The value of data science for the implementation of the Sendai Framework for Disaster Risk Reduction 2015–2030

Authors include Arofan, Dan, Jay, Rebecca and Virginia

Acknowledge all of the participants

**UN member states should take advantage of the global data science systems, mechanisms, networks and resources in order to implement the Sendai Framework**

Abstract

**Introduction**

**Background (Includes Sendai and Data Science - missing Theory of Change)**

Three years have passed since the adoption of the Sendai Framework for Disaster Risk Reduction 2015–2030 by 187 United Nations Member States at the Third United Nations World Conference on Disaster Risk Reduction in Japan in March 2015 (UNISDR 2015). Three other UN landmark agreements linking directly to the health aspects within the Sendai Framework were made in 2015—the Sustainable Development Goals (United Nations 2015), the Paris Climate Agreement (UNFCC 2015), and the Habitat III New Urban Agenda (United Nations Habitat III 2016). Margareta Wahlström, Former Special Representative of the Secretary-General for Disaster Risk Reduction and Chief of UNISDR stated that. “Access to information is critical to successful disaster risk management. You cannot manage what you cannot measure.” (UNISDR, 2012) Therefore all must be ‘able to monitor and review implementation’ calling for ‘a data revolution, rigorous accountability mechanisms and renewed global partnerships’ (UN, 2015).

The Sendai Framework aims to reinforce the shift in policy and practice of governments and stakeholders from managing disasters and other events to managing disaster risk. The Framework’s success will be assessed through action at all levels—local, regional, national, and global (Wahlstrom 2015). Rather than focusing exclusively on the response to emergencies, the Sendai Framework recognizes that by reducing and managing conditions of hazard, exposure, and vulnerability—while building the capacity of communities and countries for prevention, preparedness, response, and recovery—losses and impacts from disasters can be effectively alleviated

