Organisation for Economic Co-operation and Development

SDD/CSSP(2018)7

Unclassified

English - Or. English

20 August 2018

STATISTICS AND DATA DIRECTORATE
COMMITTEE ON STATISTICS AND STATISTICAL POLICY

Cancels & replaces the same document of 30 May 2018

Which Strategies for NSOs in the Digital Era? Towards 'Smart Data' Strategies

15th meeting of the Committee on Statistics and Statistical Policy (CSSP)
20-21 June 2018
Palais des Nations, Geneva

Following the 15th Meeting of the Committee on Statistics and Statistical Policy on the 20-21 June in Geneva, a request was made to make this document available to a broader public by written procedure. The CSSP agreed to declassify this document on 31 July 2018. This declassified document therefore cancels and replaces the original document of 30 May 2018.

The increasing popularity of the internet and social media has generated an explosion in new data, coupled with increasing computing capacity to exploit them. These data have provided significant scope to generate new and improved timely and granular evidence for analyses and statistics, but have also generated new demands challenges. NSOs need to adapt to the digital era, seizing opportunities and re-define their strategies in a new data ecosystem. This note elaborates on the issues and presents elements for a conceptual framework, based on the approach followed by the OECD when developing the 'OECD Smart Data Framework'.

This note is submitted to delegates FOR DISCUSSION under item 5 of the Agenda.

Mr. Eric Anvar, Head of Smart Data Practices and Solutions Division, Statistics and Data Directorate (eric.anvar@oecd.org)

JT03435041

Note by the Secretariat

The increasing popularity of the internet and social media has generated an explosion in new data, coupled with increasing computing capacity to exploit them. These data have provided significant scope to generate new and improved timely and granular evidence for analyses and statistics that can better inform policy makers and citizens. NSOs need to adapt to the digital era, seizing opportunities and re-define their strategies in a new data ecosystem.

The new environment, new demands and the associated challenges are complex, moving and interrelated. This calls for explicit strategies and strategic frameworks to relate the various pieces to each other, to sequence resource deployment and to communicate on the way ahead. This note elaborates on the issues and presents elements for a conceptual framework, based on the approach followed by the OECD when developing the 'OECD Smart Data Framework'. CSSP delegates are invited to **PROVIDE COMMENTS** on the note, and in particular to address the following questions:

- How do NSOs characterise opportunities and challenges they face in relation with the new digital environment - do they have broad strategies and if so which are their main components?
- How do NSOs implement strategies and achieve the necessary culture change in their business model?
- How do they see 'statistical ecosystems' evolve at national and international levels? And what are the areas for cooperation in these 'ecosystems'?
- Do Delegates have any particular feedback on the Smart Data Framework described in the document?

1. NSOs in the Digital Era – the Issues

1.1. Background

1. The increasing popularity of the internet and social media has generated an explosion in new data, coupled with increasing computing capacity to exploit them. These data have provided significant scope to generate new and improved timely and granular evidence for analyses and statistics that can better inform policy makers and citizens.

2. These new sources of information complement, and sometimes compete with official data sources, but also open up new streams of statistics in a range of areas where the collection of robust data was previously prohibitive. NSOs need to adapt to the digital era, seizing opportunities and re-define their strategies in a new data ecosystem. Agility and responsiveness are called for while remaining committed to independence and quality.

3. The international statistical community has actively discussed the challenges brought about by the digital era, including the emerging new institutional roles for NSOs in 'ecosystems of data'. Fora for these deliberations include the UNSC, the UN Global Working Group on Big Data for Official Statistics, ESSNet Big Data project, the OECD and the CES. Significant steps have also been taken to co-operate and share resources internationally, in particular through the High Level Group for the Modernisation of Official Statistics, the SDMX Initiative or the Statistical Information Systems Collaboration Community (SIS-CC) around the OECD's *.Stat Suite*. Each of these and other initiatives cover certain institutional, technical or data-related aspect of producing official statistics. In parallel, discussions have taken place at the national level and there is increasing recognition that NSOs need to actively rethink themselves and their business model. Squeezing new approaches into old structures will not be enough – new partnerships, competencies and strategies are needed.

