Analogies and Comparisons for STM Data Bodies

Phillip M. Cunio (pcunio@exoanalytic.com / 321-412-7863)
Brien R. Flewelling (bflewelling@exoanalytic.com / 979-220-3793)

ExoAnalytic Solutions, Inc., Mission Viejo, CA
February 2019

Abstract
Space Situational Awareness (SSA) has already demonstrated its potential to be extremely data-intensive. The natural extension thereof into Space Traffic Management (STM), necessitated by the increase in volume, variety, and complexity of the space traffic population, will have requirements that go far beyond the space object identification and conjunction avoidance goals of today. One can expect a commensurate increase in data to support these new requirements. The large number of objects on orbit today, if observed constantly, could produce a staggeringly large number of observations that might in turn generate large numbers of orbits as well as space object feature and behavioral information. Orbit data with a lengthy time history can be used to produce estimates of maneuver frequency, susceptibility to natural forces such as drag or solar radiation pressure, and (if combined with photometric data) assessments of behavioral patterns of life. Understanding all these key elements of information will be critical to future STM endeavors of maintaining flight safety, spacecraft anomaly support and resolution, as well as forecasting and forensically analyzing space domain traffic incidents.

Global launch rates are accelerating and will usher in a future with multiple mega-constellations and a growing number of nations and organizations with assets on orbit. This will make it likely that the amount of resident space objects (RSOs) will scale by some substantial factor, with a commensurate increase in collected STM data. While the potential for increase in the number of potential space object conjunctions scales geometrically with the rapid population increase, there are additional factors for effective STM which may additionally compound the need for the ultimate amount of data to support a global STM support and regulatory regime; e.g., many countries are considering their need to create sovereign solutions for SSA and STM.

Examining factors for near-future datasets of a comparable level of richness and size may identify insights for managing the future growth of STM data. STM data will not only become more challenging to utilize as volume growth occurs, but needs for data storage, dissemination, and processing will become steadily more ambitious as the user base for bodies of future STM data grows. This will likely suggest different technology and material solutions to address this area than traditional approaches which are more common to other domains where solutions were predicated on principles of big data from their inception.

This paper seeks to assess domains where such deep data sets are present, including from other data-intensive human endeavors, and map their general contours of development and use, with an eye to deriving lessons and insights for the future of STM.

Introduction

Although the current prevailing regime in orbit features an apparently tranquil status quo, there are still occasional incidents which prefigure possible negative futures, such as the twin failures of geosynchronous communications satellites AMC-9 and Telkom-1 in the summer of 2017.

Maintenance of some approximation of the current state (with no completely-unrecoverable failures and no indications of a rapidly-impending Kessler cascade state) is often assumed as a de facto goal by future space commerce operators and would not be an unreasonable objective for a Space Traffic Management (STM) enterprise. However, the current state exists in an environment incorporating no mega-constellations, with a limited number of inexperienced or low-margin operators of spacecraft active at any time, and with no widely-acknowledged need to maintain full and constant surveillance of all objects on orbit. As these factors change, the challenges facing an STM enterprise will multiply.

Among these challenges will be the management of the volumes of data that must be collected, managed, and utilized in order to support the proper functioning of the overall STM enterprise. The volume of data will scale with the total number of objects on orbit and the rising need to maintain control of all of them, with the corresponding need to maintain appropriate knowledge. This paper makes a simple assessment of the volume of future data that might be required for such an enterprise and compares it to other data-rich enterprises. There is little question that the scale of data involved will be considered a big data problem.

Definition of Space Traffic Management

The International Academy of Astronautics (IAA) Cosmic Study on Space Traffic Management defines Space Traffic Management (STM) as the set of technical and regulatory provisions for promoting safe access into outer space, operations in outer space, and return from outer space to Earth free from physical or radio-frequency interference [1]. Effectively, STM provides supporting services to groups which operate on orbit in a fashion and with a purpose resembling the support Air Traffic Control (ATC) provides to groups that operate in the air.

