Augmenting traditional networks with data buys can support science, as well as operations

Jennifer Gannon[1] and Noé Lugaz[2]

[1]CPI
[2]University of New Hampshire

November 21, 2022

Abstract

Science is fueled by data. Throughout history, scientists have operated sensors-from astronomical observatories to particle accelerators-that accumulate observations for analysis or to evaluate a hypothesis. However, as available technologies have increased both the volume of data and the efficiency of data storage and transmission, a new model of data access has emerged. The concept of a data buy is where an entity purchases access to a set of data or a data stream, instead of operating the sensors themselves. But why might a data consumer, whether a researcher or an end-user, prefer this kind of data access over the more traditional methods of running a network themselves? The simple answer, in some cases, is efficiency and, possibly, cost. Space weather forecasting and analysis has a growing private sector, and the extension to data gathering can be considered as a natural next step in the maturation of the field and the growing public-private partnerships. Operational applications require consistent, clean, and (in some cases) real-time data access that can be hard to support through the existing model of sensor deployment. Even in scientific applications, where access to raw information can be critical to discovery, there are benefits to the data buy model. Consistent access to a trusted data set allows more time to be spent on the scientific analysis, instead of maintaining machines that require consistent development, maintenance, and monitoring. The outsourcing of data infrastructure and pipelines can be particularly beneficial when the sensors are in distributed networks, spread over wide areas, and when there is a need to provide local data in observational gaps in existing networks. In the ideal case, a data buy can supplement the traditional observational networks in a beneficial and symbiotic way. It is important to note that data buys should not replace traditional observational networks, nor compete for funding with future observatories and infrastructure that the scientific community has deemed necessary (for example, through decadal survey processes).

Augmenting traditional networks with data buys can support science, as well as operations

J.L. Gannon, N. Lugaz

Science is fueled by data. Throughout history, scientists have operated sensors - from astronomical observatories to particle accelerators -- that accumulate observations for analysis or to evaluate a hypothesis. However, as available technologies have increased both the volume of data and the efficiency of data storage and transmission, a new model of data access has emerged. The concept of a data buy is where an entity purchases access to a set of data or a data stream, instead of operating the sensors themselves.

But why might a data consumer, whether a researcher or an end-user, prefer this kind of data access over the more traditional methods of running a network themselves? The simple answer, in some cases, is efficiency and, possibly, cost. Space weather forecasting and analysis has a growing private sector, and the extension to data gathering can be considered as a natural next step in the maturation of the field and the growing public-private partnerships.

Operational applications require consistent, clean, and (in some cases) real-time data access that can be hard to support through the existing model of sensor deployment. Even in scientific applications, where access to raw information can be critical to discovery, there are benefits to the data buy model. Consistent access to a trusted data set allows more time to be spent on the scientific analysis, instead of maintaining machines that require consistent development, maintenance, and monitoring. The outsourcing of data infrastructure and pipelines can be particularly beneficial when the sensors are in distributed networks, spread over wide areas, and when there is a need to provide local data in observational gaps in existing networks. In the ideal case, a data buy can supplement the traditional observational networks in a beneficial and symbiotic way. It is important to note that data buys should not replace traditional observational networks, nor compete for funding with future observatories and infrastructure that the scientific community has deemed necessary (for example, through decadal survey processes).

Enacted in October 2020, the "Promoting Research and Observations of Space Weather to Improve the Forecasting of Tomorrow" (PROSWIFT) Act (see Space Weather, doi: 10.1029/2020SW002628), includes a data buy provision called the "Pilot program for obtaining commercial sector space weather data". This pilot program aims to evaluate ''the viability of assimilating the commercially provided data into National Oceanic and Atmospheric Administration space weather research and forecasting models", "whether, and by how much, the data so provided add value to space weather forecasts of the National Oceanic and Atmospheric Administration and the Department of Defense; and "the accuracy, quality, timeliness, validity, reliability, usability, information technology security, and cost-effectiveness of obtaining commercial space weather data from commercial sector providers." This pilot program has been modeled after a similar terrestrial weather data pilot that started in 2017. Under this program, two companies – Spire and GeoOptics – have been providing streams of radio occultation data to NOAA for assessment. GeoOptics continues to provide these important data under an on-going data buy. The success of that program (see for example Bowler, 2020, doi.org/10.1002/qj.3872) supports expanding the data buy model for space weather. The PROSWIFT act includes examples of space weather data that includes "space radars, lidars, magnetometers, neutron monitors, radio receivers, aurora and airglow imagers, spectrometers, interferometers, and solar observatories."

Under the PROSWIFT data pilot, a significant emphasis is placed on data quality, standards, and metrics. The on-going evaluation of data quality is important for both data consumers and providers, and should be used as an objective assessment of data quality, to build trust between data providers and data consumers. This may be especially important to give reassurance to the scientists used to working on their own data. In addition to supporting a solid data infrastructure, the use of modern – but stable – technologies such as machine learning or similar advanced statistical techniques can support these quality goals, as well.

Finally, if the data procured from a data buy is to advance science, it must be publicly available. Debates in the weather community on how best to achieve this goal have occurred over the past 40 years (see for example Cirac-Claveras, 2019, doi.org/10.1016/j.spacepol.2018.08.002) . *Space Weather* requires a data availability statement for any publication. The underlying data required to understand and evaluate the research needs to be available to the peer reviewers as well as the readers. Data calibration and instrument description should also be published, as is commonly the case for "traditional" observatories and missions. In order to maintain both private company viability and data availability , a model where data is made publicly available for research purposes when it has not "real-time"/forecasting value may be considered. However, for an operationally-focused data buy, like that outlined in the PROSWIFT data pilot, it is also important to the success of the program that critical data streams be maintained without interruption. As the model for data buys evolves, new solutions for public access to data archives that do not compromise operational integrity will need to be developed. With these pieces and caveats in place, data buys will be mutually beneficial to operations, research, and data providers.