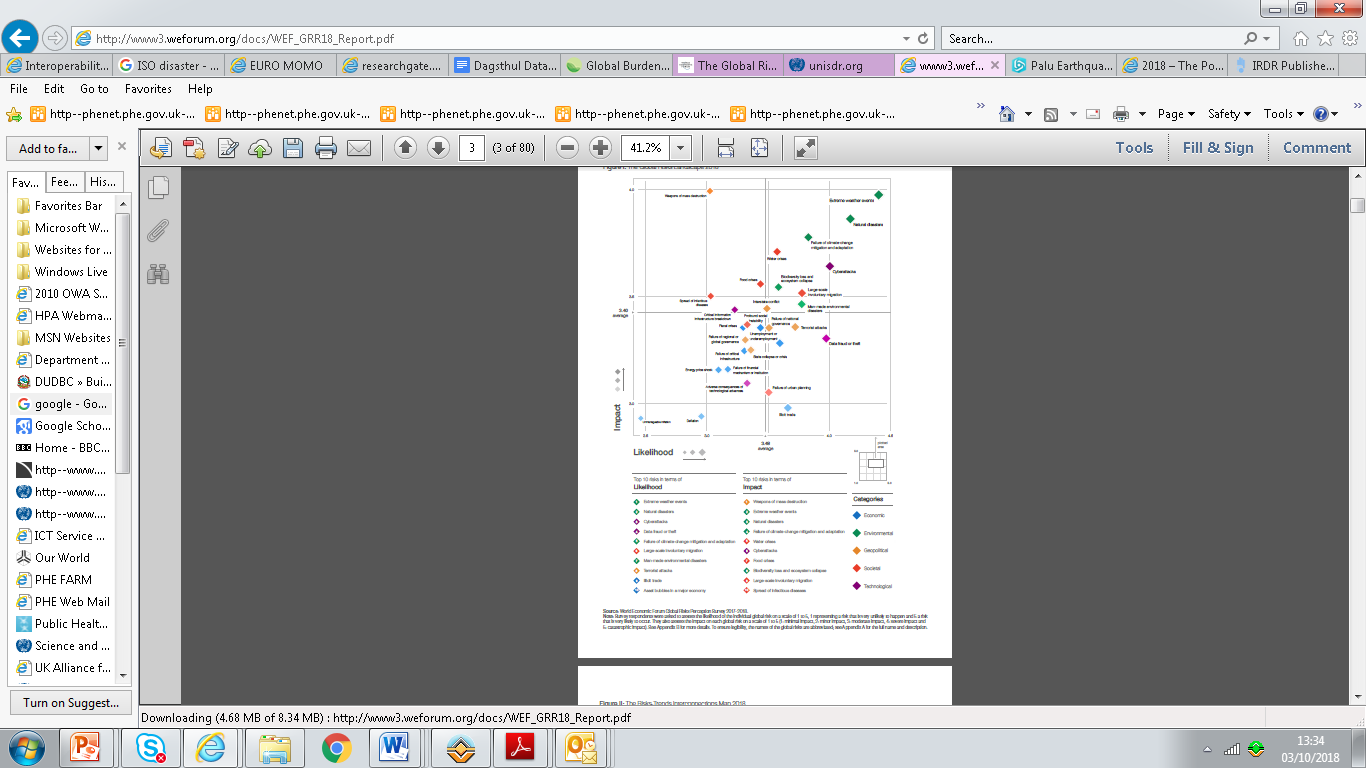
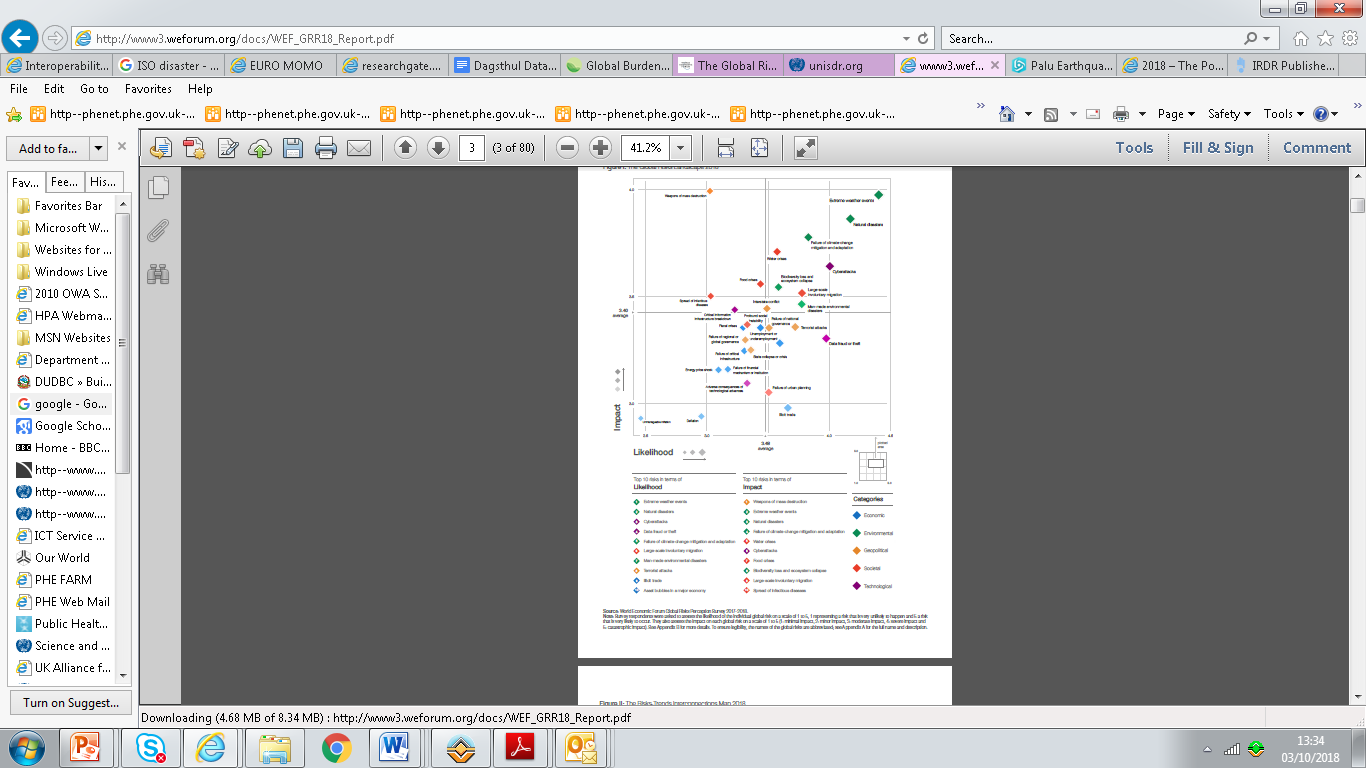
Adoption of the Sendai Framework by the UN Member States included agreement on seven global targets with 38 underlying indicators to assess global progress in disaster risk reduction. Paragraph 18 of the Sendai Framework states that:*‘these targets will be measured at the global level and will be complemented by work to develop appropriate indicators.*’ The seven global targets (UNISDR 2015, p. 12) are:

1. Substantially reduce global disaster mortality by 2030, aiming to lower the average per 100,000 global mortality rate in the decade 2020–2030 compared to the period 2005–2015;
2. Substantially reduce the number of affected people globally by 2030, aiming to lower the average global figure per 100,000 in the decade 2020–2030 compared to the period 2005–2015;
3. Reduce direct disaster economic loss in relation to global gross domestic product (GDP) by 2030;
4. Substantially reduce disaster damage to critical infrastructure and disruption of basic services, among them health and educational facilities, including through developing their resilience by 2030;
5. Substantially increase the number of countries with national and local disaster risk reduction strategies by 2020;
6. Substantially enhance international cooperation to developing countries through adequate and sustainable support to complement their national actions for implementation of the present Framework by 2030;
7. Substantially increase the availability of and access to multi-hazard early warning systems and disaster risk information and assessments to people by 2030.

