GEOSPATIAL DATA AS A SERVICE:

A VITAL INVESTMENT IN AMERICAN ENTERPRISE

A Report of the National Geospatial Advisory Committee

September 2018

Executive Summary

In 2017, the National Geospatial Advisory Committee (NGAC) received guidance from the Federal Geographic Data Committee (FGDC) to provide forward-looking advice and guidance to maximize the value and utility of data-as-a-service (DaaS) as a delivery mechanism for Federal data repositories (e.g., GeoPlatform, Data.gov, etc.). A subcommittee of the NGAC has prepared this report to examine and summarize the opportunities and challenges of key importance to federal agencies using DaaS, or considering the new opportunities that DaaS may provide, to support the growing spatial data demands from across public and private sectors. In the report, several recommendations are presented to inform federal decision making, to maximize federal investment in DaaS to support and broaden the geospatial community, and to enhance public-private partnerships for expanded innovation across all sectors. The subcommittee believes that the recommendations present opportunity for deeper exploration on the topics of greatest interest to federal DaaS stakeholders.

Introduction

As modern society becomes increasingly driven by information technology, expanding access to federal geospatial data investments has tremendous potential for American enterprise. This new information and innovation economy is about putting data to work, supporting smarter and faster decision-making. To meet the growing appetite for data, many government agencies and industries are employing "big data" analytics that are supplied by data-as-a-service (DaaS) infrastructures. These DaaS infrastructures are enabling initiatives across all public and private sectors, such as smart cities, intelligent transportation, precision medicine, and on-demand service delivery. DaaS also compliment, or can be considered a component, in supporting the development of spatial data infrastructure.

Given the trend toward DaaS, what are the implications to the geospatial market? The government's geospatial community has a history of making its data assets broadly available through a blend of public sector capabilities such as the Federal Geographic Data Committee's (FGDC) Geospatial Platform and National Oceanic and Atmospheric Administration's (NOAA) Big Data project, as well as private sector collaborations such as Google Maps, Esri's ArcGIS Online, OpenStreetMap, and cloud offerings from Amazon, Google, and Microsoft that provide infrastructure as a service for hosting large repositories of government data.

This paper examines current and future needs for DaaS from the perspective of federal data repositories that provide online data access and discusses the benefits and opportunities for maximizing these data investments.

Benefits provided by DaaS

The DaaS model provides numerous benefits over traditional on-site data storage and access or models which require downloading to work with data locally, including:

- **Minimization of data silos**. DaaS allows data to be available to users from a common location, with one point of contact to manage, maintain, and update any given data source.
- **Reduced management and delivery costs.** Centralized, common access points for data management and delivery to reduce storage needs, minimize geographic constraints on data access, and enhance ability to maintain authoritativeness of data (versus countless duplicates of a dataset on local drives with altered geographies or attributes).
- **Easier to manage, secure, and maintain.** Centralized DaaS access and management is more cost effective, secure, and easier to manage than the in-house IT services and administration needed to manage and respond to customer requests for individual datasets.
- **Support for scientific research and innovation.** DaaS makes it more feasible to work with large and/or real-time datasets available as inputs in multiple applications. Providing the massive volumes of data collected available in an online system supports efforts for large-scale modeling, data mining and machine learning, to improve and augment human decision making.
- **Supports Federal Data Strategy.** DaaS offer a coordinated and integrated approach for data delivery and discovery in support of the Federal Data Strategy principles. DaaS can maximize the benefit and value of government data through creation of federated marketplaces that directly fuel applications for development of derivative geospatial applications and services. These new services are anticipated to accelerate geospatial decision making and knowledge discovery.

Current State of DaaS

DaaS is a data distribution model in which data files are made available on demand over an internal or external network. This differs from the traditional model for distributing spatial data where access to data was restricted to downloads to the local machine or access via physical media (e.g., DVD, external hard drive, etc.). DaaS facilitates machine to machine interactions, where the data do not live on the local user machine; instead, the user accesses data directly from the remote machine. While it may be possible for a user to download all or part of a dataset from a service, not all data services provide this functionality.

