

# BUILDING NATIONAL DATA BACKBONES

Empowering countries through capacity development and technology

## 2018 Workshop Highlights Report



**SIS-CC**

*Statistical Information System  
Collaboration Community*

**SIS-CC** 2018  
WORKSHOP

09-10 APRIL 2018  
OECD PARIS, FRANCE

We are pleased to present the **OECD's Statistical Information System 2018 Collaboration Community workshop report**, that took place on April 9<sup>th</sup> to 10, 2018 in Paris, France.

### What's included in this report?

- Some facts and figures
- The participating organisations
- The workshop in brief
- Links to all presentations made over the 2 days including the live recordings
- Concluding remarks
- The case for Building the National Data Backbone

### Some Facts and Figures



**A total of 93 participants** from 39 organisations

More than **8,000 interactions on Twitter**

**115 watched the live stream** from 10 countries

### The Participating Organisations

#### Community Members

Australian Bureau of Statistics  
Statistics Estonia  
International Labour Organisation  
Italian National Institute of Statistics  
National Bank of Belgium  
OECD  
Tunisian National Institute of Statistics  
UK Data Service

#### *Including new members joining this year:*

Statistical Institute of the Autonomous Province of Bolzano  
National Institute of Statistics of Chile  
National Statistical Office of Luxembourg  
UNICEF

#### Partners

PARIS21  
EUROSTAT  
UNSD

#### Interested Organisations

IMF  
ECB  
Pacific Community  
National Institute of Statistics of Cambodia  
Central Statistical Office of Ireland  
Statistics Finland  
INEGI  
INSEE  
ONS (UK)  
National Bank of Denmark  
German Central Bank  
Central Bank of Lithuania  
Swiss National Bank  
FPS Economy Belgium  
Ministry of Finance of Turkey

#### Technology Partners and Experts

Helmes AS  
Tata Consultancy Services  
Metadata Technology  
Community Systems Foundation  
QuantCube Technology  
Positium LBS  
Open Knowledge  
Gartner  
Smile

## The Workshop in brief

This years workshop theme was “**Building National Data Backbones - Empowering Countries through capacity development and technology**”.

The **keynote** on day 1 was given by Martine Durand, OECD Chief Statistician, Director of Statistics and Data Directorate, and chair of the SIS-CC Strategic Level Group (SLG), highlighting the increased importance of the work of the Community to enable collaboration on statistical solutions between organisations, both at national and international levels, to improve the relevance and efficiency of our statistical products and processes.

Martine reminded participants that the Community continues to align to a [5-year strategy](#) with overarching governance providing mechanisms to pool resources and ensure alignment to each organisation’s strategic directions and the wider international initiatives. **The initial directions have been somewhat stretched to take into account a changing landscape**, as well as the full redevelopment of .Stat as we know it, with the wide extension of its functional scope to cover the full data cycle, and with the community going fully open source – this is the vision with the new .Stat Suite. This new paradigm will be effective in 2018, with the release of the new .Stat Suite as open source, which will enable the community to bring its solutions to the world at large.



Martine went on to clarify the connection between this Community and **the broader data for development and Sustainable Development Goals (SDGs) agenda**, endorsed by all UN States in September 2015. Data and indicators are playing a big role in the implementation and monitoring of progress towards the SDGs. The OECD is supporting the United Nations in ensuring the success of the 2030 Agenda for Sustainable Development by bringing together its existing knowledge, and its unique tools, data and experience.

With UNICEF, now a full member of the SIS-CC, and UNSD joining forces with Paris21 and OECD on several SDG reporting pilot projects, this commitment takes now a very concrete shape. In this regards, Martine especially thanked Claes Johansson, who represented UNICEF at the workshop, and Abdullah Gozalov, who represented UNSD, and Panhara Oukchay, who represented the Cambodian Statistical Office.