UNISDR reported that on [2 February 2017, in adopting Resolution A/RES/71/276, the United Nations General Assembly endorsed the Report of the Open-ended Intergovernmental Expert Working Group on Indicators and Terminology Related to Disaster Risk Reduction](https://www.unisdr.org/we/inform/publications/54970) (A/71/644), and the recommendations for indicators and terminology relating to disaster risk reduction contained therein. In the Report of the OIEWG, Member States requested the United Nations Office for Disaster Risk Reduction (UNISDR) to undertake technical work and provide technical guidance inter alia to:

1. Develop minimum standards and metadata for disaster-related data, statistics and analysis with the engagement of national government focal points, national disaster risk reduction offices, national statistical offices, the Department of Economic and Social Affairs and other relevant partners.
2. Develop methodologies for the measurement of indicators and the processing of statistical data with relevant technical partners.

In March 2018 UNISDR has prepared and published [Technical guidance for monitoring and reporting on progress in achieving the global targets of the Sendai Framework for Disaster Risk Reduction (New edition).](https://www.unisdr.org/we/inform/publications/54970) This document was a preliminary draft for consultation, developed in response to the request of Member States. It builds on the recommendations and deliberations of Member States in the OIEWG, on the technical documentation produced by the Secretariat at the request of Members of the working group, on the deliberations of the Inter-agency and Expert Group on SDG Indicators (IAEG-SDGs), and on technical consultations with Member States and experts since the submission of the Report of the OIEWG and the Report of the Inter-agency and Expert Group on Sustainable Development Goal Indicators (E/CN.3/2017/2).The document goes onto provides technical suggestions and considerations of Member States, relevant technical partners and the UNISDR in respect of applicable definitions and terminology, possible computation methodologies, data standards and critical issues.

The first target to be delivered by each country by 2020 is Target e) when all countries are expected to have national and local disaster risk reduction strategies (UNISDR National Disaster Risk Assessment 2017). However even agreement on the risks each country faces is complex but can be summarised by the [World Economic Forum 2018](http://www3.weforum.org/docs/WEF_GRR18_Report.pdf) report figure 1 below. Of note those which are considered to have the greatest likelihood are 

Extreme weather events, natural disasters, cyberattacks, data fraud or theft, failure of climate change mitigation or adaptation and man-made environmental disasters are currently thought to be those most likely to occur and those considered to have the greatest impacts include weapons of mass destruction, extreme weather events, natural disasters, failure of climate change mitigation or adaptation and water crises.

However, the disaster data landscape is a complex one, though information on loss data is rapidly growing. When mortality or morbidity affect people along with monetary or environmental losses occur as a result of a disaster, extensive loss data are often collected and stored by different organizations, but the thoroughness and accuracy of the data vary from country to country and even among local entities. While many methods do exist, there is no standard that introduces a reasonable level of comparability into the resulting assessment results. This leads to gaps and overlaps in the data, and biases that ultimately affect the quality of research conducted and policies made on the basis of the data. Several systems exist to document disasters and these range from those recommended by UNISDR and used by individual countries such as [Disinventar](https://www.desinventar.org/) to those collecting data such as [EMDAT](https://www.emdat.be/explanatory-notes) but such systems are not able to provide all the data for reporting on the global targets and indicators. Therefore this paper considers how the UN member states should take advantage of the global data science systems, mechanisms, networks and resources in order to implement the Sendai Framework

The uses of global data science systems, mechanisms, networks and resources in order to implement the Sendai Framework

First principles for the uses of data science include:

* In the Sendai framework local data matters because it feeds up through national, regional and international levels
* There is no single solution that can assure the quality and completeness of the local data
* Data science can contribute to assuring the quality and completeness of local data by partially automating:
  + data quality,
  + data integration across diverse and locally specific data sources and
  + the actual production of the SDG Indicators

Data science, in the context of creating aggregate data like the SDG indicators, is able to provide information gathering algorithms. [Adaptive algorithms](https://en.wikipedia.org/wiki/Adaptive_algorithm), for example, can navigate through a field of data resources called a [domain model](https://en.wikipedia.org/wiki/Domain_model) on their way to creating statistics like SDGs.

Because domain models and the data resources that compose them are [spatiotemporally specific](https://www.merriam-webster.com/dictionary/spatiotemporal), adaptive algorithms for creating statistics produce local solutions that are subject to change over time.

Many of the examples below are the product of natural experiments in which adaptive algorithms have played a role. In different circumstances each one has been a survival of the fittest. In the future we might run some actual experiments to create data quality, data integration and SDG indicators using information gathering algorithms.