The current state of the space domain represents a key inflection point, with rapid increase in risk for space traffic incidents concerning both physical and spectral interference. The traffic population is anticipated to grow at rates not before imagined due to the concept of mega-constellations being embraced and the significant reduction in cost to orbit afforded by recent advances in the space launch industry. Considering each member of this traffic population as a node in a complex network, it is easy to consider the monitoring of objects and connections in this network as a big data problem, given their sheer number.

Mega-constellations represent such a significant shift in the nature of the low earth orbit population. The NASA Orbital Debris Program Office 3rd quarterly newsletter for 2018 [2] quotes 19,137 objects being tracked as of 4 July 2018 by the U.S. Space Surveillance network. Approximately 1,232 of these objects are active LEO spacecraft, 558 are active GEO spacecraft, and on the order of 16,000 comprise the full population of LEO objects leaving on the order of 3000 for the complete GEO population according to the Union of Concerned Scientists [3]. Therefore, the fraction of active spacecraft as compared to the total population today, though already accelerating, is on the order of 7.7% of the objects in LEO being active as compared to

18.6% of the GEO population being active as compared to objects tracked by today's basic data and services that are larger than 10 cm. Geosynchronous spacecraft and low Earth orbit spacecraft are flown very differently. Every active spacecraft in GEO can perform its own station keeping with propulsion and ultimately disposing of itself in the graveyard orbit. In contrast, the vast majority of LEO spacecraft do not carry onboard propulsion. This makes the preponderance of LEO space objects essentially just active debris, in that their ability to do anything about an advisory on physical or radio-frequency interference is incredibly limited. This has led to a dichotomy of SSA and STM data support for these two populations. Where deep space traffic has essentially self-organized into a "near real-time flight safety services" model, space traffic in low earth orbit has used the 25-year rule to de-orbit and a more "hazard advisory" approach. These approaches are characterized by significant differences in the volume of data, timeliness of advisories, and the richness of information contained in derived conclusions. The single addition of just one mega-constellation on the order of 5,000 satellites (of which there are multiple planned) suggests that approximately one-third of the future LEO space object population will be active satellites. Using the distribution and behavior of spacecraft operators in deep space as an analogue, we may argue then that the data supporting LEO space operations needs to progress to a near real-time flight safety services model, especially if these mega-constellations plan to use onboard propulsion. Such a change will rapidly bring the suggested Big Data challenges being discussed here.

Definition of Big Data

Big data is an evolving term that describes a large volume of structured, semi-structured, and unstructured data that has the potential to be mined for information and used in machine learning projects and other advanced analytics applications. For a problem domain and its supporting data to be considered a big data problem, it typically must display the 3 V's and 1 C most often associated with big data. That is the volume, velocity, veracity, and complexity of the data must be substantial. In the case of STM, there is much opportunity to collect sufficient volume of external data from proliferated sensor architectures. Enhancing velocity and veracity depends on the modality and collection rates of these architectures, which are also steadily evolving. Additional sources of rich data (often unexplored) include such adjuncts as the onboard telemetry of each active space object. The complexity of a big data solution must also be matched to reflect the complexity of the solution space the data is intended to inform. This leads to a need to be very deliberate in the data curation process.

Centrality of Big Data

The modern world, especially in the last fifteen years, has seen a de facto proliferation of two previously-unexpected items, to the point where both are now commodities. One of these is sensors (especially imaging sensors, with positioning sensors – including both GPS receivers and accelerometers or magnetometers – a close second); the other is internet connectivity infrastructure (with wireless access, especially via internet browsers, perhaps the most common).

The prevalence of sensors and the ability to move data means that humans have an unprecedented capacity to collect quantifiable data on many objects, persons, activities, and events in the world, and the ubiquity of connectivity means that they in turn have the unprecedented capacity to connect and assemble these datasets.