Martine concluded by stressing the importance that the outcome of this workshop – the first two days open to Community members and interested organisations and partners, the following three days for Community members only – with an expectation that is not only a consolidated vision for the .Stat Suite product roadmap, but also initial **thoughts and ideas to guide the strategy for the upcoming 2019-24 cycle**.

## Session by session presentations

| Session                  | Title  | Presenter and Organisation  | Link(s)                      |
|--------------------------|--|---|------------------------------|
| SIS-CCommunity in action | Business Case for joining the Community  | Claes Johansson, UNICEF   | <a href="#">Presentation</a> |
|                          |  | Christian Welter, National Statistical Office of Luxembourg                 | <a href="#">Presentation</a> |
|                          |  | Timon Gaertner, Statistical Institute of the Autonomous Province of Bolzano | <a href="#">Presentation</a> |
|                          |  | Andrew Carter Enrione, National Institute of Statistics of Chile            | <a href="#">Presentation</a> |
|                          | Member Project updates   | Jeremy Michell, Australian Bureau of Statistics                             | <a href="#">Presentation</a> |
|                          |  | Kadri Naumanis, Statistics Estonia  | <a href="#">Presentation</a> |
|                          |  | Edgardo Greising, International Labour Organisation                         | <a href="#">Presentation</a> |
|                          |  | Domenico Scalzo, Italian National Institute of Statistics                   | <a href="#">Presentation</a> |
|                          |  | Aurélien Chabriel, National Bank of Belgium                                 | <a href="#">Presentation</a> |
|                          |  | Kamel Abdellaoui, Tunisian National Institute of Statistics                 | <a href="#">Presentation</a> |
|                          |  | Victoria Moody and Richard Wiseman, UK Data Service                         | <a href="#">Presentation</a> |
|                          | .Stat Suite Product roadmap  | Jens Dossé, OECD  | <a href="#">Presentation</a> |
|                          | SDMX Collaboration Activities  | Nadezhda Vlahova, Eurostat  | <a href="#">Presentation</a> |
|                          | .Stat Suite Community Driven Architecture                                      | Bruno Urban, OECD   | <a href="#">Presentation</a> |
| Setting the stage        | Keynote - Frictionless Data an introduction                                    | Rufus Pollock, Open Knowledge   | <a href="#">Presentation</a> |
|                          | Methodology and Introduction   | Eric Anvar, OECD  | <a href="#">Presentation</a> |
| Assessing the issues     |  | Francesco Rizzo, Italian National Institute of Statistics                   | <a href="#">Presentation</a> |
|                          |  | Ann Vanderbusse, National Bank of Belgium                                   | <a href="#">Presentation</a> |
|                          |  | Kamel Abdellaoui, Tunisian National Institute of Statistics                 | <a href="#">Presentation</a> |
|                          |  | Panhara OukChay, National Institute of Statistics of Cambodia               | <a href="#">Presentation</a> |
|                          | Panel discussion   | Chaired by Abdulla Gozalov, UNSD  | None                         |
| Exploring solutions      | Reference Open Source Community  | Jonathan Challener, OECD  | <a href="#">Presentation</a> |
|                          | The sectorial angle, experience in the area of labour statistics               | Edgardo Greising, ILO   | <a href="#">Presentation</a> |
|                          | Data flow framework – methodology for assessing countries' needs and solutions | Rajiv Ranjan, PARIS21   | <a href="#">Presentation</a> |
|                          | Break-out groups   |   | None                         |
| Envisaging disruption    | Big data journey at the IMF  | El Bachir Boukherouaa, IMF  | <a href="#">Presentation</a> |
|                          | Smart Data Series for Strategic Intelligence                                   | Thanh-Long Huynh, QuantCube   | <a href="#">Presentation</a> |
|                          | Human Mobility Statistics from mobile phones                                   | Siim Esko, Positium   | <a href="#">Presentation</a> |
|                          | Q&A with the floor   | Chaired by Christian Reimsbach-Kounatze, OECD                               | None                         |
|                          | Final Panel discussion   | Chaired by Claes Johansson, UNICEF  | None                         |

Each session where presentations were made are available through Video-on-demand from the links below:





## Concluding remarks

Paul Schreyer, Deputy Director, Statistics and Data Directorate, OECD, closed the day highlighting the excellent discussions that took place over the 2 days. It is clear from the number of participants that attended the workshop, along with the increased membership, the community is both thriving and growing.