**Examples**

## Data Collection, Documentation, and Dissemination in the Developing World and Its Impact on Research: The International Household Survey Network (IHSN) Example

Within the international organizations which concern themselves with statistical data in low and middle income (LMI) countries is a collaboration involving several of the UN agencies: the International Household Survey Network (IHSN). The World Bank headed an initiative under this collaboration to build tools and build capacity for data collection by the national statistical agencies, with the majority of adoption being in Africa, Southeast Asia, and (increasingly) in the Americas. This effort is a good example of how effective capacity-building at the national level can be, even in countries where the official statistical organizations are often under-resourced. We also see research efforts in Africa and elsewhere in the world adopting the same tools, methods, and approaches in such fields as epidemiology based on the success of these efforts.

The IHSN has developed a number of tools which, used together, have proven effective in establishing and sustaining high-quality data collection and dissemination efforts in many countries. These include:

* Free tools for documenting data which has been collected, in a processible form which produces not only rich, human-readable documentation, but also machine-actionable metadata which can be repurposed to support discovery, access, dissemination, and analysis in a variety of ways.
* Detailed guidelines for how to design, implement, disseminate, and document surveys.
* On-site training, consulting, and related capacity-building efforts.
* Free tools for operating data portals/catalogs at the national level, based on the data and documentation collected using the IHSN toolkit.
* A centralized catalog of data sets collected within the IHSN, including search capabilities, fine-grained documentation of data sets, information about access, and a wealth of other information.

Unlike some efforts to support data collection in the LMI countries, the IHSN has taken a very practical approach, taking into account the high rates of turnover experienced by some statistical organizations and the low availability of specialized skills required to perform data collection effectively. Despite these challenges, the members of the IHSN network have produced more than 6000 data sets to date, searchable in the catalog hosted at the World Bank.

Both the technology and guidelines employed by the IHSN leverage international standards, reference models, and best practices. The metadata editing/publishing toolkit is based on an application developed by data archives in Canada, Norway, and the UK. This tool – Nesstar – has been widely adopted in Europe and the Americas by data archives and similar organizations. It is based on an open standard developed and maintained by the DDI Alliance, focused on data documentation and related metadata in the social, behavioral, and economic sciences. That standard – DDI – is also widely employed not only in these domains but also in the world of official statistics and increasingly in public health, childhood cohort studies, and other health-related research fields.

The guidelines published by the IHSN also employ a reference model for statistical production developed by the statistical arm of UN/ECE, the Generic Statistical Business Process Model (GSBPM). This model enjoys broad popularity among national and international statistical agencies internationally, and has been adapted for use with longitudinal studies in the social sciences and other related research domains (the Generic Longitudinal Business Process Model, GLBPM).

Based on the merits of the IHSN tools – notably the high level of support for sustainable use in the LMI countries – they have been adopted by many research organizations in those countries which are not involved in the collection of data for official statistics purposes. An excellent example of this is seen in the InDepth network, and the related Alpha Network, which concern themselves with epidemiological monitoring and demographic surveillance. Many of the research sites which are members of this network use the IHSN tools, and are (in the case of the Alpha Network) expanding the role of the GSBPM and the DDI metadata standard to provide highly detailed documentation of not only the data sets themselves, but also of their provenance and processing. The researchers who work with the compiled data will be provided with a remarkable level of information about data provided by these sources – one which is truly world-class in terms of its quality.

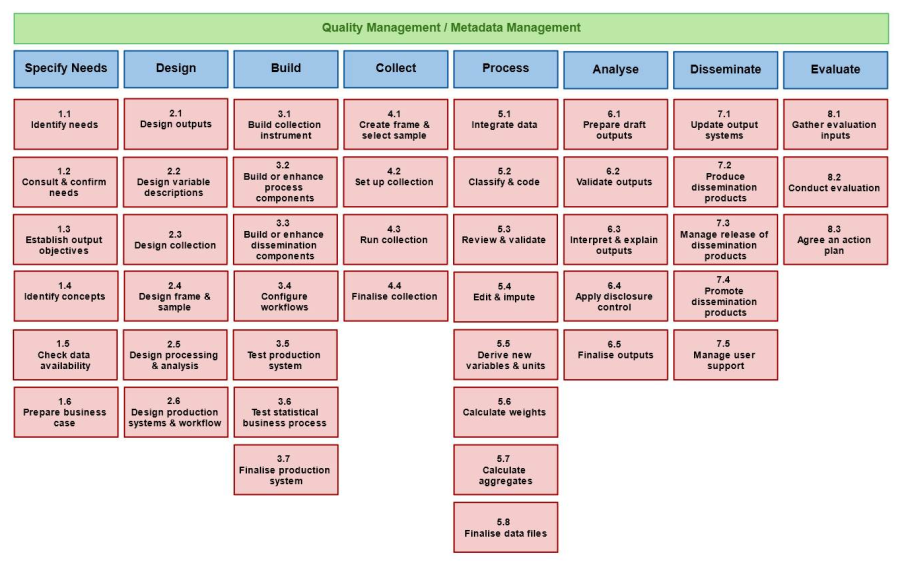
Overview of Generic Statistical Business Process Model (GSBPM)

The Generic Statistical Business Process Model (GSBPM) is a standard produced by the national statistical institutes and international statistical offices under the auspices of the UN Economic Commission for Europe (UNECE). The GSBPM was the first among several UNECE information standards for official statistics that address process, data, integration, and interoperability.