4. This note elaborates on the issues and presents elements for a conceptual framework. Part 1 of the note asks what kind of strategic frameworks are needed to prepare and accompany culture change and transformation in NSOs. Part 2 describes a Smart Data Framework, based on the OECD's own approach. While there are of course significant differences between the objectives and workings of an international organisation and an NSO, there are also many communalities, in particular when it comes to developing a strategy for the digital era and framing the necessary change.

1.2. The New Demand

5. With the digitalisation of our societies, and people's ability to obtain real-time information (reliable or otherwise) on virtually every aspect of life, expectations have also evolved regarding statistics. At least four characteristics mark this new demand:

> **(i) Minimum 'time-to-market'** is underlying much of the demand: the readiness to accept delays in the provision of official statistics has declined. Policy-makers and citizens want to tune in in real time with the pulse of society and are more and more reluctant to accept statistics that, like astronomers, visualise phenomena that took place light-years ago. People expect quick analyses of policies and how they impact their lives. The acceleration of information supply in a digitalised world carries over to statistics.

(ii) Granularity: beyond timeliness, there is also an expectation that data are granular, for example locally relevant ('how is my community doing?'), that they distinguish different socio-economic groups and that they cover a broad array of social, economic and environmental subject matters. Many private 'grass root' initiatives that produce indicators for local and everyday use speak to this aspect of new demand.

(iii) A quest for 'trusted quality': confronted with an ever increasing (and overwhelming) data supply, there is an expectation from the part of users that NSOs make sense of the noise associated with 'big data', cut through sometimes contradictory information and add value and context to the orphan data that may be dropped in social networks, the media or other new sources. Adding value means differentiating knowledge from data by bringing the statistical expertise (structure, rich metadata, documented and proven methodologies and analysis); ethical standards (transparency and respect of data privacy); findability and openness (machine readable data, available codes) – all of which contribute to build 'trust', a precious and rare asset in the post-truth era of fake news.

(iv) Expectations have also risen in regards to **a wider range of new data services and products**, to which experts (such as academia, NGOs, businesses) or perhaps citizens at large can contribute. These new services may cut across traditional boundaries for NSOs, such as nowcasting data, engaging in modelling work, or sharing algorithmic knowledge and code. Engagement with different types of data users with various profiles and capacities is also required for dissemination. Innovation in data products and services requires NSOs to establish structures, cultures and ecosystems where useful ideas can be experimented and nurtured to become projects and services.

1.3. Challenges

6. Established NSO cultures, structures and approaches towards producing statistics may not be well-adapted to respond to the new demand. In traditional structures, a vertical value chain is often being operated by one entity, within one organisation, which designs and operates the process to collect data, and transform it into a final product for dissemination. Such vertically produced data is characterised primarily by its uniqueness (there is only one census in the country) and its core value proposition is trusted quality. But this responds only to one of the facets of the new demand outlined above and NSOs are challenged to transform their data cycle by tapping into multiple new data sources, adopting new data science techniques, acquiring new platforms, making algorithms available to users – in short, moving from a vertical value chain towards an integrated value chain and step up the readiness to experiment with new sources, methods and partners.

7. NSO 'production lines' are under strain to transcend their historic model to overcome gaps and respond to the new demand. This may not be possible with the current business model because several gaps have to be overcome:

- *Data Sourcing Gap*: while NSOs continue to collect primary data as part of their core value proposition (Census, LFS, enterprise surveys, etc.) the capacity to combine these data with other sources, or create new products based on new data sources, becomes a critical part of the value proposition. Statistical organisations have very much focused on opening the data they produce; they need now to also

focus on becoming users of data available out there, produced by third parties. If this is the case, access to new data becomes a strategic question that needs to be formulated through a global *data sourcing strategy*. This may be a new activity to many statistical organisations, and calls for a fundamental shift in the NSO's culture – from pure data producer to data producer-user. It also entails forging new partnerships, in particular with the private sector.