DaaS opens new possibilities and innovation by improving access to and use of data. Users and producers across government, industry, open source, education, and the sciences are turning to DaaS to meet their huge appetite for information, exponential storage needs, data archiving and to minimize data duplication and reduce costs. To illustrate these points, we examined four case studies that provide more in-depth use cases from across Federal and local agencies and the private sector. These case studies showcase real world scenarios with large, distributed data sources to support a broad range of end user needs. This is not intended to be an exhaustive list of examples of DaaS in use. There are many other relevant examples and programs of DaaS across the stakeholder community. The full case study documents are available from the NGAC.

To demonstrate use of DaaS for local governments and emergency response, we highlight the work of the Missouri Task Force One (MO-TF1) and Boone County Fire Protection District (BCFPD). MO-TF1 and

BCFPD are using public cloud for basemap creation, data access, and delivery to enable a common operating picture of critical data during deployments. (Appendix 1 - Data as a Service & Cloud Computing: Disaster Response Just-in-Time Basemap Creation for Deployments).

To demonstrate the applicability of DaaS for serving a broad range of state-level geospatial data, we highlight the Montana State Library's use of Cloud based GIS DaaS web map services to complement their traditional, downloadable data and growing collection of web applications (Appendix 2 - Montana Spatial Data Infrastructure).

To demonstrate applications of DaaS for large volume data, we highlight the NOAA Big Data Project. NOAA has been experimenting with the Big Data Project to improve accessibility and usability of environmental data. This project currently exceeds 20 terabytes of data captured per day (from a variety of sources, including satellites and weather stations) and is expected to exceed 80 petabytes of archived data per year by 2021 (Appendix 3 - NOAA Big Data Project). To further highlight the extreme size of many geospatial datasets, NASA now produces 7-20 petabytes a year per project.

To demonstrate commercial use of DaaS, we highlight the private sector 'Living Atlas' platform provided by Esri (Appendix 4 - Living Atlas of the World – The Data You Need) that serves data to a variety of customers and applications – many of which are embedded into mobile devices such as smartphones and tablets.

Future of DaaS

Data are of no value unless they contribute to understanding, decisions, and actions. The future of DaaS is not simply to serve data faster and at lower costs but to enable answers-as-a-service, driving data to decisions and actions by providing new opportunities for enhancing data accessibility and utility (making more data more accessible in more places). DaaS is evolving into an ecosystem that provides a framework for efficiently integrating data from multiple sources and opening new possibilities and innovation. This growing DaaS marketplace will continue transforming beyond data supply and processing toward data analytics, applications tools (APPs) and APIs that can be customized for targeted decision-making and action-taking. This next generation of DaaS will also facilitate rapid elicitation of useful insights from data to improve situational understanding, make informed decisions, and take proper actions.

The expanding information and innovation economy will rely more and more on geospatial-enabled DaaS offerings to contextualize and customize answers to user queries. For example; a vehicle driver may ask an APP for the shortest route on a trip. This DaaS-enabled APP of the future will likely contextualize the uses request and provide the shortest route and safest route based on traffic accidents and social media postings on a stormy day. In another example; a user might be interested in walkability options for their community and query an APP to understand the possibilities. This future DaaS-enabled APP will likely produce a walkability map with community health, environmental factors, and other local insights. In the walkability example, the DaaS-enabled APP shows the user how walkability correlates to community health and what factors may relate to their positive and negative relationships. Therefore, the user can prioritize areas and take actions to enhance positive walkability factors and improve community health.

People need data to make decisions, and the future of DaaS-enabled APP development, moving into the future, will provide more actionable data and insights for decision making. The next generation of DaaS will also continue to support and grow the innovation economy with APPs that leverage open and easily accessible data and be supported by geospatial information. Machine learning and artificial intelligence will inform the future of DaaS in its evolution toward answers-as-a-service. Artificial intelligence (AI) incorporates knowledge representation and can reason, plan and adapt. Machine learning accelerates the processing power of DaaS to discover information faster. Incorporation of AI and machine learning into DaaS will be a promising development as it advances toward more geospatial-enabled problem solving and support for decision making and action taking.