The purpose of the GSBPM was to create a common language and framework to understand the process of producing official statistics. Every statistical office conducts its business in the same general way. Therefore, a common process framework, as a standards producing exercise, was feasible.

The GSBPM contains an outline of the processes needed to produce statistics. It contains 8 main phases, 2 overarching processes that cross all phases, and several sub-processes under each phase. The ordering in which specific sub-processes are organized to describe how each statistical activity is carried out is not prescribed. Statistical activities include censuses, surveys, and other data acquisition projects, and each generally uses the phases and sub-processes identified. However, a language for naming each of the sub-processes is provided, and this provides a way for any statistical program to communicate with any other about what it is doing.

Here is a pictorial view of the GSBPM with the overarching processes, the 8 phases, and each of the sub-processes named and organized:



Quality and metadata management are the over-arching processes identified in GSBPM. Both quality and metadata are issues inherent in each process conducted by a statistical office. Quality is a set of measures describing how well some process is designed, built, and conducted. Metadata describe the inputs, outputs, and steps a process contains. Metadata also include the quality indicators applied to each process.

The GSBPM can also be used in conjunction with an information model for statistics. A process model organizes the things that happen. An information model organizes the inputs, outputs, and steps used for each process. In this sense, process and information models inter-relate, and they depend on each other. The GSBPM may serve the process model role for official statistics.

Many statistical offices around the world use GSBPM as part of their business. When adopting the GSBPM, a statistical office will adapt its own language to the phases and sub-processes identified in the model. This results in a business process model specifically for the agency that uses it, but one that is easily translatable into the standard. This is called a local profile.

Uses of the GSBPM mostly include the management of the software and IT resources needed to produce that software for conducting statistical activities. Many offices classify their IT systems, software development projects, IT resource acquisition, and strategic IT plans in terms of the local profile. The common and repeatable terms from which to classify these items is vital to communication, both within the office and across offices.

Another benefit is to identify the potential for duplicating software development or maintenance. Many statistical activities perform processes that are nearly identical, and statistical offices traditionally devote separate IT staffs to each major statistical activity within the office. The result is writing systems separately for each activity that uses them, and this is a huge waste of resources. Reducing duplication is important for preserving resources in times of limited budgets, including easing development and maintenance costs.

The GSBPM is a cornerstone for communication within and among statistical offices precisely because the underlying process model is commonly shared. Statistical activities in official statistical offices around the world are conducted in generally the same ways.

## 

## The Statistical Data and Metadata Exchange Standard (SDMX) and its use in the MDGs and SDGs

Much of the reporting infrastructure for official data flowing from national statistical organizations to international and regional organizations is supported by an infrastructure based on the SDMX standard. Developed by the BIS, World Bank, OECD, IMF, Eurostat, ECB, and the UNSD, this standard was first used as the technology infrastructure for the Millenium Development Goals Indicators (MDGs) and now for the Sustainable Development Goals Indicators (SDGs).

The basic workflow of SDMX infrastructures is designed to minimize the reporting burden on the national level, to support the standardized collection of data and metadata within a particular domain, to ease technology implementation by supporting the use of generic software tools, and to allow for the retention of control by the appropriate organizations over their versions of the data.

The workflow can be understood at a high level as illustrated below in Figure x:

