

**POLICY BRIEF:**  
**BIG DATA AND  
THE SUSTAINABLE  
DEVELOPMENT GOALS:**

**Innovations and Partnerships to Support  
National Monitoring and Reporting**

**TR****ENDS**  
Thematic Research Network  
on Data and Statistics



Partners  
for  
Review

**JULY  
2021**

The Sustainable Development Goals (SDGs) present an unprecedented monitoring challenge for governments and National Statistical Offices (NSOs). As a result of increasing demands for data from users, declining budgets, and rising data-collection costs, interest has grown in harnessing data from new partners in the national “data ecosystem.” The potential for “big data” to support SDG monitoring has incited considerable enthusiasm with many emerging experiences and use cases which underscore the need for increased collaboration and partnership. However, there is poor information-sharing on how partnerships can support national SDG monitoring.

This paper provides guidance for countries aspiring to leverage partnerships that harness big data for national SDG monitoring. We have drawn together experiences from countries that have used big data partnerships as well as insights from the latest research collaborations that are deriving innovative datasets from big data to support SDG monitoring.

This policy brief was prepared by Dr Cameron Allen, Grant Cameron and Hayden Dahmm from the United Nations Sustainable Development Solutions Network’s Thematic Research Network on Data and Statistics’ (SDSN TReNDS) workstream on data governance, with financial support and oversight from Partners for Review (P4R). P4R was initiated in 2016 on behalf of the German Federal Ministry for Economic Cooperation and Development and the Federal Ministry for the Environment, Nature Conservation and Nuclear Safety and is implemented by GIZ. P4R’s objective is to help foster robust follow-up and review mechanisms at all levels and to ultimately contribute to achieving the SDGs.

## The Big Data Value Proposition for SDG Monitoring

---

Big data describes large volumes of high velocity, complex, and variable data. This covers commercial, sensor, mobile, and online data from searches and social media. Within national statistical systems (NSS), big data sources have a strong value proposition, including increased coverage and granularity, cost reductions, improved timeliness, and reduced respondent burden. The associated opportunities, challenges, and risks will depend on the type of big data, the legal and regulatory context, and the institutional capacity settings.

## Big Data Use Cases for SDG Monitoring

---

There have been major developments over the past decade in how countries use big data for statistical production and decision-making.

Mature big data applications include web scraping or scanner data for price statistics, social media data for consumer confidence indexes, mobile data for mobility and tourism statistics, and satellite data for agricultural statistics. There are also recent case studies of countries experimenting with big data sources to support national SDG monitoring, including in the Philippines, Colombia, and Ghana, among others.

For less mature big data applications, NSOs must look to others to complement their skills and data to produce statistics such as researchers and academia, international organizations, geoscience and space agencies, the private

sector, and civil society. In particular, the research community has pioneered new uses of big data sources to support SDG monitoring. Our paper identifies and reviews 100 recently published datasets derived from big data, highlighting considerable coverage of the SDGs corresponding to 15 goals, 51 targets, and 69 official indicators. Satellite or Earth Observation (EO) data was the dominant big data type, and publicly available satellite imagery combined with cloud computing infrastructure has largely driven advancements in SDG applications.

Newfound datasets derived from big data can serve SDG monitoring by filling data gaps as well as producing more timely and disaggregated estimates. Often, NSOs must rely on global datasets, generated by other organizations, to produce national-level statistics. To assure quality outputs, NSOs must bridge the gap between themselves and the global collective of “big data for SDGs” innovators through new forms of formal collaboration. Some examples of good collaborations involve global platforms that enable countries to produce their own national datasets through the provision of analysis-ready big data, open access methodologies, and training. Greater use of these scalable applications and tools is a steppingstone to expand the number of countries taking advantage of research and new global datasets derived from big data to improve national-level SDG reporting.

However, while providing open-source global datasets is a critical factor to improve SDG monitoring, issues often remain around interpretation, verification, understanding, and ultimately the use of these datasets by national stakeholders. To gain the trust of national stakeholders, global datasets must be developed alongside national capabilities and information systems to create the best statistics possible. And from a technical perspective, building awareness and transferring new skills and capabilities from global experts to NSOs will be critical to mobilizing EO data and other big data sources for SDG monitoring.

## Partnerships for Harnessing Big Data for SDG Monitoring

Interviews with NSO officials indicate that integrating big data sources to support SDG monitoring have upended the traditional models for producing statistics and confirm that NSOs must work with new partners to tap non-traditional data sources, learn and adopt new techniques, and utilize new platforms.

### Understanding the Partnership Landscape

Partners can engage with the NSO in big data projects through a range of different roles and working modalities. In addition to collaborating with big data providers, NSOs may need to partner with technical service providers or technology partners to gain access to capabilities not available in-house. Knowledge brokers or convening organizations might also be needed as partners to raise awareness of big data’s potential and build trust in new methods and processes. Additional funding is often needed to get big data projects off the ground, which can benefit from an executive sponsor / partner to champion a business case for internal funding allocations, or an external donor willing to provide funding. Finally, a legal or data privacy expert to help navigate the regulatory landscape is another important partner that is often overlooked.