The automated collection of datasets has in turn enabled the use of machine learning, which necessarily relies on datasets that would be unwieldy by traditional standards. Many of these datasets would present challenges to human understanding – simply reviewing every image in a collection of millions of similar images would be extremely challenging for a human, but algorithms often perform the task repeatedly of necessity.

Accordingly, as long as readily-available collections of superhumanly-complicated datasets by similar devices are a feature of the world, it is conceivable that these datasets will play a role in the operation of systems in the world. Some such datasets are already being compiled and utilized; however, it is not likely that the full possible depth of such datasets (meaning the fullest possible meaningful collection of such data) has been reached – wherefore it may be surmised that further growth in the richness of these datasets is very possible.

Notable Big Data Endeavors

In order to provide some context for the potential size of the data bodies that an STM enterprise may encounter, we estimate the size of data bodies that would be encountered in other data-intensive human endeavors. Three areas of human attention generally associated with big data are used as benchmarks: global overhead imagery, human performance monitoring, and information technology network traffic monitoring. These three areas are not uncommonly seen as exemplars for the use of big data. All are also seen as examples of deep data collection, albeit with some additional growth possible.

Rich Deep Datasets

For the purposes of this paper, a deep dataset is any which gathers at or near a level of thoroughness that captures nearly all variation possible in the data. This estimate may be rooted in some assessment of the physical constraints which allow variation that generate data in the first place and is at some level in contrast to a shallow dataset, which only captures a convenient level of data, at some cadence or degree of detail made possible with simple equipment. (Note also that "rich" and "deep" are used interchangeably in some sections of this paper.) As an example, a shallow dataset may easily describe collecting only the necessary data to support catalog maintenance of deep space objects assuming zero maneuvers and no confounding events will occur. While this data easily supports maintenance of orbit estimates for slowly evolving objects doing relatively few maneuvers, this does not comprise a rich and deep dataset which would for example support detailed continuous characterization of more complex motion and potentially inform possible indications and warnings of anomalous behavior.

Although typically a deep dataset is more valuable than a shallow dataset, we do not propose to address that distinction here, and note only that deep datasets are larger, yet generally more desirable. Furthermore, commercial offerings of space object observation and support services are being developed with these deeper goals in mind with sizing of data storage and handling solutions being driven by anticipation of what can be learned by addressing STM as a big data problem. We posit that analysis of a fully-deep dataset, where nearly all the possible variation that might be meaningful is captured along temporal, spatial, and other axes, might be instructive. As such, datasets considered in this paper are analyzed at fully-deep levels, which will be achievable soon and likely augmented by additional technology which provides complimentary information on similar timescales.

When considering transition between a shallow dataset-based solution and a rich and deep dataset-based solution, it is useful to exploit a combination of human subject matter expertise in processing as well as automated processing such that a hybrid architecture may be utilized to effectively maintain the existing capabilities of the former while progressing to enhanced capabilities of the latter. One such approach is discussed here, the "data depth on demand" architecture, which enables the collection of data that is rich and deep at the request of the user.

Global Overhead Imagery

For the purposes of this paper, we imagine global overhead imagery as the collection and regular updating of visible imagery of the Earth's surface with some meaningful resolution. While traditional purposes for overhead imagery include national defense purposes, bodies of overhead imagery data are increasingly of great commercial use and will remain of national civil use for purposes of mapping and resource management for the foreseeable future.

Commercial uses of overhead imagery include personal navigation services, as well as real estate development, remote archaeology, and commodity throughput tracking. Entire commercial entities such as Planet, Spire, and DigitalGlobe exist and compete solely in this growing market.