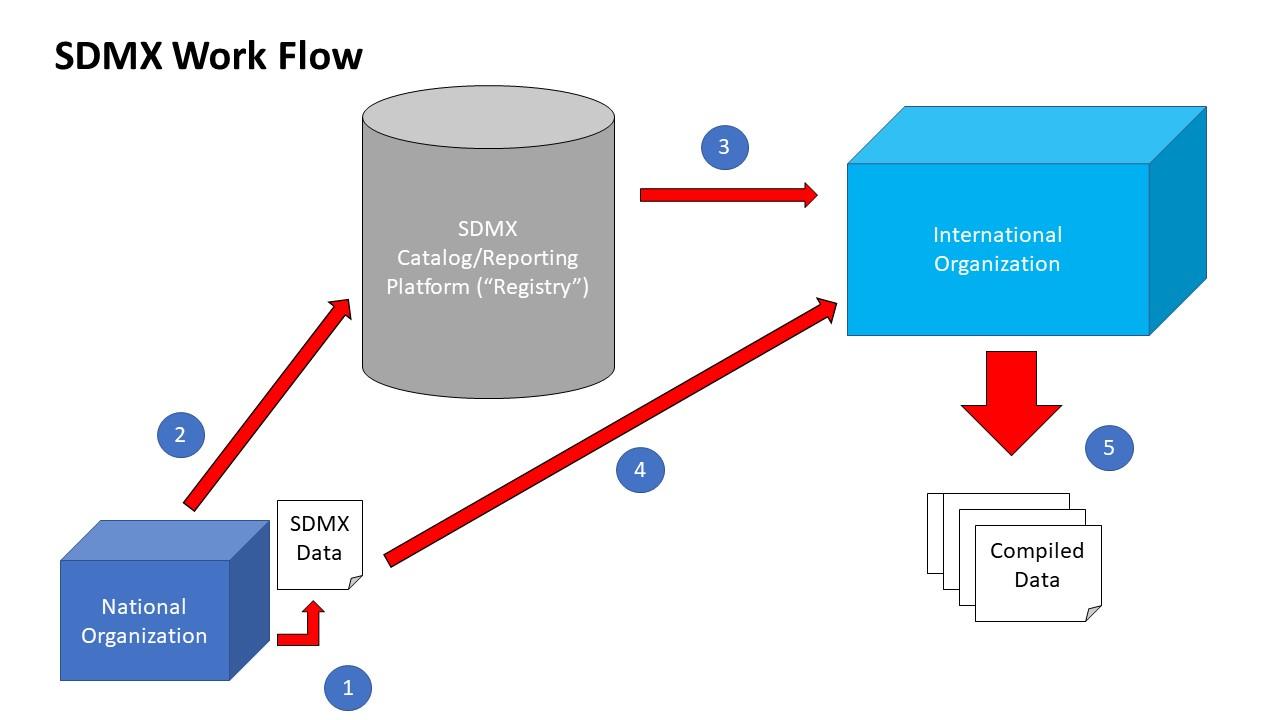


Figure x. The SDMX workflow: 1) National data is published by the national organization on its own site. 2) The national organization alerts the SDMX reporting platform that the data has been published. 3) A notification is passed to the international organization which is compiling the data coming from the reporting countries. 4) The international organization pulls the published national data from the national site, and performs the processing needed to include it in the broader compiled data set. 5) The compiled data set is published by the international organization on its own, centralized site.

One important aspect of this infrastructure is that the national version of the data – sometimes adjusted during the process of compilation at the international level – is maintained by the national organization which produced it. This version of the data is not copied, stored, and managed separately at the international level. Any changes to the national data are thus controlled by the national organization which produced it.

The international version of the country data – which is sometimes different from the national version – is published and maintained by the compiling organization. This model provides clear ownership and responsibility for the two versions of the data; there is no duplication of storage or management to cause confusion or problems. (The differences between these two versions – if any – are documented and explained as part of the metadata provided by the central site which publishes the compiled version of the data.)

All data and metadata reported using this infrastructure follows agreed SDMX structures, giving end users a consistent and comparable presentation of the data, and a rich set of metadata/documentation. (These structures are agreed by the domain organizations, as in the case of the SDGS – the structural definitions are described at [Link to SDG website]).

The technologies employed by this infrastructure are designed to support the development of generic tools which may be used by international or national organizations, including standard protocols for exchanging data and metadata, and standard services for querying, retrieving, and performing other common technology functions. The standard provides for the “SDMX Registry”, which can be understood as a central catalog or reporting platform for coordinating the flow of information through the system.

While the implementation of these standards may vary from domain to domain, to support the specific needs of the counterparties within that domain, the underlying technology is the same across all of them. The SDMX Reference Infrastructure provides a set of tools which can be combined in different ways to support the needs of various domains and their information flows.

The value of this technology approach is that it uses a set of standards which are familiar to the national statistical organizations and the international and regional organizations. While the infrastructure used for SDG reporting may not meet national needs exactly if deployed as a system for collecting data within a country, it can be used as a good example of how large-scale data can be implemented from a technology perspective. The same tools developed for use in compilation of data at the regional and international level will also work at the national level, once suitable data and metadata structures have been agreed among those organizations reporting and collecting data.

## 

## IMF – Dissemination Standards Bulletin Board

The International Monetary Fund (IMF) manages the dissemination of national statistical economic and social indicators with its National Summary Data Pages (NSDP) within the Dissemination Standards Bulletin Board website. The indicators from each participating country are contained on a page managed by that country and linked from the NSDP home page. These indicators are transmitted to the IMF from the national statistical offices producing them through the use of the Statistical Data and Metadata eXchange (SDMX) standard. This section describes this program.