- *Data Platforms Gap*: seizing these opportunities will require investing in new platforms to build 'data lakes' and to provide data science services. More and more, platforms will be hosted remotely ('on the cloud'), which requires adapting financial models for IT (from capital expenditure to operating expenditure models) and enforcing privacy policies. Emerging AI techniques will reshape our work; many existing tasks will disappear, and jobs will be reshaped in fundamental ways. Blockchain could possibly redefine the way trust in data is being built and there will be other developments that are yet hard to anticipate. All this exposes existing data value chains to accelerated obsolescence and requires a strategic approach towards conceiving and integrating data platforms.
- *Data Skills Gap*: The shortage of data science expertise is a widely known challenge. Statistical operators, the work of whom should be ultimately largely assisted or automated thanks to AI-enabled procedures, should become more of data scientists, picking up tasks in activities such as data sourcing, data modelling, data exploration, data analytics. This gap calls for a *data skills strategy* and a review of existing structures.

8. Given the level of investment required to overcome simultaneously these three gaps, statistical organisations will have to rethink how they are structured, what is their business model and how to leverage broader ecosystems. Arguably, the same question applies to national statistical 'ecosystems' (e.g., division of labour between NSOs and other entities), and to international ecosystems and the international organisations acting therein (e.g., joint investments, partnerships and regulatory efforts).

1.4. Frameworks and Strategies

9. The new environment, new demands and the associated challenges are complex, moving and interrelated. This calls for explicit strategies and strategic frameworks to relate the various pieces to each other, to sequence resource deployment and to communicate on the way ahead. A number of questions arise here:

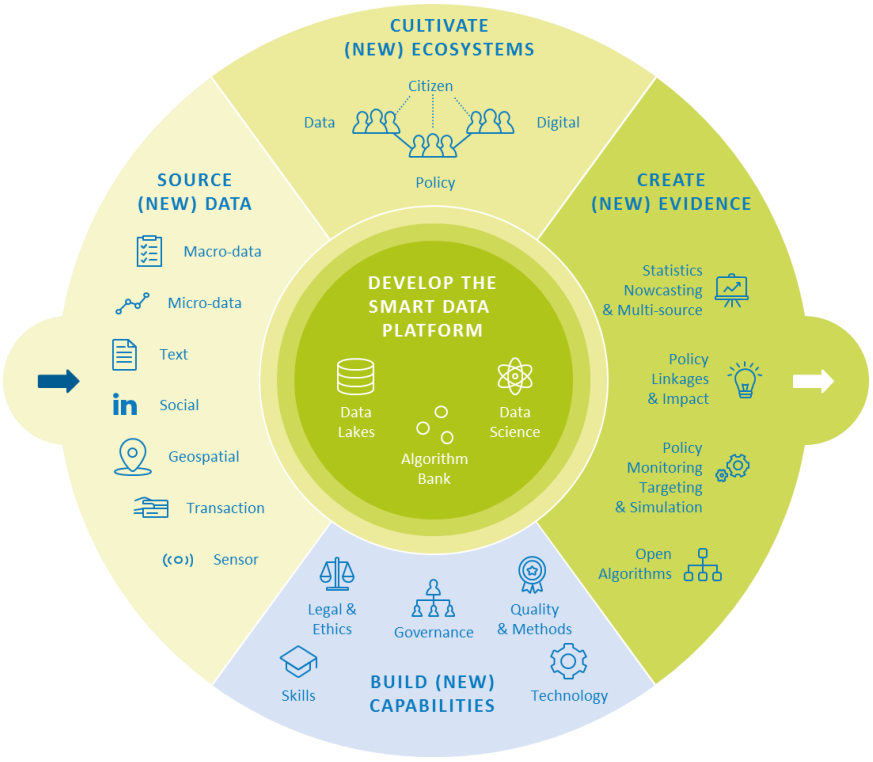
- How do NSOs characterise opportunities and challenges they face in relation with the new digital environment – do they have strategies and if so which are their main components?
- How do NSOs implement strategies and achieve the necessary change in their culture and business model?
- How do they see 'statistical ecosystems' evolve at national and international levels in a way that supports the implementation of their strategies?

10. The next section presents the approach followed by the OECD to address these questions. While this approach aims at responding to the specificities of statistical activities at the OECD, a number of its elements may be wide-ranging and also apply to the situation of NSOs and other organisations.