During day 1 we were presented with the Business case for joining the Community from our newest members: Institute of Statistics of Chile; Statistical Office of Luxembourg (STATEC); ASTAT (Statistical Institute of the Autonomous Province of Bolzano, Italy); and UNICEF, with the single most quoted reason being '*Community*'. But not just any community, a community around **open source**, very different from a community around a commercial product, providing an environment for shared decision making, modern, sustainable, and flexible enough to adapt to a continuously changing environment.



Paul exchanging the MoU with Andrew Carter from INE Chile

The main topic of the workshop "**Building National Data Backbones**", with the term 'backbone' at first bringing a vision of something rigid, strict, medical focus, but in fact, 'backbone' can be seen as being highly flexible, and concentrates the most important flows of information between brain and rest of the body in orderly way. Many useful discussions took place addressing some of the technical issues mainly about the **role of NSOs** in building data backbones.

**It is clear the role of NSOs** needs to shift, from coordinators of national statistical system (NSS) and near monopoly producer of official statistics, to more of the coordinator and animator of national statistical ecosystem and data steward. This implies at least four key changes:

1. Dealing with a broader set of actors, public and private;
2. Embracing new sources, 'Smart Data';
3. Taking on a much bigger role in leading on data modelling so that the data system is more than the sum of all its parts;
4. Dealing with an increased number of institutional and legal issues: privacy, confidentiality, data ownership, sustainability of sources.

The SIS-CC has been developing both a model and a Suite of tools. ".Stat Suite", to help make this transition happen.

Two aspects that need to be considered in the planning ahead: Delivering on some important functions – Embargo: this is not only a technical issue, there is a need for conversation and agreement on what exactly we mean by 'embargo'; Reflecting on content-related capacities that are key for implementation: frameworks for training of data modellers and pooling information on **user research**.

In summing up Paul confirmed that the Community has come a long way since it's inception some years ago, with a strong and positive outlook for the future – many of the ideas brought together here will make it into the strategic planning for the next 5 years. For the Community members, they will spend the next 3 days to discuss and start to plan the next 5 year cycle.

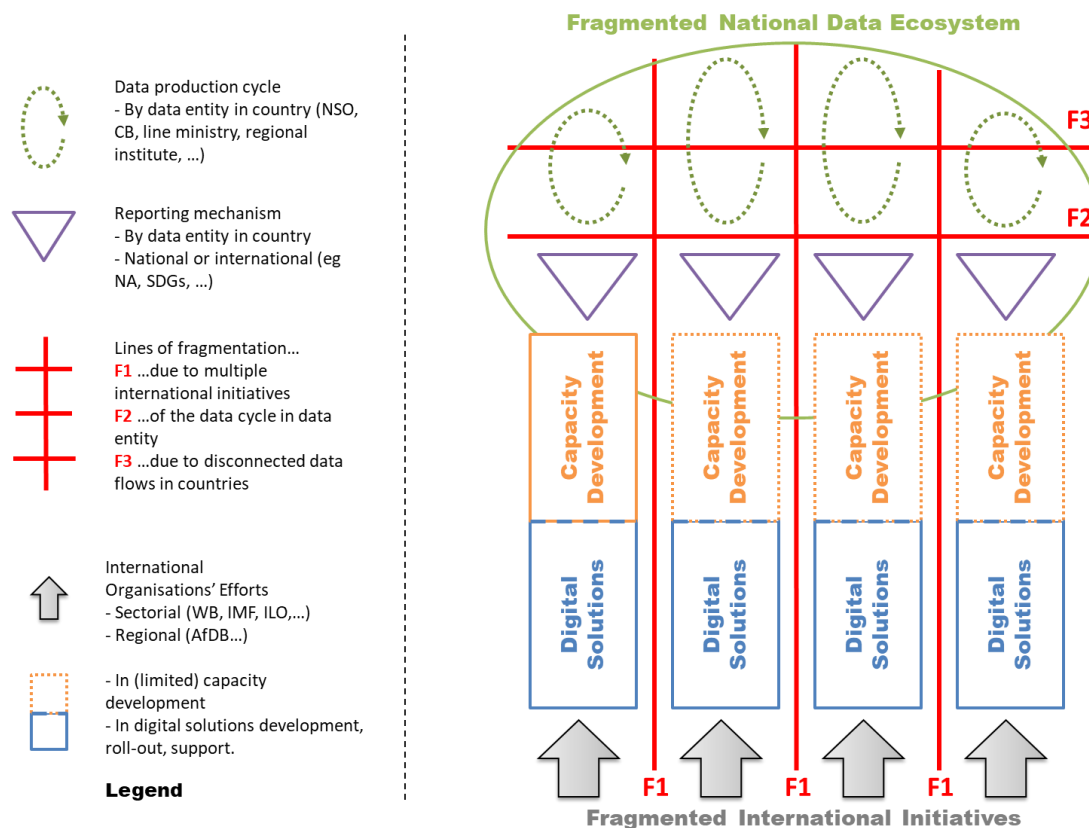
Paul thanked all participants for joining this year, and bringing their valuable contributions over the last 2 days and looking forward to seeing everyone again at the next year.

