data that have, say, a monthly resolution are not expected to be available on a near-instantaneous basis on production. On the other hand, the increase in time resolution of the GOES-17 and later satellites was accompanied by ground system latency requirement of 50 seconds for processing the full disk to calibrated radiance at 5-minute resolution; the latency requirement is a mere 23 seconds for mesoscale (regional) mode, which has a latency of 30 seconds (Kalluri et al., 2018). The data Velocity is sometimes complicated by the desired processing mode: land cover classification, for instance, often benefits from a Time-First, Space-Later approach in which the temporal history of a pixel is used to classify the land cover or land use, meaning that the system must retain enough pixel history to see the parameter's autocorrelation. On the other hand, the spatial

distribution of the data helps to parallelize the processing in this case. For example, Assis et al. (2017) show an architecture based on MapReduce and Hadoop streaming to process data from the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument, resulting in a streaming analytics system for remote sensing imagery. Similar approaches are likely to proliferate as the number of satellites available increases the effective time resolution and latency demands of the user communities.

Future Big Earth Data Analytics architectures will also need to address the data management complexities that come with restructuring the data. This restructuring often consists of two main aspects: preprocessing the data to make it "analysis-ready" and reorganizing the data to optimize them for cloud analytics processing. The criteria for analysis-ready data varies to some degree with the user community and the type of analysis desired. The creation of analysis-ready data may include applications of various corrections to increase the interpretability of the data, such as calibration, atmospheric correction, terrain correction, quality filtering and geophysical parameter retrieval. For data comparisons or time series analysis, spatial and/or temporal regridding is often also applied. Analysis-ready data are then reorganized into a structure and stored in a framework that makes the data faster to analyze. Both of these processes make changes to the data from the typically file-based data products from which they were derived. This results in a formidable data provenance challenge: science users need to be able to tell what the original data were that went into an analytics result, as well as how the analysis-ready preprocessing and data organization affected the data content. While some provenance frameworks have been developed for processing remote sensing data, they also tend to work with large packages of data, such as files, images or coverages (Jiang, 2018). Spreading those same data values across a distributed filesystem or database for fine-grained access by computational algorithms presents a more difficult problem, particularly with respect to how to make the provenance documentation usable by the end user, who may want to trace back the contributions going into a particular pixel time series. In short, it is not clear that existing frameworks from the International Organization for Standardization (ISO) or the World Wide Web Consortium (W3C) scale well with this data reorganization. Also, the frameworks themselves are not yet well supported by tools that can render the provenance information usable by scientists. Data archives and analytics systems will need to work together to provide provenance mechanisms that are usable by Earth scientists in the face of cloud-based data-parallel analytics.

The data management community will face similar challenges in managing data archives for Big Data analytics. The reorganization of data needed for high-speed analytics likely implies that at least two copies of data, the archive version and the analytics-optimized version, need to be managed. The cost of maintaining two or more versions of a given dataset may lead data management organizations to limit the residency time of the analytics-optimized data, which are often on more expensive storage. Another challenge will be to create an analytics user experience working with the rest of the data archive to encourage the user community to shift from a mindset of download-and-analyze to analyze-in-place.

The very existence of several influential communities of practice springing up in the area of Big Earth Data Analytics, such as the Apache Science Data Analytics Platform (https://sdap.apache.org), is likely to have the most significant effect on the future of Big Earth Data Analytics. Open source frameworks abound for doing complex data analysis. The general machine learning community has exploded due to a combination of commercial applications, high-quality accessible online training, and robust, easy to use packages and platforms. This

community overlaps with the Python data analysis community centered around the data processing capabilities of Python along with a wealth of add-on packages for analysis and even many particularly suitable for EO use (e.g., h5py). The geoscience world has developed its own communities, such as Pangeo, which develops and curates a Python-based platform for analyzing geophysics Big Data, particularly in cloud and cluster environments. This is not to say that the future of Big Earth Data Analytics will be based on Python. The nature of computer languages is such that there will likely always be other languages with particular strengths, such as the statistical analysis capabilities of R or the parallelization available from Julia.

With so many communities growing so fast in response to different drivers (other than the desire to analyze more data than can be done currently), a legitimate question is whether the evolution of these communities will continue to enable them to build on each other, or whether they will eventually become stovepiped into non-interoperable communities whose work is difficult to leverage in other communities. There is reason to hope for and expect continued interoperable evolution. Many of the communities emphasize the role of open Application Programming Interfaces (APIs). Also, some standards communities are turning their focus toward supporting analytics. The Open Geospatial Consortium maintains two standards that support analysis of geospatial data, the Web Coverage Processing Service (Baumann, 2016) and the Web Processing Service (Kazakov et al., 2015).