While considerable opportunities exist for partners to support NSOs and governments with their national SDG reporting obligations, NSO capability and experience gaps can undermine effective partnerships and slow the production of new statistics. These gaps include a lack of awareness of big data’s potential, difficulties in initiating big data projects, legal and privacy issues to access big data, and gaps in technical capability. These gaps can vary from one NSO to the next and dictate the type of partnership that must be formed.

## A Big Data Maturity and Partnerships Model for NSOs

Some NSOs have yet to commence their big data journey, while others are quite advanced. A formal maturity model provides an intuitive framework to understand the current state and progress of NSOs in terms of deriving value from big data (Figure 1). This is structured along three dimensions: i) Organizational characteristics: big data awareness, literacy, and support for big data use cases; ii) Technical capabilities: both human and infrastructure; and iii) Governance: the legal and regulatory context. Big data maturity across these dimensions is categorized in one of five stages: nascent, pre-adoption, early implementation, proficient, and mature. As NSOs move through these stages, they require different partners and partnership models to gain greater value from their investments.

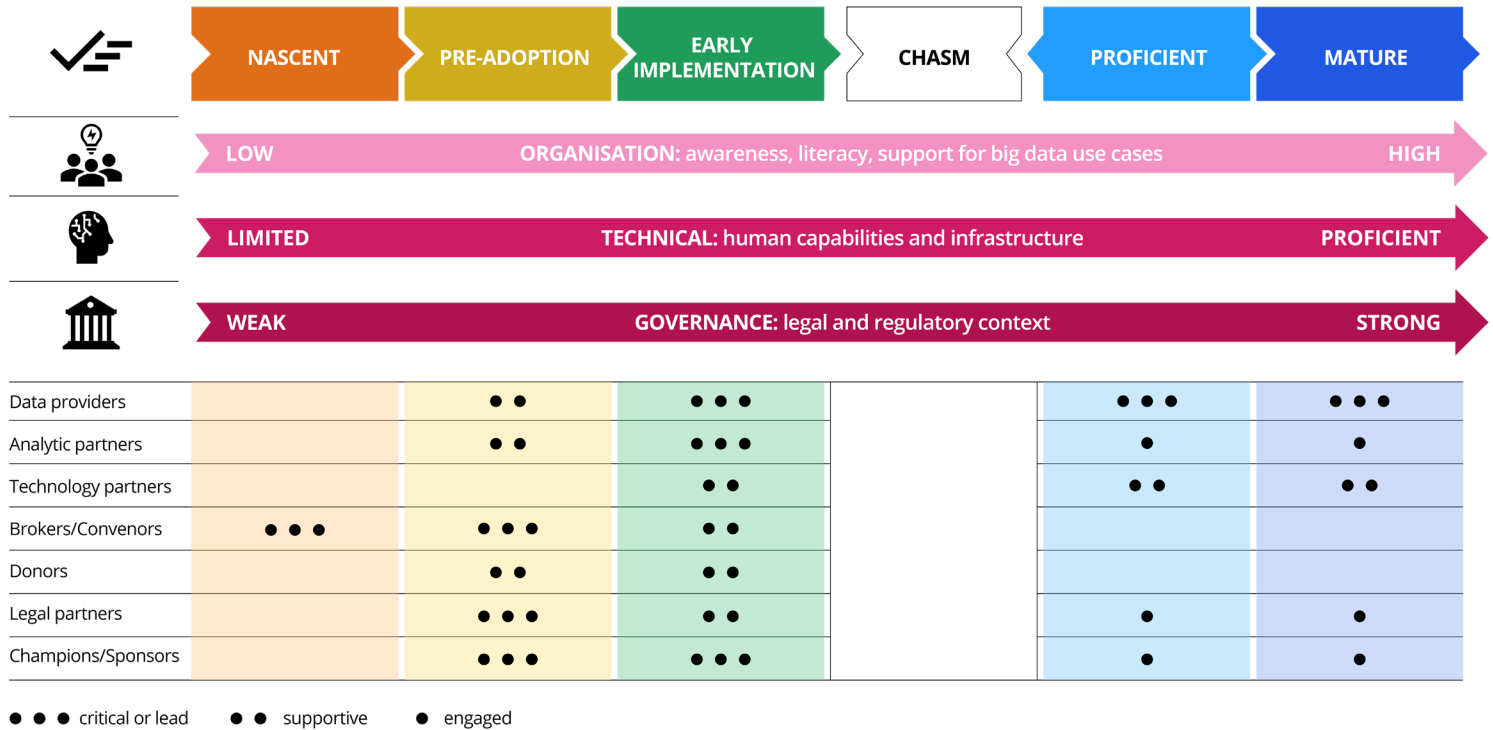


Figure 1 A Big Data Maturity and Partnerships Model for NSOs<sup>1</sup>

<sup>1</sup> Adapted from: 73.Marcovecchio, I., et al., Capability Maturity Models as a Means to Standardize Sustainable Development Goals Indicators Data Production. Journal of ICT Standardization, 2018. 6(3): p. 217-244, 75.Halper, F. and K. Krishnan, TDWI Big Data Maturity Model Guide. 2013, The Data Warehousing Institute (TDWI).