## The case for Building the National Data Backbone

In this section, the concept of 'National Data Backbone' as introduced during the workshop, the interrogations it raised and some of the thoughts and ideas expressed during the workshop – are captured, as a first contribution to this emerging concept.

### What are the issues?

At SIS-CC workshop in 2017 – [Sustainable Data for Development – Building an inclusive data ecosystem](#) – we explored options for a change of paradigm in building a truly global, open and inclusive data ecosystem to deliver on SDGs but also, more broadly, data for development agenda. The problem that was identified as a series of fragmentations of the international and national data ecosystems, preventing objectives at both levels being delivered synergetically:

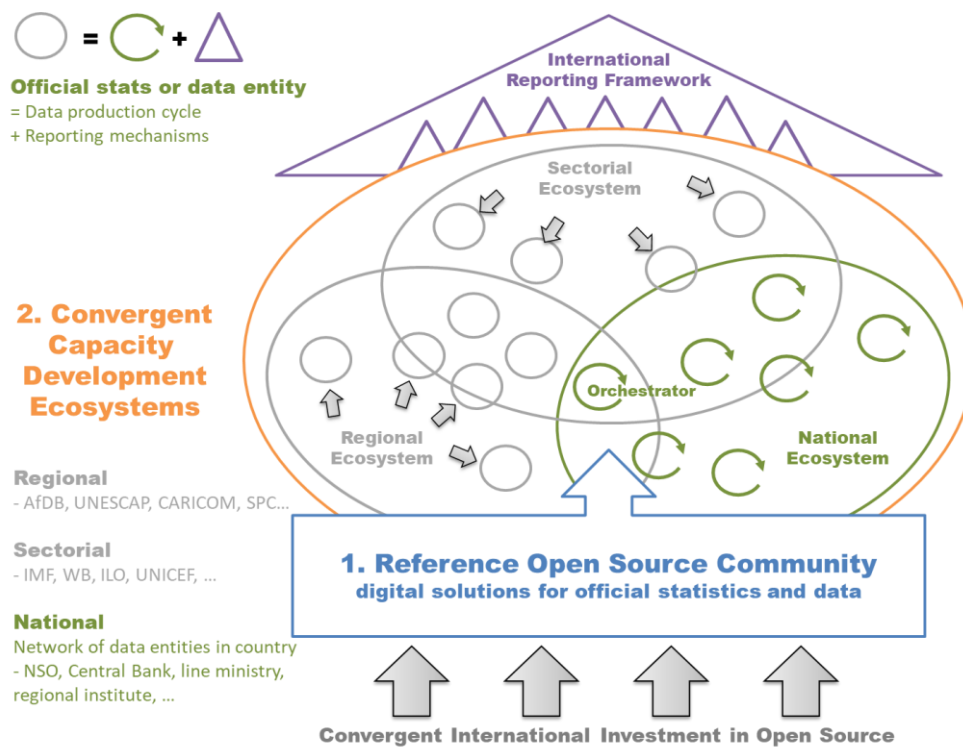


**Figure 1:** Multiple lines of fragmentation disrupt the data value chain at national and international levels

- **Fragmentation due to multiple international initiatives damages the data user experience.** As analysed by PARIS21 (see [report](#)) many initiatives, over the past years, have pushed for various platforms with limited coordination and synergies at international levels – leading to a complexity overhead at country level, in managing multiple platforms sometimes disseminating the same data, and confusion or damage to the data user experience. Also, emphasis has been put more on technical platforms rather than capacity development.
- **Fragmentation of the lifecycle in each entity damages its capacity to produce data**, as much focus has been put on the reporting and dissemination of data and lesser attention given to the process for creating the data product. This has led to limited human and IT capacity being built in the background in order to address the deeper process issues at national and regional levels, where data is being collected and produced in the first place. As a result, the level of productivity, and quality of the data produced can be sub-optimal.
- **Fragmentation of the national data landscape due to disconnected data flows** partially results from the above, but mainly results from the inner dynamics at work in countries – multiplicity of players (NSO, Central Bank, line ministries, local governments,...) and the challenges (in terms of technical solution, capacity development, data governance and political support) to build a consistent framework where these data flows can be orchestrated and connected.