7 Conclusions

Earth Observations are likely destined to continue to grow apace. EOSDIS, for instance will see an order of magnitude increase to over 200 PB of data over the next 6 years with the launch of SWOT and the NASA-Indian Space Research Organization Synthetic Aperture Radar Mission (NISAR). The expansion of Earth Observations into commercial sectors will similarly increase EO data Volume, as well as Velocity and Variety. Fortunately, advancements in the field of analytics algorithms, particularly in the machine learning area, and the increased access to big computing offered by cloud computing offer good prospects for keeping up with the massive data growth. The most challenging areas are likely to be in areas that do not scale well with technological advancements, such as the Variety of data sources and the resultant datasets. In these areas, concerted efforts toward standardization of such aspects as data structures, formats and metadata will be important to take full advantage of the new diversity of data.

References

de Assis, L.F.F.G., de Queiroz, G.R., Ferreira, K.R., Vinhas, L., Llapa, E., Sanchez, A.I., Maus, V. & Câmara, G., (2017). Big data streaming for remote sensing time series analytics using MapReduce. Revista Brasileira de Cartografia, 69.

Baumann, P., (2016). A voyage through dimensions: Recent innovations in geospatial coverages. In Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International, 3599-3601.

Brandt, M., Rasmussen, K., Hiernaux, P., Herrmann, S., Tucker, C.J., Tong, X., Tian, F., Mertz, O., Kergoat, L., Mbow, C. and David, J.L., 2018. Reduction of tree cover in West African woodlands and promotion in semi-arid farmlands. Nature Geoscience, 11, 328, doi:10.1038/s41561-018-0092-x.

Durand, M., Fu, L.L., Lettenmaier, D.P., Alsdorf, D.E., Rodriguez, E. & Esteban-Fernandez, D., (2010). The surface water and ocean topography mission: Observing terrestrial surface water and oceanic submesoscale eddies. Proceedings of the IEEE, 98, 766-779.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D., Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. Science. 353, 790-794, doi:10.1126/science.aaf7894

Jiang, L., Yue, P., Kuhn, W., Zhang, C., Yu, C. and Guo, X., 2018. Advancing interoperability of geospatial data provenance on the web: Gap analysis and strategies. Computers & Geosciences, 117, 21-31.

Johnson, B.A., Iizuka, K., Bragais, M.A., Endo, I. & Magcale-Macandog, D.B. (2017). Employing crowdsourced geographic data and multi-temporal/multi-sensor satellite imagery to monitor land cover change: A case study in an urbanizing region of the Philippines. Computers, Environment and Urban Systems, 64, 184-193.

Kalluri, S., Alcala, C., Carr, J., Griffith, P., Lebair, W., Lindsey, D., Race, R., Wu, X. & Zierk, S., (2018). From Photons to Pixels: Processing Data from the Advanced Baseline Imager. Remote Sensing, 10, 177, doi:10.3390/rs10020177.

Kazakov, E., Terekhov, A., Kapralov, E. and Panidi, E., 2015. WPS-based technology for client-side remote sensing data processing. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences.

Kontgis, C., Schneider, A., & Ozdogan, M. (2015). Mapping rice paddy extent and intensification in the Vietnamese Mekong River Delta with dense time stacks of Landsat data, Remote Sensing of Environment, 169, 255-269, doi: 10.1016/j.rse.2015.08.004.

Lary D.J., Zewdie, G., Liu, X., Wu, D., Levetin, E., Allee, R., Malakar, N., Walker, A., Musse, H., Mannino, A., Aurin, D. (2018). Machine Learning Applications for Earth Observation. In: Mathieu PP., Aubrecht C. (eds) Earth Observation Open Science and Innovation. ISSI Scientific Report Series, vol 15. Springer, Cham. doi:10.1007/978-3-319-65633-5_8.

Liu, Y., Yang, Y., Jing, W. and Yue, X., (2018). Comparison of Different Machine Learning Approaches for Monthly Satellite-Based Soil Moisture Downscaling over Northeast China. Remote Sensing, 10, 31, doi:10.3390/rs10010031.

Moussavi, M.S., Abdalati, W., Scambos, T. and Neuenschwander, A., 2014. Applicability of an automatic surface detection approach to micro-pulse photon-counting lidar altimetry data: implications for canopy height retrieval from future ICESat-2 data. International Journal of Remote Sensing, 35.5263-5279, doi:10.1080/01431161.2014.939780.