2. A Smart Data Framework

11. Over the past few years, analysts and statisticians at the OECD have increasingly made use of non-conventional data sources in their work such as geospatial data, social media information or data scraped from websites. About 100 relevant projects were analysed. At the same time, new demands as outlined in Section 1 and technical developments made it necessary to reflect on how traditional data are collected, processed and disseminated. The OECD has come up with the concept of 'Smart Data Framework' (Figure 1), as a way to structure strategy in this field. This section uses the structure and basic approach of the OECD Framework but couches them in the context of NSOs.

Figure 1: Smart Data Framework



The Smart Data Framework comprises 4 axes: (1) Create (new) evidence; (2) Source (new) data; (3) Develop the smart data platform; and (4) Cultivate (new) ecosystems. In a value-driven approach, the (new) capabilities to build (skills, legal & ethics, governance, quality & method, technology) are key success factors to achieve objectives set under these 4 axes.

12. Why 'Smart Data' rather than 'Big Data'? As observed in the projects analysed, value is not necessarily about the data being 'big'; rather, value results from taking creative angles at data, sometimes combining existing (small and big) data sources in new ways, as well as tapping into un-conventional sources and forging new methods and algorithms. Value also arises from combining the more traditional statistical approaches with the newest data science techniques – rather than opposing them. 'Smart Data', from that perspective, can be seen as a semantic advocacy for a value proposition that combines and transcends both, bringing together the statistician and the data scientist.

13. We call 'Smart Data Framework' the interlinked areas organisations need to invest in to explore and realise expected outcomes. It is a representation of the value chain and broader ecosystem that we need to create, in order to bring together approaches that today tend to be more local to a project or team and of limited impact, and would gain from being mainstreamed across all organisation's domains, or even the broader ecosystem. This should accelerate the pace of adoption of emerging practices, and create opportunities for cross-fertilisation and horizontality around data within and beyond each organisation. Four areas of work are constitutive of this Framework:

i. *Create (new) Evidence* for policy: policy innovation should be the driver. Too often, big data strategies in organisations take either an opportunistic angle ('I have these data, what can I do with them?') or opt for an approach driven by constrained capabilities ('how to resolve privacy issues?' is a critical question, but it should not be the driver). The perspective taken by the OECD is that the approach should be value driven: which are the policy problems we want to solve, which cannot be solved with current data sources and practices? Which are the new products that could help resolve policy issues?

ii. *Source (new) Data*: Access to data is the fundamental obstacle to overcome, especially when the majority of new data are with private sector, and their use is subject to privacy issues. This access can be direct ('pull the data') or indirect ('push the code to the data'). Statistical organisations being primarily data producers may have difficulty in seeing data sourcing as a strategic topic: however, in the future the capacity to use external data sources will probably be more important than the data we produce ourselves. Design and execution of a global data sourcing strategy, leveraging various business models, and forging new partnerships in the area of data sourcing, are therefore the cornerstones of a smart data strategy.

iii. *Develop the Smart Data Platform* for statisticians, data scientists and analysts to work with the data more efficiently and creatively, including aptitude to easily access the data, explore them, share them, combine them, including by augmenting regular statistics. The platform consists in data core(s), where regular statistical data are managed, data lakes for any other data, and data science services, leveraging high performance algorithm banks. The platform is connected to a network of partner platforms, with a view to sharing algorithms and data, and facilitating open research and innovation.

iv. *Cultivate (new) Ecosystems* of multidisciplinary (from various policy, digital and data domains of expertise) communities of practitioners, who share common interests in similar data sources or techniques (for example, environment, agriculture, urban specialists working on the same set of geospatial data and creating a shared layer of horizontal metadata and knowledge attached to the same spatial unit). Ecosystems can be developed at different levels: within an organisation, with other institutions and partners at national level, and as part of international, sectorial or regional networks. Ecosystems could ultimately connect citizens as well, in as much they are offered to become active in sharing their data through various digital channels.

14. Working on these levers requires, and in return contributes, to (v) *Build (new) Smart Data Capabilities* in the areas of skills, legal and ethics, quality and methodology, technology, governance – and overcome challenges attached to each of these capabilities.

Legal and ethical challenges are of particular significance as experienced with the privacy issue attached to administrative and other micro-data over the past decade.