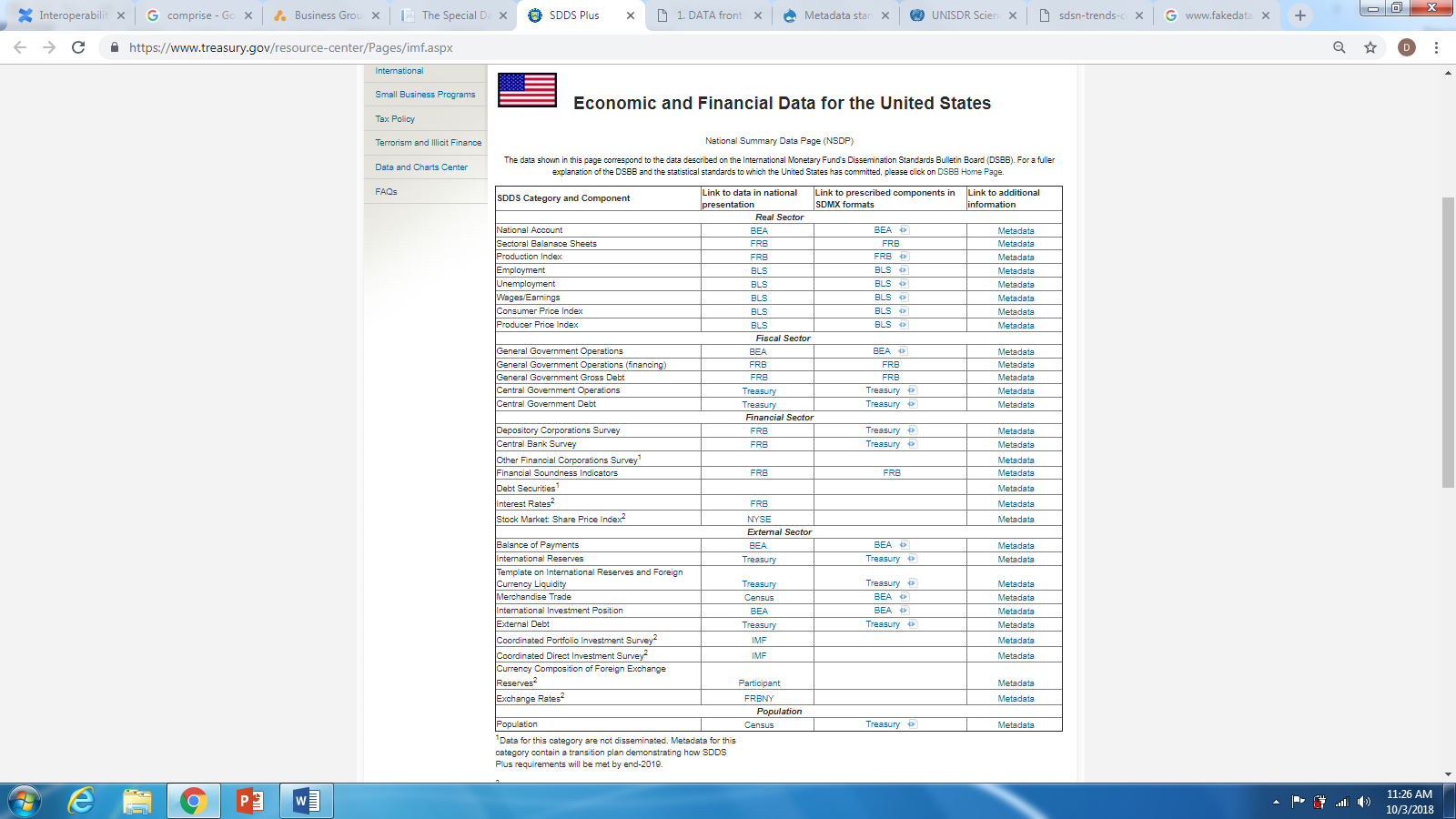
The National Summary Data Page ([NSDP](https://dsbb.imf.org/nsdp)) is a data portal for obtaining national indicator data by country. Countries participating in the SDDS, SDDS Plus, e-GDDS, metadata standards and DQRS data quality standard provide their data to IMF via SDMX. IMF established the Data Standards Initiative for the purpose of promoting the transparency of economic and financial data. The relevant standards are:

* SDDS – Special Data Dissemination Standard was established in 1996 by the IMF to guide members that have or seek access to international capital markets in providing their economic and financial data to the public.
* e-GDDS – extended General Data Dissemination Standard was established in 2015 to guide countries in data dissemination by supporting transparency, encouraging statistical development, and helping create strong synergies between data dissemination and surveillance. The e-GDDS superseded GDDS, which had been established in 1997. The GDDS was developed to support countries with less well-developed statistical systems.
* SDDS Plus – Special Data Dissemination Standard Plus was established in 2012 by IMF to guide member countries on providing economic and financial data to the public in support of domestic and international financial stability. The SDDS Plus builds on the SDDS to account for economies that play a leading role in international capital markets and has institutions interconnected, for example, through interbank lending, security lending, repurchase agreements, and derivatives contracts.
* DQRS – The Data Quality Reference Site is part of the IMF's ongoing effort to develop a common understanding of data quality. The site introduces definitions of data quality, describes trade-offs among different aspects of data quality, and gives examples of evaluations of data quality.

