SDMX STARTER KIT FOR NATIONAL STATISTICAL AGENCIES

Version: 23 September 2015

DRAFT

FOREWORD

The aim of the *Starter Kit for National Statistical Agencies* (or Starter Kit) is to provide a resource for national statistical agencies in countries contemplating the implementation of the Statistical Data and Metadata Exchange (SDMX) technical standards and content-oriented guidelines for the exchange and dissemination of aggregate data and their related methodological information (metadata). It outlines a structured process for the implementation of SDMX by agencies that have little knowledge about SDMX and how they would go about the implementation of the standards and guidelines. In order to avoid duplicating the work of the SDMX sponsoring agencies the Starter Kit makes extensive use of links to existing SDMX background documents and artefacts that have been developed at the global and national levels.

The Starter Kit was developed initially under the auspices of the joint 2014 Asian Development Bank (ADB) – United Nations Economic and Social Commission for Asia and the Pacific (UN ESCAP) SDMX initiative to improve the efficiency of data and metadata exchange between national statistical agencies in the Asia and Pacific region and international organizations through the ongoing use of SDMX standards. The initiative also aimed at enabling national agencies to determine which, of a range of available tools for SDMX implementation, best meets the needs of the organisation. Finally, it sought to identify national skill capacity development needs in the context of such implementation. However, almost all of the content of the current document is relevant for countries in all regions around the globe.

The basic premise of the Starter Kit is that SDMX implementation must be seen in the context of a wide range of corporate institutional, infrastructure and statistical initiatives currently underway in almost all statistical agencies around the globe to improve the quality and relevance of the service they provide to government and non-government users of their outputs. Such corporate initiatives are often undertaken within the framework of a national strategy for the development of statistics (NSDS), statistical master plans or similar strategic planning process. A number of corporate initiatives relevant to SDMX are outlined in the *Implementation Plan for the Regional Programme for the Improvement of Economic Statistics in Asia and the Pacific* endorsed by the ESCAP Committee on Statistics in December 2012. Planned improvements from the Implementation Plan covered areas such as advocacy, coordination, statistical infrastructures, and skills.

A key feature of the Starter Kit is that careful consideration of a range of institutional, statistical and skills related issues needs to be undertaken well before detailed planning of IT-related issues related to the implementation phase of the SDMX technical standards takes place. Key decisions that need to be include careful thought about include identification of improvement in the efficiencies in data and metadata reporting and dissemination / reporting that could be gained through SDMX implementation as well as the statistical domains that will be included in such implementation. The outcomes of decisions in these areas will help ensure that appropriate choices are made with respect to the selection of appropriate data structure definitions, data and metadata exchange scenarios and selection of implementation tools best suited to the needs of the national organisation. Statistical domain experts must play a key role in these decisions.

SDMX implementation entails considerable use of resources, expertise and effort at the national level and these will only be forthcoming if a substantive business case of the potential benefits can be made for SDMX implementation, particularly in the context of both broader national strategic goals and alternative technical standards. The Starter Kit presents the process for developing the business case for SDMX implementation from the perspectives of both international and national statistical agencies. The intended audience are senior management in national agencies as well as statistical subject matter experts.

The final key theme throughout the Starter Kit is that SDMX implementation by national statistical agencies is not just an IT issue, but one that requires close collaboration between different parts of the

organisation, in particular, management and statisticians working in the domains being implemented. The intended audience for the Starter Kit within national statistical agencies therefore includes, national statistical agency: management; statistical domain specialists; coordination units; data dissemination units; and IT specialists.

The current version of the Starter Kit is still work in progress and feedback and suggested improvements from other international organisations and national statistical agencies are welcomed. The document was prepared by Denis Ward, Independent Statistical Consultant and incorporates extensive comments and suggestions from the European Commission's Eurostat, the Organisation for Economic Cooperation and Development (OECD), and the United Nations Statistics Division (UNSD). Further input was obtained from the International Labour Organisation (ILO), the International Monetary Fund (IMF), Metadata Technology, the United Nations Educational, Scientific and Cultural Organisation (UNESCO) and the World Bank.