Pradhan, R., Aygun, R.S., Maskey, M., Ramachandran, R. and Cecil, D.J., (2018). Tropical Cyclone Intensity Estimation Using a Deep Convolutional Neural Network. IEEE Transactions on Image Processing, 27, 692-702.

Rodriguez-Fernandez, N.J., Richaume, P., Kerr, Y.H., Aires, F., Prigent, C. and Wigneron, J.P., (2017). Global retrieval of soil moisture using neural networks trained with synthetic radiometric data, Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International, 1581-1584.

Schutz, B.E., Zwally, H.J., Shuman, C.A., Hancock, D. and DiMarzio, J.P., (2005). Overview of the ICESat mission. Geophysical Research Letters, 32.

Song, X.P., Potapov, P.V., Krylov, A., King, L., Di Bella, C.M., Hudson, A., Khan, A., Adusei, B., Stehman, S.V. & Hansen, M.C., (2017). National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. Remote sensing of environment, 190, 383-395.

Sundararajan, M., Taly, A. and Yan, Q., 2017. Axiomatic attribution for deep networks. arXiv preprint arXiv:1703.01365.

van der Made, P.A. and Mankar, A.S., Brainchip Inc, 2017. Neural processor based accelerator system and method. U.S. Patent Application 15/218,075.

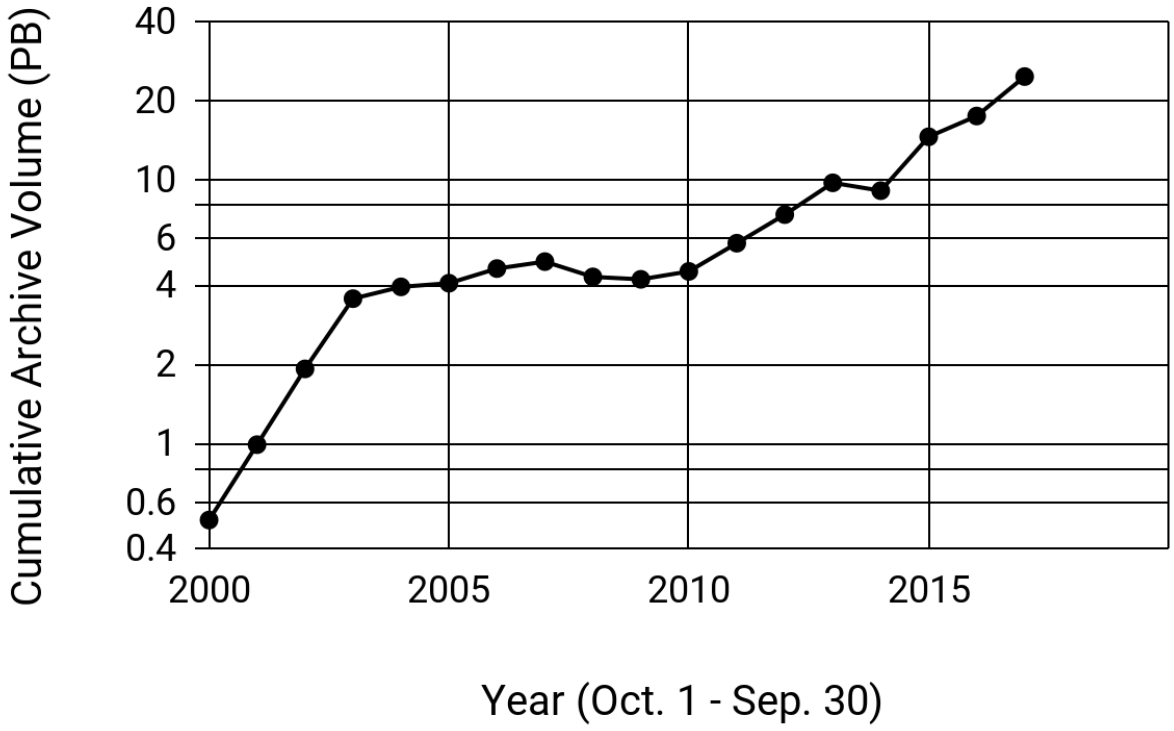

Fig. 1. Growth of Cumulative EOSDIS Archives since 2000, shown on a semi-logarithmic graph. Metrics reporting years begin in October and end in September, so the value for 2017 covers the cumulative archive on 30 September 2017.