A roughly canonical estimate of the required information in a 1-cm resolution image of the entire Earth is 1 zettabit, or 10^{21} bits [4]. A more precise and potentially useful estimate can be derived by assuming that the entire Earth need not be imaged. For instance, because 95% of the world's population resides on about 10% of the Earth's land surface [5], we can assume that surface monitoring would focus on some fraction of the world's surface. While this fraction may be dynamic, we estimate it at around 20% of the land surface (the most populous 10% plus an additional 10%), and a comparable 10% of the (overall much larger) Earth's ocean surface. This equates to approximately 66,082,200 square km, or 6.6E13 m^2. Resolution of a meter would meet most needs for mapping, building-level archaeology, large-scale commodity tracking, and many other associated needs (e.g., automobile presence or absence on highways or in parking lots). If all of this area is imaged at a bit depth of 64, then a total of 4.2E15 bits are needed for a full image refresh.

However, in any data enterprise there is also some need for regular updates, and because many of the active material or vehicles encountered under this application do not move more than 100 m/s under many conditions (and many of the buildings do not move at all), we may assess that an update rate of 0.1 Hz would allow tracking these objects across fields of view with relative ease.

Accordingly, an overarching estimate of the fully-deep per diem data body produced by a relevant global overhead imagery enterprise is 3.65E19 bits/day.

Full-Body Human Performance Monitoring

Human bodies are extremely complex functional systems, and yet utterly pivotal for a great many human endeavors (automated device and software operation being perhaps one of the few large-scale systems running without a need for a human body engaged somewhere). Achieving optimal performance has long been an obsession for professional or Olympic-level athletes and capturing and managing relevant performance parameters is of great and growing interest to any human wishing to practice better medicine or surgery, manage or eradicate any disease, or even simply live a more healthful life.

The growing wearable fitness tracker industry testifies to this interest, but only barely indicates its scale. For a full running assessment of human performance, it might be necessary to monitor humans at a cellular level.

Humans have 3E10 cells, and each of these cells may have 1000 or more mitochondria, 100 or more lysosomes, and as many as 1E6 proteins per cubic micron [6]. Even admitting that these are only perhaps the most interesting and active components of a human cell (and nowhere near the full list of elements or organelles), there are very many elements to monitor.

Additionally, human cellular metabolism proceeds at a comparably very high rate, as estimated by assessing the estimated adenosine triphosphate usage rate [7], set here at around 6.8E7 operations per second. If we assume that each of these elements can be captured by a 32-bit data package, then we can arrive at a total estimate of 1.1E34 bits/day for a human per diem monitoring exercise.

A few somewhat simplifying assumptions reduce the data volume. For instance, monitoring all neighboring cells in a cell group (or even in a given muscle fiber) may not provide much valuable information when a single cell in the cluster could be monitored as a stand-in for the whole. We may assess that something like 3000 cellular data-capture points (estimated here as 1000 muscular, 500 connective-tissue, 500 cardiovascular/respiratory, and 1000 endocrine/cognitive collection features) can serve as a first-order assessment, and we may further presume to capture the operations of only 1% of all mitochondria and lysosomes, and 0.1% of all proteins. Finally, we may also note that many perceptible differences in cell performance appear on human time scales, and thus decrement the data collection rate by a factor of 1E5, arriving at a collection rate of approximately 680 Hz, which is well within feasible instrumentation operating parameters as well.

Accordingly, a reasonable assumption for the data collection we may perform to understand a human's full functioning can be set at 1.13E19 bits/day.

<u>Large Information Technology Network Traffic Monitoring</u>

The world is also increasingly straddled and supplied by networks of devices that connect to other devices (or servers) for purposes of process management, financial transactions, personal relationships, etc. The volume of data flowing through these networks represents a long-anticipated and ever-increasing problem.

For instance, if an estimate of 50 billion internet-connected devices by 2020 is accurate, and we assume each generates a packet with an average size of 1500 bits at a rate of around twice per second, then a total of 1.3E19 bits/day represents the per diem data body size.

Roughly supporting this estimate is an estimate of all data that will be generated on the internet in 2020 [8]: 1.7 MB/s per person, or a total of 9.36E21 bit/day [9].