Denis Ward
Melbourne, Australia
September 2015

Email: teedward@gmail.com

CONTENTS

ABBREVIATIONS AND ACRONYMS

ABS	Australian Bureau of Statistics
ADB	Asian Development Bank
AfDB	African Development Bank
API	Application Programming Interface
ASEAN	Association of South-East Asian Nations
BIS	Bank for International Settlements
BoP	Balance of Payments
CDC	Cross-Domain Concept
CES	Conference of European Statisticians
COG	Content-Oriented Guidelines
COMSEC	Commonwealth Secretariat
CSPA	Common Statistical Production Architecture
CSV	Comma-separated values
DDI	Data Documentation Initiative
DFID (UK)	Department for International Development (United Kingdom)
DRS	Debtor Reporting System
DSD	Data Structure Definition
DQAF	Data Quality Assessment Framework
ECB	European Central Bank
EMEAP	Executives Meeting of East Asia Pacific Central Bank
ESMS	Euro-SDMX Metadata Structure
ESS	European Statistical System
ESS MH	ESS Metadata Handler
ESQRS	ESS Standard for Quality Reports Structure
EU	European Union
FDI	Foreign Direct Investment
GDDS	General Data Dissemination Standard
GESMES	Generic Statistical Message
GESMES/TS	GESMES for Time Series
GFS	Government Finance Statistics
GSBPM	General Statistics Business Processing Model
GSIM	General Statistical Information Model
GUI	Graphical User Interface
HLG-BAS	High Level Group for Strategic Directions in Business Architecture in Business Statistics
HTML	HyperText Markup Language
ICT	Information and Communications Technology
ILO	International Labour Organisation
IMF	International Monetary Fund
INEGI	Instituto Nacional de Estadistica y Geografia (Mexico)
I/O	International organisation
ISO	International Organisation for Standardisation
ISTAT	Instituto Nazionale di Statistica (Italy)
IT	Information technology
KODAPS	Korean Data Provision System
KSIS	Korean Statistical Information Service
MCV	Metadata Common Vocabulary
MDGs	Millennium Development Goals
MSD	Metadata Structure Definition
MSIS	Management of Statistical Information Systems (Meeting of)
NSDP	National Summary Data Page

NSDS	National Strategy for the Development of Statistics
NSO	National statistical organisation
NSS	National statistical system
ODP	Open Data Platform
OECD	Organisation for Economic Cooperation and Development
PDF	Portable document format
R&D	Research and development
SAARC	South Asian Association for Regional Cooperation
SDDS	Special Data Dissemination Standard
SDMX	Statistical Data and Metadata Exchange
SDMX-RI	SDMX Reference Infrastructure
SEACEN	South-East Asian Central Banks
SIMS	Single Integrated Metadata Structure
SNA	System for National Accounts
SOAP	Simple Object Access Protocol
SPC	Secretariat of the Pacific Community
STES	Short-term Economic Statistics
UNCTAD	United Nations Conference on Trade and Development
UNDG	United Nations Development Group
UNECLAC	United Nations Economic Commission for Natin America and the Caribbean
UNESCAP	United Nations Economic and Social Commission for Asia and the Pacific
UNESCO	United Nations Education, Scientific and Cultural Organisation
UNFAO	United Nations Food and Agricultural Organisation
UNSD	United Nations Statistics Division
XML	Extensible Markup Language

A. OBJECTIVES OF STARTER KIT

The Starter Kit outlined in the current document is a resource for national statistical agencies in countries contemplating the implementation of the Statistical Data and Metadata Exchange (SDMX) technical standards and content-oriented guidelines for the exchange and / or the dissemination of aggregate data and their related methodological information (metadata).