The Opportunity

Recent estimates indicate that the geospatial industry is generating significant economic value to the US economy and projected for continued high growth both domestically and globally into the near future. A 2012 study by the Boston Consulting Group (BCG) estimated that the geospatial industry was providing jobs for over half a million people and was driving $1.6 trillion in revenues throughout the global economy. A more recent 2018 study by Geospatial Media shows similar productivity for the geospatial industry: a sustained high growth projected at 13.6% through 2020, current 2018 economic output for the US economy at $300B, and projected US economic output to grow to $500B by 2020. The public sector represents only a fraction of this growing market. While public sector market share continues to decline over time and the federal government is no longer the dominant producer of geospatial information, this point simply demonstrates the healthy and vibrant state of the geospatial market. Government agencies still provide valuable geospatial data that supports a variety of consumer applications and services (across the public and commercial sectors) and that drives innovation in the growing information economy. Maximizing the value and benefits of these government data assets has been an ongoing challenge. Often government may lack the funding, technical resources, and mix of talent for putting these data to work; while other innovators such as businesses, citizens, academia and others may have the talent and resources necessary to innovate but lack access to the government's data assets. So, what are the implications of the growing commercial market in DaaS and geospatial data to the government? What should be the government's role in this new DaaS and geospatial marketplace? What are the opportunities for strengthening public sector geospatial data assets and the benefits to society?

Public-private partnerships (P3s) provide the government a mechanism to leverage the unique capabilities from both the public and private sectors that are flexible and intended to produce benefits to all parties in the collaboration. P3s achieve this by distributing risk and sharing resources with opportunity for non-government parties to develop new products and services. These partnerships are good for the government in expanding the talent pool, adding scarce resources, increasing access to and improving distribution of valuable government data, and promoting new scientific discoveries and operational efficiencies. Non-government partners benefit from improved access to government data, scientists, learning through collaboration, opportunities to support public interests, and projects that can spur new revenue and markets for their enterprise.

The most common P3 arrangements are cooperative research and development agreements (CRADA), cooperative agreements, and joint ventures. The P3 model has been a successful instrument for bringing private capital and expertise to public services and has expanded from traditional infrastructure projects (i.e., highways, bridges, dams, etc.) to government big data and other services. NOAA is using the CRADA

model for its Big Data project to make their scientific data more accessible by leveraging cloud infrastructure and compute processing from the private sector. The U.S. Census Bureau is using cooperative agreements to create federal statistical research data centers that provide non-government researchers secure access to otherwise restricted government data. NASA has used joint ventures numerous times with both for-profit and not-for-profit partners. These have included joint ventures with SpaceX for new space launch capabilities and the German government to improve infrared astronomy.

P3s offer the geospatial community a real opportunity to maximize its value and benefits to the public good and the commercial economy. They provide government an opportunity to overcome the challenges of what it can afford and expertise it has to deliver by leveraging the deep, rich technical capabilities and capital from non-governmental partners.

Challenges Introduced by DaaS

While DaaS present many benefits, there are also significant challenges to be addressed. DaaS introduces novel challenges, while still being plagued by many of same long-standing challenges that the online geospatial data services industry face.

Novel challenges

- **Accessibility of data in offline situations.** Though the data service itself is not likely to be impacted by power or connectivity losses due to the replication of data across geographically dispersed servers, the lack of connectivity by an end user is a challenge that must be addressed. This is particularly important for disaster response and recovery situations. There may need to be niche products and services available in the market to support disconnected use of online data that easy synch with resumption of service.

- **Legality and licensing.** Since many geospatial information data sets are licensed, a user must comply with the licensing terms of each of the data sets that are used, even if the data is licensed under an Open Data license. A license is a legal document. While most people see a license agreement as a transfer of intellectual property rights, it also serves to allocate the risk between the parties. For example, who is responsible if the data quality is insufficient, if the data is used for a purpose for which it is not suited, or if the data was collected, or used, in violation of laws? As a result, the terms of a license will vary. Since many lawyers are not familiar with geospatial information or how it is used, and the legal and policy framework is confusing and unclear, a lawyer is more likely to say no than yes. Even if they want to say yes, when faced with a long list of licenses that they need to review, things tend to get bogged down.