Scale of Data Bodies

Table 1 captures the estimates made in the prior section and adds an assessment of data rates for an STM enterprise. Global overhead imagery and human performance columns both include a prior estimate for a full-scale deep dataset that is somewhat larger than the most practically useful estimate, for a full context.

Note that there appears to be some approximate parity in the order-of-magnitude size for all the three regimes selected. This suggests the possibility that there may be some characteristic size for big data bodies; here estimated as on the order of 1E19 bits/day, or approximately 20 TB/sec.

Table 1. Data demands per diem.

Regime	**Data demand per diem (bits)**	
	Full-depth	Representative
Global Imagery	1.13e26	3.65e19
Human Performance	1.69e34	1.13e19
IT Net		1.30e19
STM	1.6e19	6.6e15

<u>Space Traffic Management</u>

An estimate of space traffic management data needs in the near future would account for the presence of megaconstellations (sometimes also called kiloconstellations) as well as more traditional constellations and debris objects.

If we estimate that five megaconstellations of about 5000 satellites each are flown, and that the number of other active satellites is approximately 5000, we can assess a number of active operating satellites. We may also assess around ten times that many debris objects above 10 cm in diameter, and perhaps a further order of magnitude more debris objects below 10 cm in diameter, for a grand total of 580,000 Resident Space Objects (RSOs) of interest.

If we assume that the minimum set of trackable information to disseminate for RSOs in an STM context includes a state vector, a photometric count, and a small image chip, then we may use estimates of the data size for each of these components (starting with a simple 32x32 image chip and update rate of 1 Hz) to arrive at a total data body size of 6.6E15 bits/day. However, it is also easy to assess the utility of a higher data rate around 250 Hz (to track things like vibrational signature of RSOs) and larger image chips of 100x100 pixels (for better context imaging) and arrive at a value of 1.6E19 bits/day for an STM enterprise. Because STM is not yet as mature or advanced an enterprise as the others considered, the lower value is used as representative in Table 1.

We may therefore assess that an STM enterprise is not as utterly strenuous a big data activity as exists in all of human endeavor, but it may well approach that level in the future.

As a means of benchmarking present-day STM, we may note that the ExoAnalytic Global Telescope Network (EGTN) collects approximately 11 TB of raw imagery per diem. (Note that raw collected imagery must be processed before products such as state vectors and photometric counts may be extracted.) This means that, within a single enterprise, the maximum STM-relevant data handling rate currently achieved is 0.00064% of the characteristically-gigantic figure of 20 TB/sec noted above. However, this still represents a data flow rate of around 127 MB/sec, and it is still growing rapidly. Moreover, we may also note that this present scale of data collection and management by ExoAnalytic alone represents a full 1.34% of the size of the data body that a future STM enterprise might hope to manage in the representative case of 6.6E15 bits/day.

Implications of STM Enterprise Data Scale

The likely future scale of data to be incorporated into an STM enterprise, even though it compares perhaps somewhat favorably to the extremely large scales of data bodies in global overhead imagery, human performance monitoring, and large IT network traffic management, is sufficiently large as to represent a notable forthcoming challenge.

One straightforward way to manage this challenge is to utilize a data-depth-on-demand architecture. This type of architecture focuses on collecting a full-sized dataset at a distributed network of points of collection, but then extracts the data of most interest at that point (via edge processing) and sends it back to a central server or observer immediately, making larger and most detailed data sets available only on a slower basis, and utilizing prior knowledge of what these larger data sets might be, and on what cadence they might be desired.

Figure 1 illustrates this concept as instantiated by the EGTN and compares it to a notional flow-through architecture. Note that there are multiple levels of data products, and in fact the data at one level is often a product of the prior level. Additionally, the data volume forwarded greatly decreases between levels. While users preferentially (and rapidly) view data from a given level, they may on demand access deeper levels. However, there is no assumption (as in the baseline flow-through architecture) that every single datum need be placed in front of a human at the maximum possible speed, and some data will never attract human attention.