## What are the possible solutions?

During discussions at the [SIS-CC 2017 workshop](#), but also at following fora (especially: the UN organised [SDMX Conference in Addis Ababa](#), October 2017; [UNSD workshop on National Reporting Platforms](#) in NY, January 2018), a new paradigm has emerged, based on two broad lines of action:



**Figure 2: Decoupling 3 layers – open source, capacity development, reporting framework – to enable convergence**

1) **Create the open source community to deliver reference digital solutions for official statistics and data.** This community should be focused on product excellence, and enforce key principles:

- **Pooling resources to mutualise costs** (vs. competing projects leading to inefficient allocation of resources and user confusion) to produce best-in-class, user friendly and standards compliant open source software – accessible to all, including lower capacity countries. This community should save resources for more capital intensive activities in the area of capacity development, which it should not be managing.
- **Component oriented architecture** (vs. monolithic approaches that limit contextual adaptation or synergies between initiatives at component level): component architecture is key in enabling organisations to assemble and customise a set of open source components into their own specific data platform, adapted to their context. This approach also enables to leverage the good work of multiple existing open source initiatives (such as [Eurostat SDMX project](#)), to build the broader solutions needed without reinventing the wheel.
- **Full data cycle coverage** (vs. only focusing on reporting needs, and not addressing the deeper ‘back office’ issues statistical organisations are facing): there is very large convergence in functional requirements across the board in statistical organisations, and across the different steps of the data lifecycle ([GSBPM](#): collect, process and disseminate). This creates the possibility of building the ‘ERP of official statistics’, progressively covering the main steps of the statistical process.