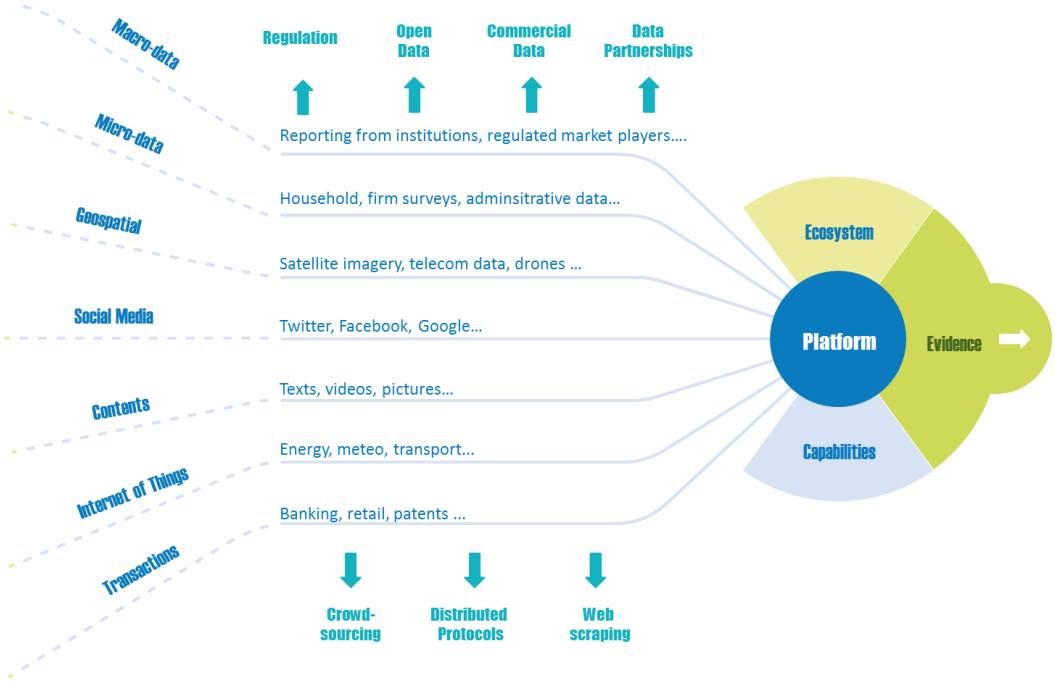
2.1. Which Business Models for Data Sourcing?

15. Historically, statistical organisations have built their products in vertically organised, fully controlled lines of production based on surveys (households, firms) and reporting from other institutions (for example, NSOs collecting data from line ministries, agencies and local governments; or central banks collecting data from financial institutions). For years, NSOs have strived to (a) increase efficiency of the reporting mechanisms, and (b) gain access to administrative sources to complement these data and possibly reduce the burden, especially on businesses.

16. The 'data deluge' has resulted in the massive emergence of new sources of data, especially (a) with the advent of large digital players serving hundreds of millions of customers (GAFAs) and, more recently, the generalisation of the 'Internet of Things'; and (b) through disruptive developments in domains where the production of data is a well-established activity but has undergone technological breakthroughs in the quality and quantity of data produced and/or the capacity to interpret it with new data science techniques (e.g. high resolution satellite imagery and pattern recognition techniques).

17. Access to data, is of vital importance, but becomes more and more complex as new sources and players multiply (Figure 2).

Figure 2: Data Sourcing



Sourcing data of multiple types, from public as well as private sector sources, and according to multiple business models, becomes a core business for statistical organisations in a fast evolving data market.

18. Statistical organisations need to acquire skills in order to harness the complexity of this market, the dynamics of it and the multiple business models to leverage:

- *Regulation*: the idea here is that, as part of regulatory arrangements (such as the provision of spectrum for mobile phone operators), market players have an obligation to report or give access to certain data and contribute data for public good. Including statistical objectives into regulatory arrangements could also create efficiencies elsewhere, e.g. through the collection of data on business transactions from a single intermediary (e,g, Uber, Air BnB, Amazon) rather than the actual service providers (e.g., households), or through banks and credit card information for measuring household consumption or cross-border trade. Could additional targeted regulation empower statistical organisations in better and faster reaping the benefits of the 'data deluge'?
- *Open data*: 'Public, but also private players share data for free, for the sake of transparency, and in order to stimulate innovation around data'. Where the open data movement concerns public organisations (local governments, national administrations) for which data are only by-products, NSOs can bring their expertise in 'trusted quality data' to further enhance the open data ecosystem to deliver more relevant and better structured data for policy.
- *Commercial data*: 'Monetisation of data is massively expanding and changing the market of data brokerage… By 2020, major consolidation will occur in the data broker market as the leading players with large scale and the ability to invest in value adding solutions/platforms and deal with increasing regulation will acquire smaller players unable to keep up.' (Gartner, 2018). Data brokers of multiple shapes and sizes should become privileged partners of statistical organisations. There is an opportunity for organisations to join forces at national and international levels to create advantageous negotiation positions, especially in areas where competition is limited, and perhaps identify opportunities for data partnerships.
- *Data partnership*: 'Public-private partnerships (PPPs) offer significant opportunities such as cost effectiveness, timeliness, granularity and scope for new indicators. There remain, however, a number of challenges, which need to be surmounted, such as technical difficulties, risks related to data confidentiality and a lack of incentives. A number of collaborative projects have already emerged and can be classified into four ideal types, namely: (a) the in-house production of statistics by the data provider; (b) the transfer of private data sets to the end user; (c) the transfer of private data sets to a trusted third party for processing and/or analysis; and (d) the outsourcing of national statistical office functions.' (Paris21, 2015).
- *Crowdsourcing*: 'Share your data to improve policy and make your life better' – crowdsourcing can target citizens at large or specific business or expert communities. The wide digitalisation of a variety of tasks via mobile devices is an opportunity to reduce the fixed cost for statistical organisations to get more data directly from target populations. Key is the incentive to participate in the crowdsourcing process, for example by combining it with a task the target population is willing to execute online anyway (e.g. in the context of e-government services). Key challenges relate to privacy and representativeness of samples. A much more systematic approach to crowdsourcing could be a

powerful way of working around the problem of access to data and develop a notion of 'data citizenship' in the population, whereby 'donation of one's data' would become an act of citizenship.

- *Distributed protocols*: 'Push your code to the data source instead of pulling the data to your code'. For a number of years, protocols have been established to give access to confidential micro-data by selected profiles of researchers at data source's premise; agreements are reached to grant data to third parties under stringent legal and technical constraints. NSOs could leverage the experience gained in this field to develop more systematically the practice of 'distributed protocol': provided data models are sufficiently documented, and relevant samples of data and algorithms are shared, data scientists could develop algorithms to be run at the source, without necessitating physical access to data. This approach is being practiced not only for administrative sources, but also for telecom and satellite data. A standardisation effort may be needed here to overcome difficulties such as IT security risks, data source sustainability over time, and fixed cost of running algorithms at the source with sufficient computing performance. NSOs could spearhead standard setting in this field, and pave the way for a market of 'semi-open' data that could extend to private sources as well. The advent of distributed protocols could also, perhaps, facilitate partnership models or not too intrusive data-for-public-good regulations.
- *Web scraping*: 'Crawl and scrape the data from public websites and APIs'. Scraping data and content from the web via intelligent search agents, including semantically enabled ones, has constituted a market since the inception of the internet. There have been many experiments to complement traditional data sources with web-based ones (e.g. real estate adds, job postings, ledgers, tweets for sentiment analysis, etc.), but success has been limited, due to the inherent instability of the sources and legal limitations. The advent of a new, AI-enabled, generation of scraping engines could represent a breakthrough, requiring however significant investments in technology, data curation and tuning. Data brokers are investing in this field, with which statistical organisations could partner (vs. directly investing in such capabilities).