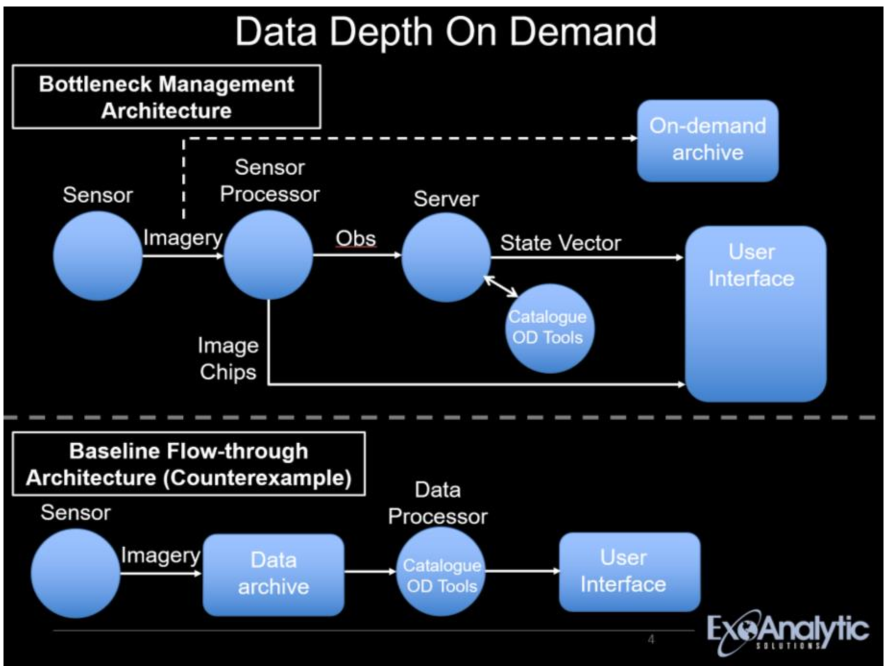

Figure 1. Example of a data depth on demand architecture.

The future path of an STM enterprise can perhaps be imagined via comparison to the current state of air traffic control. Although the global ATC enterprise is comparatively mature, its scale is not quite equal to the future scale that may be expected from an STM enterprise, due in part to the fact that aircraft tend to return to the surface within hours after taking off, while many RSOs remain on orbit for decades or centuries.

For instance, the largest number of aircraft in flight at one time is approximately 19,000 [10]; the current number of active and inactive RSOs on orbit is approximately 5000 [11], not counting various fragments and debris objects, including some launch vehicles.

The overall ATC problem is thus of a scale roughly comparable to STM, with the substantial caveat that nearly every single aircraft is, if not occupied by an active human pilot, in the active custody or under the attention of a human, far fewer spacecraft at a time can make this same claim. Additionally, due to the presence of tens of thousands of debris objects on orbit, none of which can maneuver under their own power, a much higher percentage of potential collisions in space are avoidable only by one of the objects concerned, rather than both (as with most air traffic).

Additionally, and critically, ATC is conducted under an interlinked but independent network of distributed sites operated by a coalition of third parties unconnected to the vehicle operators, with full legal and regulatory ability to compel changes in air traffic patterns. None of these characteristics are presently applicable to STM.

Recommendations

The sheer volume of data that may be produced by a near-future global STM enterprise augurs against immediately implementing a single centralized space operations command center for each nation, institution, or organization requiring STM knowledge. If a full-knowledge STM data stream requires constant access to and utilization of a 20 TB/s data channel, creating just one such operations center is likely to generate a significant strain on local data-transfer infrastructure, if it is even possible. As of the time of writing, speeds of this class are around two orders of magnitude slower than those attested in even advanced laboratory settings. Creating multiple such centers would only exacerbate the challenge.

Shipping a crate of 30-TB hard drives [12] (assumed size 5’ on a side) results in around 240 PB of data making transit; even if this requires a full day to arrive at its destination, it represents a data transfer rate of about 2.7 TB/s.