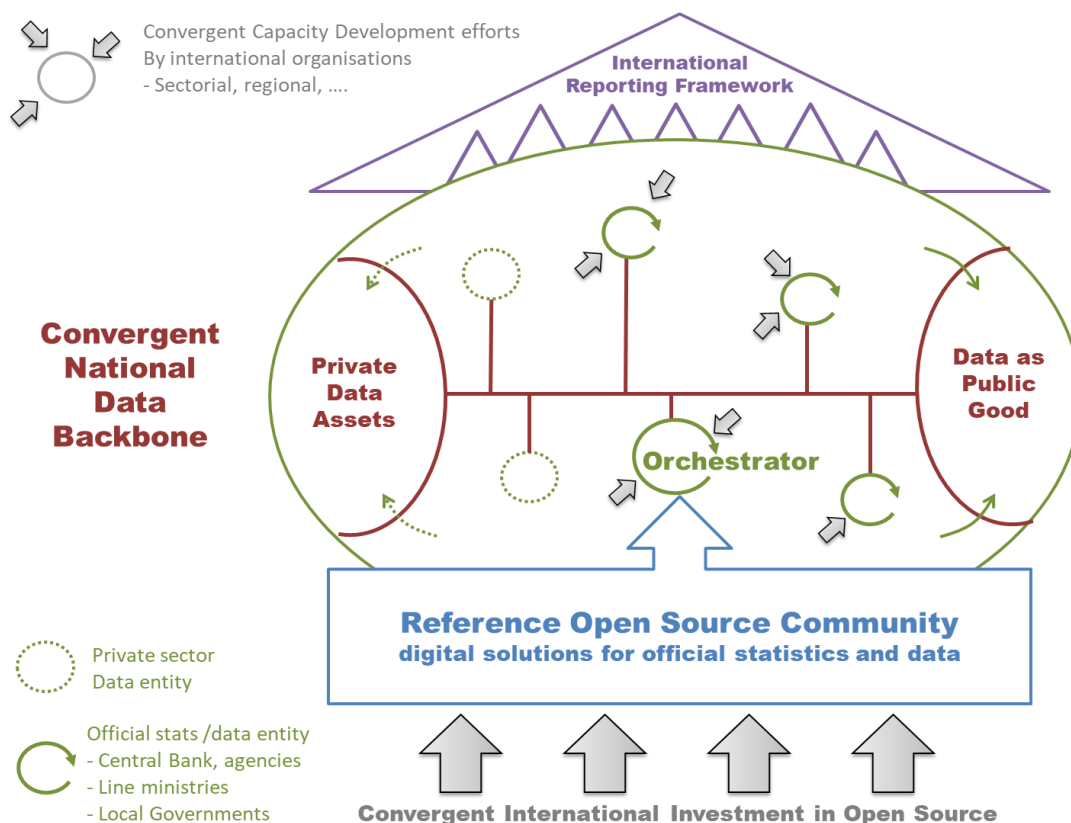
2) **Enable and develop the federation of capacity development communities:** not one single central community can support all the various entities, of various size and degrees of capacity, at local, national and regional levels, in their efforts to streamline statistical and data processes and increase the scope and quality of outputs. Conversely, with the assumption of an existing reference open source community for official statistics and data solutions (see previous action) – the possibility of concretely federating efforts within and between countries becomes more achievable. Regional entities (such as [African Development Bank](#) or [UNESCAP](#)), as well as sectorial entities (such as [ILO](#) or [UNICEF](#)), and of course national reference players (National Statistical Offices) could develop approaches whereby they can create synergies in their various capacity building efforts – which could never be achieved so far, due to the previously mentioned fragmentations. From the countries’ perspective, this approach creates the unique opportunity to federate and align all the different contributions and supports into one global, consistent national approach.

## How could these solutions be actioned?

The 2018 Community Workshop presented an opportunity to cement the view that **it is now time for a paradigm shift whereby SIS-CC could become in the future the reference open source community to bring about common solutions in the area of official statistics and data.**

It was clear from the discussions that the advent of such community would greatly help synergise the international efforts and make sure that at country level, multiple international initiatives do not compete or create overhead, but rather work in sync with each other.

During the workshop it was explored as to how the paradigm which emerged at international level could also possibly be applicable at national level, with the concept of National Data Backbone, as a means to achieve both the national and international agendas:



**Figure 3:** Transposing the previous model from the country perspective: Building a National Data Backbone

The directions explored mirrored the ones explored at the 2017 workshop, but this time looked at from the country perspective, to understand the national ecosystem dynamics and how it could be further empowered to deliver on national and international agendas.