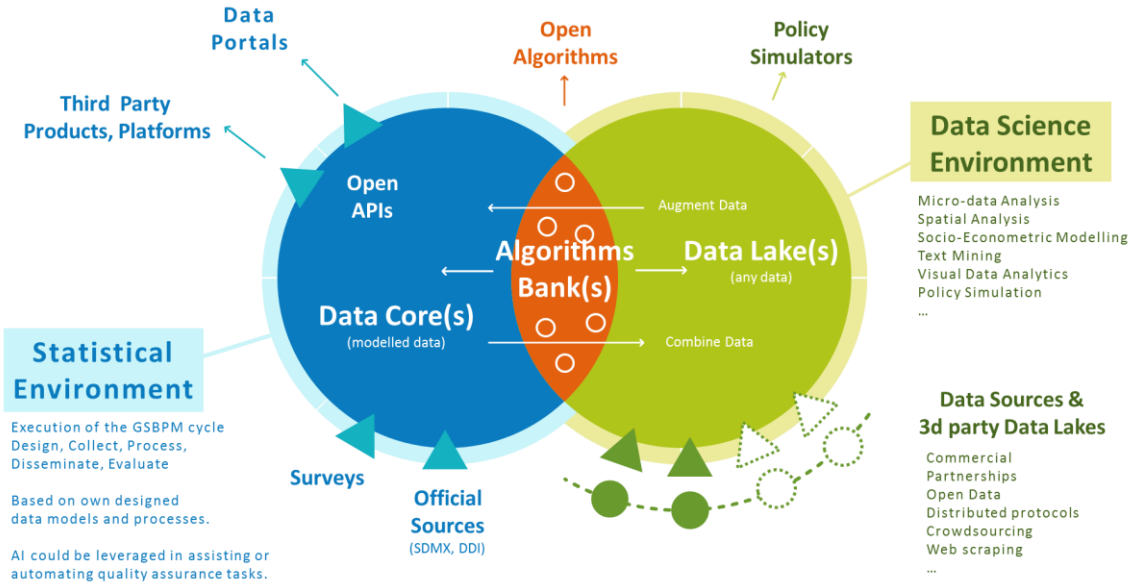
19. The above list of business models in the area of data sourcing is certainly incomplete; their full description and analysis from an NSO perspective would require a substantial research effort, which is necessary to inform coordination efforts within national ecosystems and international ones. These coordinated efforts could deliver values such as:

1. Raise attention and focus on the 'data for policy' dimension in the national open data strategies;
2. Develop standards in the area of distributed protocols and crowdsourcing, including via joint investment in national and international programs;
3. Establish a collective negotiation mechanism with some large digital players and commercial data providers, through joint data procurement and global partnership approach.

2.2. Next Generation Platform: AI-enabled Statistics

20. The mastery of platforms and technology is of constant strategic concern for statistical organisations since decades. Quality assurance framework are crystallised in IT solutions and procedures; well performing computation and storage capabilities are critical for the work of statisticians and data workers; standardisation efforts (such as SDMX) ultimately translate into open source software assets. This is all the more true in the era of big data. The shift of focus from statistics only, to statistics + data, calls for a shift in the global platform architecture supporting each organisation's data cycle (Figure 3).

Figure 3. Platform Architecture



The next generation of official statistics platforms should combine the highly structured data from statistical environments (data cores) with the loosely structured data sourced from the outside (data lakes). High performant algorithm banks should serve both: AI-enable (and, perhaps, blockchain-enable) the statistical cycle (assistance or automation of quality assurance tasks) and support data science exploration.

21. Principles guiding the design of next generation platforms include:

- *Statistical environments to evolve towards AI-enabled statistics*: Statistical organisations have heavily invested in common standards (GSBPM, SDMX, DDI) to streamline data models and processes, as well as the data and metadata exchange and reporting mechanisms through open APIs. These efforts should not be discontinued to fully reap expected benefits, in achieving fully integrated National Statistics Systems[1] and international reporting mechanisms. More efforts also are envisioned in developing quality assurance by design in systems (VTL). AI-enabled (and, perhaps, blockchain-enabled) statistics holds the promise of pursuing this path, by automating lower-level tasks in the statistical cycle, and assisting the statistical operators especially in quality assurance tasks. If this

[1] For example, by building 'National data backbones', as discussed at the recent SIS-CC workshop – see report here.

promise is indeed actualised, AI-enabled statistics could well prove to be a critical factor for statistical organisations to achieve the productivity gains needed for them to invest in becoming data science organisations.