Long-standing challenges

- **Findability and discoverability of content.** Users rarely know all the content available nor what other content may be available to support their needs. Findability and discoverability affect two distinct end user needs: finding specific information and discovering new information. With the explosion in geospatial data, emergence of data intermediaries, and pivot toward DaaS, it will be important to ensure that end users are getting to the right information at the right time for right

needs. This will likely involve continued investment in knowledge management initiatives across the different market sectors: government, commercial, scientific, and open data community.

- **Access, data provenance, versioning.** Many applications or analysis pipelines that rely on DaaS are continuous processes, and not one-off projects. Support for long-term analytic processes requires retention and access to historical data; datasets must be continuously available and provide dynamic metadata to track version updates. Many models require specific data versions (e.g., current weather forecast or 1990 Census boundaries). Developers and users need to be able to easily set up automated data access based on what version they need to utilize, and have available information on provenance to document the historical record of the data. Additionally, it is critical to consider the impact of the format obsolescence that often occurs with advances in geospatial data collection and distribution.

- **Format and interoperability standardization.** There are multiple formats and exchanges standards for distributing geospatial DaaS (e.g., ESRI Rest services, WMS, WFS, etc.), with the likelihood of additional formats to be developed. DaaS providers and end users need to be able to work with any number of formats and must translate these to multiple exchange standards depending on the intended use.

- **Trustworthiness of data and applicability of use.** Questions of trustworthiness and applicability of data are longstanding issues faced by the digital geospatial data community and are not unique to the issue of DaaS. With DaaS, these issues may be heightened due to the increasing proliferation of geospatial data from a wide variety of sources, for instance, from professionals in the geospatial community to John Doe on the street with a smartphone. This variety of potential sources and data distributors leads to questions about trustworthiness of data and applicability for any given use.

  Authoritative data may not be the best available data for the intended use and purpose. Users need to be able to 'look under the hood' to understand the structure of the data, attributes, format, scale, lineage, etc. Perhaps of greater importance than the authoritativeness issue is that of data provenance. This need is especially acute in regards to data intermediaries that serve to fuse multiple sources of information into common DaaS access points or serve to function as data distributors repackaging government data into their platform services.

  Accessibility alone should not raise a red flag when evaluating DaaS - if a user readily knows the source of the data and has the ability to determine if the data comes from a reliable source and is applicable for his or her purposes. Any DaaS system should provide users with ready access to the necessary metadata to evaluate the applicability of the data before making use of data services. One concern is that metadata may not adequately capture data provenance of the DaaS.

  Ensuring that sufficient metadata is provided is imperative in the growing DaaS market. Identifying information is mandatory in the content standard for digital geospatial metadata and should also be mandatory for DaaS offerings. It may also be necessary to examine and define or redefine common terms and concepts that may be applied to datasets such as authoritative, certified, trustworthy, and appropriateness for use to achieve a common lexicon.

Path Forward: Recommendations to the Federal Geospatial Community

The recommendations discussed in the path forward are intended to inform Federal decisions on maximizing the impact of DaaS to support and broaden the geospatial community. Realizing the benefits of DaaS and capitalizing on the opportunity will require federal commitment to the following steps:

1. Renewing leadership (public, private, academic, non-governmental, etc.) to drive success and facilitate action across all stakeholders;

2. Improving access and availability of government geospatial data through DaaS implementation mandates;

3. Emphasizing interoperability and common exchange standards that promote competition and spur innovation across all sectors (public and private);

4. Leveraging shared-capital through public-private partnerships to expand use and innovation in government geospatial data; and

5. Displaying leadership with Federal geospatial in transforming the Federal Data Strategy. For example, merging existing online Federal geospatial data repositories with Data.gov.

This paper was prepared by a subcommittee of the NGAC that included the following members: David Alexander (chair), Sarah Battersby (vice-chair), Stu Davis, Kevin Pomfret, Rebecca Somers, Jennie Stapp, Harvey Thorleifson, Jason Warzinik, and May Yuan.