The **Nascent Stage** represents a pre-big data environment. In this stage, the NSO and wider NSS have a low awareness of big data and the potential value for official statistics and SDG monitoring, but convenors and brokers can play a key role in raising awareness levels.

In the **pre-adoption stage**, the NSO is starting to explore big data analytics, and the organization may also be developing a business case for big data applications, with buy-in from an executive sponsor. However, broad-based support from leadership is still lacking. Sources of financing for big data projects may also be explored, including from international donors or government allocations. In this stage, the organization is beginning its big data journey, and the mindset is around experimentation.

The **Early Implementation Stage** is typically characterized by a few pilots that become more established. There is generally at least one executive sponsor involved, however broader interest is likely growing as pilots deliver successful outcomes. This may lead to the establishment of a team to start to plan and strategize for a wider big data scope. During this stage, the organization will be collecting big data from data providers in different formats and may be investing in infrastructure to take advantage of the growing data volumes or outsourcing this to technology partners. Data access and sharing agreements will be in place where needed to enable data sharing while safe-guarding privacy. The organization will also be building data analytics capabilities, usually starting with single data sources and more mature use cases and methods. Technical and analytical partners may be providing services to complement and help build in-house capacity. Measures are also being taken to support the recruitment and retainment of human resources for in-house data science capabilities.

As organizations move from early adoption to the more mature production stages, they need to overcome a **“chasm.”** While pilot applications have been deployed, gaps may remain in terms of capability development and the governance context, including securing ongoing funding. The incentives and business models may not be adequate to move beyond pilot projects, and a big data management and governance strategy or roadmap may be needed. Another key challenge for NSOs relates to building and maintaining capable technical staff, in-house data science capabilities and infrastructure. Moving to mature stages requires that big data projects become a planned and budgeted initiative treated on par with other statistical production and data integration programs.

Reaching the **Proficient and Mature** stages means an organization has well-established, ongoing big data programs that are executed as budgeted. Planned initiatives and funding is secured. During these stages, innovation in data and data analysis becomes a core value of the organization, and it is making use of many forms of big data and other technologies. Data analytics may be operationalized and automated as part of the business process, and governance arrangements will be in place such as a steering committee to oversee progress and a center of excellence to provide training.

# Practical Guidance for NSOs on Partnership Processes

This maturity model highlights areas where NSOs and their partners can target joint efforts, including developing big data awareness and literacy, human capabilities and infrastructure, and governance and regulatory frameworks. We propose a series of practical steps that NSOs can take to develop their big data maturity and identify the partners who can help along the journey.

**Getting big data partnerships off the ground:** for countries and NSOs that are yet to commence their big data journey, there is a general lack of big data literacy. Engaging global and regional statistical conferences and partnering with brokers or convening organizations will be critical at this stage. Initial steps could include convening a national dialogue on big data which invites experts from across sectors to explore real world use cases, and finding a big data champion with influence in government to help raise awareness.

**Building a business case and aligning incentives:** moving through the pre-adoption to early implementation stage requires the elaboration of a sound business case for piloting big data applications for national statistics, generally starting with one or two promising priority applications. A range of partners is often necessary in developing the business case, including NSOs, data providers, regulators, technical service providers, privacy officers, legal counsel etc. The business case should clearly articulate the technical project requirements, costs, and benefits and create a win-win situation to encourage all partners to continue to commit resources to the project.

**Navigating the governance and regulatory landscape:** privacy and security risks may arise in big data projects with the transfer, storage, and sharing of data between stakeholders, and the potential for the unauthorized disclosure of personal data. To help navigate the regulatory landscape and privacy requirements, early engagement should be undertaken with data privacy agencies, regulators, and officers, as well as legal counsel units within the NSO or government. Building trust with these partners is critical during the pre-adoption and early implementation stages and can lead to the development of standardized data sharing agreements or MoUs for different big data partnership arrangements.

**Building the technical and human capabilities:** The early implementation stage provides an opportunity to begin building the necessary human capabilities and infrastructure for big data applications. Analytical and technology partners are critical at this stage to provide technical support services, training, and advice to help complement and grow in-house capacity. Establishing relationships with national geoscience or mapping agencies, space agencies, academic and research institutes, and domain experts are essential for developing national capabilities. Furthermore, given the specialized nature of datasets and skillsets, NSOs will need to partner with multidisciplinary teams comprising subject matter experts, data scientists, statisticians, computer scientists, among others to harness big data sources.

**Sustaining big data partnerships:** scaling up from pilot activities to production and sustaining these activities over the long-term will likely require NSOs to make significant investments in human capital and skills development, such as appropriately skilled data scientists to collect, process, and analyze big data sources, and to integrate new data-collection and production methods into their work programs. This will take time and require ongoing capacity-building, training, methodological advancements, and financing. These efforts will also need to be backed by ongoing big data programs that are planned and budgeted with secure funding. Data providers, analytical partners, technology partners, and donors will therefore be critical in supporting NSOs to cross the chasm towards big data proficiency.

**TRENDS**  
Thematic Research Network  
on Data and Statistics



Partners  
for  
Review