As such, transferring all of this data on a constant basis would be an infeasible requirement. As such, an architecture affording data depth on demand, with immediate transmission of extracted critical information at much higher speeds and later transmission of full data volume for spot loci only at times and places of need (e.g., for full incident forensic analyses) is a much more reasonable approach.

Accordingly, one of the first goals of research supporting such an STM enterprise should be the development of a specific scheme for this kind of architecture.

Summary and Conclusion

A brief assessment of the possible growth of bodies on orbit in the near future was provided in this paper, and the total volume of data collection that a management enterprise focused on such a regime might require was estimated, using as comparison bases three other noted data-rich fields of endeavor. Although neither STM nor any of these fields are noticeably close to a state wherein they have reached full depth of data collection, an interesting possible convergent factor was noted in that a full-depth data collection for each of these enterprises could be said to exist at a per diem data collection of approximately 1E21 bits per day, or a data collection rate of about 20 TB/s.

This data collection rate would most likely represent an infrastructural challenge for any central STM data and operations control center. As such, the paper recommends a collection architecture based on the concept of data depth on demand, wherein immediate processing is applied to the data at the point of collection, and only certain pre-identified subsets of data are sent to a point for further analysis. Human-supported intervention or identification of non-standard cases can then be used to retrieve already-collected fully deep data sets on an as-identified basis. ExoAnalytic's Global Telescope Network operates on such an architecture and reduces data from the as-collected scale to approximately three orders of magnitude smaller [13].

References

[1] "IAA Cosmic Study on Space Traffic Management." International Academy of Astronautics, 2006. URL: https://iaaweb.org/iaa/Studies/spacetraffic.pdf.

[2] "Orbital Debris Quarterly News" National Aeronautics and Space Administration, 2018. URL: https://orbitaldebris.jsc.nasa.gov/quarterly-news/pdfs/odqnv22i3.pdf.

[3] "UCS Satellite Database" Union of Concerned Scientists, 2018. URL: https://www.ucsusa.org/nuclear-weapons/space-weapons/satellite-database#.XE9pPlxKiUk

[4] "Data topography for pervasive, proliferated space situational awareness." Cunio, P. M., Corbin, B., and Flewelling, B. R. AMOS 2018.

[5] "Urbanization: 95% of the World's Population Lives on 10% of the Land." URL: https://www.sciencedaily.com/releases/2008/12/081217192745.htm.

[6] Milo, R. "What is the total number of protein molecules per cell volume? A call to rethink some published values." Bioessays 35: 1050-1055, 2013. DOI: 10.1002/bies.201300066.URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3910158/.

[7] "How many ATPs are produced per second by the average human?" URL: https://www.quora.com/How-many-ATPs-are-produced-per-second-by-the-average-human.

[8] "Data Never Sleeps 6.0." URL: https://www.domo.com/learn/data-never-sleeps-6.

[9] "Projections of population growth of World (2020)." Population of the world website. URL: https://www.livepopulation.com/population-projections/world-2020.html.

[10] Satherley, D. "Record number of planes in the air at once." Newshub, 02/017/2018. URL: https://www.newshub.co.nz/home/travel/2018/07/record-number-of-planes-in-the-air-at-once.html.

[11] "How many satellites are orbiting the Earth in 2018?" Pixalytics web site, 22 August 2018. URL: https://www.pixalytics.com/sats-orbiting-the-earth-2018/.

[12] "Samsung Electronics Begins Mass Production of Industry's Largest Capacity SSD for Next-Generation Enterprise Systems." 20 February 2018. URL: https://news.samsung.com/global/samsung-electronics-begins-mass-production-of-industrys-largest-capacity-ssd-30-72tb-for-next-generation-enterprise-systems.

[13] Cunio, P. M., Therien, B., Ingram, C., and Hendrix, D. Presentation at SSA Data Science Workshop, 24-25 October 2018, Austin, TX.