Each NSDP page allows users to access data, view metadata, or browse links to online datasets for all available categories for a given country. This is true even if these categories are compiled by multiple statistical agencies, such as in the USA. For countries participating in SDDS Plus and e-GDDS, the NSDP enables automatic exchange and sharing of statistical data and metadata in SDMX, a standard for machine-to-machine transmission.

Almost all countries in the world participate in the IMF program. Fewer than 20 use SDDS Plus, and about half the rest use e-GDDS and the other half use SDDS. Subscribing to the service obligates a country to supply the indicators required. Some others are encouraged. SDDS comprises four main categories: real, fiscal, financial, and external. Subscribers to SDDS Plus have to include nine additional categories. E-GDDS subscribers have fewer categories demanded of them, as this standard addresses emerging economies.

The following is a screen shot of the page from the US:



Note, there are almost ten different agencies reporting data. Each uses an SDMX Data Structure Definition (DSD) to transfer their data and metadata to the US Department of Treasury (DOT), who compiles the page illustrated. DOT built and maintains each DSD needed.

For each country, indicators are updated on a regular basis, but the timing of those updates varies depending on the indicator. An Advanced Release Calendar (ARC) is maintained to alert IMF to the release of updated indicators. This ARC is a required resource to be maintained by each subscribing country.

As the economy improves within some country, that country can elevate its reporting requirements. Each reporting standard (e-GDDS, SDDS, and SDDS Plus) allows for adding non-mandatory indicators, but once those new ones are a permanent part of the reported set, a new reporting level can be adopted.

# PCORnet and PopMedNet

PCORnet, the National Patient-Centered Clinical Research Network, is an innovative initiative of the Patient-Centered Outcomes Research Institute ([PCORI](https://www.pcori.org/)). It is designed to make it faster, easier, and less costly to conduct clinical research than is now possible by harnessing the power of large amounts of health data and patient partnerships. In the process, it is transforming the culture of clinical research from one directed by researchers to one driven by the needs of patients and those who care for them (from [https://pcornet.org](https://pcornet.org/)).

PCORnet together with its technical backbone PopMedNet is a United States approach for making clinical data hosted in US clinical settings available to responsible researchers in a way that maintains the data sovereignty of patients and their providers. It is a network of networks.

As an example, a PCORnet partner network was recently used by advocates working with Chicago’s Homeless Management Information System (CHMIS), a federally funded and locally administered technology system that is used to collect client-level data on homeless individuals and families. CHMIS collaborated with CAPriCORN, a clinical research network that is part of PCORnet. The goal was to link homeless agency data to the CAPriCORN network’s clinical data to generate some real numbers to show how people experiencing homelessness access and use different healthcare facilities. If, as expected, the data reveals a high degree of fragmentation or excessive use of hospitalization, these numbers would be influential in motivating health systems’ leaders to contribute to the housing pool.

PCORnet partner networks roll up patient clinical data from provider member electronic health records into a patient-level [Common Data Model](https://pcornet.org/pcornet-common-data-model/) (CDM). Each PCORnet partner network maps data to the same consistent format (i.e., with the same variable name, attributes, and other metadata). By undertaking this step, PCORnet create a platform that enables much more rapid responses to research-related questions.

The PCORnet CDM leverages standard terminologies and coding systems for healthcare (including ICD, SNOMED, CPT, HCPSC, and LOINC) to enable interoperability with and responsiveness to evolving data standards.

PopMedNet is the technical backbone of PCORnet as well as other health networks including [Sentinel](https://www.sentinelinitiative.org/), the [Cancer Research Network](https://crn.cancer.gov/) (CRN) and the [National Institutes of Health (NIH) Collaboratory](http://rethinkingclinicaltrials.org/) among others.

PopMedNet is an open-source application used to facilitate multi-site health data networks. It uses a distributed network design that enables data holders to retain full control of their data. Investigators send questions to data holders for review and response. PopMedNet eliminates the need for assembling patient records in a centralized repository, thus preserving patient privacy and confidentiality.

PopMedNet is made up of three primary components:

* **PopMedNet Query Tool:** The web-based application from which queries are distributed by the requestor
* **DataMart Client:**A small application installed by the data holder through which queries are reviewed, executed and responses are sent back to the requestor.
* **Web Service API**– API used to implement all application logic that processes requests against a network.

(from <https://www.popmednet.org/tools-and-features/standard-functionality>)

Using PopMedNet and a Common Data Model, it is possible to implement a network or network of networks for the rapid and secure exchange of clinical data across the internet anywhere and everywhere. In this way de-identified cohorts as well as aggregate data can be exchanged for use in clinical research.