**Open Source as a key lever:** The existence of a reference open source community for official statistics and data solutions, covering the full data cycle, can bring about a significant change, enabling an open source approach to be leveraged at national level with the possibility of a globally hosted platform and potential larger volume of contributions to the open source project.

This of course is not without some key considerations. For one, developing open source solutions is not free. Ongoing maintenance (bug corrections, security issues, etc.) and support of user communities come with a cost. Contributors need appropriate skills. Also, component oriented architecture will allow countries, and organisations and entities within countries, to capitalise on the open source assets while also being able to assemble blocks and tailor them appropriately to their own context and process.



**Harmonized International initiatives:** By bringing together the different currently fragmented efforts (international support and programs— whether sectorial or regional) to support capacity and platform development, the country-level efforts to build a national data backbone can be reinforced. With this goes also the necessary shift from the focus being mainly on data dissemination / SDG reporting, to investing more into data production and human capacities.

### **Two proposed actions**

1. At National level: A layer to be added to National Strategy for Developing Statistics (NSDS) to define international coordination, and increase the consistency of the approach to capacity development for data providers and data modellers.
2. At International level: Define a National Data Backbone charter, binding international initiatives to reduce the overhead rather than increase it, at national level, and synergise systematically with other initiatives through agreed channels (such as: invest in reference open source solutions).

**The National Data Backbone:** With an emphasis on designing the system and the national framework for data flows to meet the national agenda (for example, aggregation from local to central governments, or amongst the different line ministries) and international agenda (SGDs and other reporting activities), and put in place the appropriate technical mechanisms to support a fluid framework of data flows.

Involve the citizens in the statistical programmes to be at the center. Have a single authority ("orchestrator") to simplify as much as possible the governance structure of the dataflows – NSOs are naturally to pick up that role in most countries. There should be a policy of 'collect once reuse multiple times'. It should be regulated but also involve and convince the different stakeholders. Reporting to international organisations has thus far been seen as a burden, but should be inversed to become more of an enabler for better organising the dataflows. This of course requires a more coordinated effort at international level to reduce the existing reporting burden and fragmentations as described previously.

**Orchestration of the National Data Backbone:** There are some distinct needs of the players in the field: NSOs, line ministries, local governments, central banks and other agencies, and each national ecosystem comes with specific challenges. The concept of National Data Backbone can emerge as an overarching framework servicing statistical reporting objectives as well as the broader open data agenda, and how new sources of (mostly private) data could be transformed into data for public good through 'smart data' practices.

The governance of this backbone, with the role of the orchestrator, should be mandated by way of law. This would provide the motivation for national agencies to be part of an efficient data ecosystem, with coordinated exit points (vs. single statistics exit point) that help overcome the various fragmentations. It is recognised that a great variety of situations from one country to another must be taken into account, and would have to be clarified at a country level with the respective line ministries, agencies, local government, as well as international institutions (sectorial or regional).

### **The 'smart data' disruption**

Multiple 'smart data' opportunities are emerging in the area of official statistics and data in general: making new evidence out of new sources of data, and/or leveraging new algorithmic techniques. New data players have appeared – mostly from the private sector – that create new data assets, corresponding to the ever growing digitalisation of our lives and society in general.

Of course this brings many challenges and a potential disruption to orchestrating the National Data Backbone, but also a high potential for value add to complement the existing and traditional statistical data production line. Clearly this cannot be ignored and further discussion will be needed to identify what role these new players will play in Building the National Data Backbone.

## WHERE TO FIND US

 [SISCC.OECD.ORG](https://siscc.oecd.org)

 [@SISCCcommunity](https://twitter.com/SISCCcommunity)

 [bit.ly/linkedin-sis-cc](https://bit.ly/linkedin-sis-cc)



**SIS-CC**

*Statistical Information System  
Collaboration Community*

**SIS-CC** 2018  
WORKSHOP 

**09-10 APRIL 2018  
OECD PARIS, FRANCE**