- *Data science environments to embrace a wide range of techniques with varied degrees of expertise*: The range of data types and techniques to develop is extremely wide (micro-data analysis, spatial analysis, socio-econometric modelling, text mining, visual data analytics, policy simulation…). Very successful open source projects (R, Python, Spark, Jupyterhub, ArcGIS…) provide free data science tools, plus free algorithms and access to large expert communities – they represent a huge opportunity for statistical organisations, provided they invest in them by making them key components of their next generation platform, contribute and are active in the relevant communities and take on decisive actions in building corresponding skills amongst their staff.
- *Data lakes to aggregate multiple data types, over hybrid infrastructures, connected to multiple third party data lakes*: Organisations' data lakes will capture data from various types and structures, sourced from the outside. Full cloud solution does not seem to be the way to go – for now – both for financial, and security/privacy compliance reasons. Statistical organisations will have to consider investing in hybrid infrastructures (on premise + cloud), with strong distributed computing and storage features that can be leveraged by properly designed algorithms, and linkages to third party (commercial and institutional partners') data lakes – including via distributed protocols. The tasks of sourcing data, storing them in the data lake, preparing and curating them, linking them to other data and cataloguing them etc. require significant know-how and resources not necessarily currently available in organisations.
- Open *algorithm banks at the heart of the new Platform*: Algorithms are key assets of statistical organisations and represent a potentially significant contribution to the knowledge economy, and a powerful way of connecting to broader data science communities – through sharing of open algorithms and associated knowledge. They are to serve, on top of highly performant distributed computing environments, both the statistical process as well as data science research work.

22. In summary, the next generation, AI-enabled platforms will be highly complex. This requires significant and sustainable investments in IT solutions and infrastructures, as well as in upgrading staff skills and dealing with legal and ethical issues. It will be difficult for statistical actors to address these challenges in isolation and new business models need to be invented to leverage cooperation in national as well as international 'data ecosystems'.

2.3. Building Ecosystems: New Forms of Cooperation

23. Given the scale of the transformation ahead, to overcome the sourcing, platforms and skills gaps, ecosystems will have to be built, amplified and connected at different levels (Figure 4):

Figure 4: New Ecosystems

Ecosystem building and connecting to existing ecosystems are key activities going forward, as no statistical organisation is able, alone, to overcome the sourcing, platforms and skills gap.

- *Within organisations*: statistical organisations need to reassess their structures and question dispersion of data expertise in multiple teams. One response can be to develop transversal communities of practice, within which knowledge and practices can be exchanged across established structures. These communities need to be enriched by external expertise such as academia, civil society, or specialised companies. The scale of investment in new data sources and platforms likely requires more centralisation or a greater degree of specialisation. As already stated, the open source model is predominant in this field, and connecting to existing, powerful open source software and algorithms communities is of crucial importance.
- *In the context of the National Statistical System*: NSOs are in a natural position to orchestrate a national Smart Data Strategy. The scarcity of skills, the necessity to develop a favourable negotiating position vs. data providers and the intensity of required investments in platforms – all of these factors converge towards the notion of a strong role for NSOs, (a) to develop global partnerships with data providers, (b) to develop the reference national data lake – connecting private as well as public sector data lakes, as well as to strategic open source communities – and (c) launch a national initiative to develop data science skills in all administrations, from the 'data for policy' perspective.

- *Within the broader international ecosystems*: International fora can offer the space where to: (a) raise awareness as regards the importance of developing smart data strategies, thanks to sharing of experience and good practices; (b) develop common frameworks and standards, for example in the area of distributed protocols or crowdsourcing; (c) coordinate dialogue with the large digital players (including regulation options) and data providers (including joint procurement options); (d) mutualise costs by jointly investing in targeted components of the strategy – AI-enabled statistics solutions could be one example; 'data for policy' training and capacity building efforts could be another example.

24. Cultivating ecosystems – within and beyond each statistical organisation – is hence a crucial component of a smart data strategy, to overcome the sourcing, skills and platforms gaps and, more fundamentally, to mutualise costs that no one organisation can afford alone.

25. In summary then, the world of statistics and data is undergoing rapid change. New demands have arisen that challenge some of the established workings of NSOs but also offer many opportunities. NSOs have started to actively embrace change, exploring the opportunities of new data sources, new partnerships and new business models. Facets of change are manifold and complex and addressing them requires strategies and frameworks to implement them. Because issues are similar across countries, international co-operation will also be helpful in this domain.