The Starter Kit was initially developed under the auspices of the joint 2014 Asian Development Bank (ADB) – United Nations Economic and Social Commission for Asia and the Pacific (UN ESCAP) SDMX initiative to improve the efficiency of data and metadata exchange between national statistical agencies in the Asia and Pacific region and international organizations, through the ongoing use of SDMX standards and guidelines[1]. These standards are currently used by only a relatively small number of national statistical agencies in the region. The aim of the joint initiative was to build the capacity of countries in the Asia and Pacific region to apply SDMX standards and guidelines. Whilst the initial impetus for the Starter Kit was to assist countries in the Asia and Pacific region, almost all of the content of the Kit is relevant for countries in all regions around the globe.

Following a discussion of issues relating to the development of a business case for implementing SDMX (in Section B), the Starter Kit (in Section C) outlines a number of practical steps, processes and activities to be undertaken by national agencies, working with partner countries and international organisations, prior to and during SDMX implementation. Such steps entail:

Step	Activity	Responsibility
I	Acquisition of a basic understanding of key SDMX artefacts by national agencies in both developed and developing countries. As will be further elaborated in Step 3 below, one of the most effective ways of learning about SDMX and its benefits is to implement a small scale pilot study with limited scope and timeframe on a statistical domain using one of the available global Data Structure Definitions (DSDs) and suites of implementation tools.	National agencies
II	Consideration of a range of issues by countries prior to embarking on the implementation of SDMX technical standards and content-oriented guidelines (COG), in particular, the mapping of key national concepts to those outlined in the COGs. These issues were identified through discussions with some of the SDMX sponsoring agencies (IMF, Eurostat, BIS, ECB, UNSD, World Bank and OECD) and with statistical agencies in several countries in the Asia and Pacific region, in particular, those involved in an Asian Development Bank (ADB) - United Nations Economic and Social Commission for Asia and the Pacific (UN ESCAP) pilot project undertaken in 2014. Some of the information derived from the analysis of these issues could be used in the preparation of the business case for SDMX implementation described in Section B of the current document.	National agencies

[1] For further information on the joint ADB – ESCAP initiative (objectives, pilot project, project focus, etc) refer paper prepared for MSIS meetings in Manila and Dublin on 14-16 April 2014 at. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2014/Topic_2_ADB_ESCAP.pdf

III	Identification of national SDMX capacity development needs (IT, statistical, institutional) and capacity development delivery.	National agencies with assistance from partner countries and international organisations as required
IV	Determining which set(s) of implementation tools to use, based on national needs, level of development, etc.	National agencies with assistance from international organisations as required
V	Participation in SDMX networks linked to existing global networks.	National agencies, partner countries, international organisations, "official" SDMX working groups that any statistical agency may apply to join

The Starter Kit complements the *Checklist for the Implementation of SDMX in a Statistical Domain* and the *Modelling a Statistical Domain for Data Exchange in SDMX* documents developed by the SDMX working groups[2].

The underlying premise of the current document is that SDMX implementation must be seen in the context of a wide range of corporate institutional, infrastructure and statistical initiatives currently underway in almost all statistical agencies around the globe to improve the quality and relevance of the service they provide to government and non-government users of their outputs. Such corporate initiatives are often undertaken within the framework of a national strategy for the development of statistics[3] (NSDS), statistical master plans or similar strategic planning process. These initiatives frequently include the modernisation of IT infrastructures.

However, SDMX implementation by national statistical agencies is not just an IT issue but one that requires close collaboration between different parts of the organisation, in particular, management and statisticians working in the domains being implemented. The intended broad national audience for the Starter Kit therefore includes, national statistical agency: management; statistical domain specialists; coordination units; data dissemination units; and IT specialists. In order to minimise duplication of effort the Starter Kit provides links to existing relevant SDMX documentation and resources developed and disseminated by the seven SDMX sponsoring agencies[4] on the SDMX.org website.

[2] Refer SDMX.org, Standards and Guidelines [Checklist document not yet publicly available]
[3] Refer Paris21 NSDS website at http://paris21.org/NSDS
[4] These comprise the: Bank for International Settlements (BIS); European Central Bank (ECB); Eurostat; International Monetary Fund (IMF); Organisation for Economic Cooperation and Development (OECD); United Nations Statistical Division (UNSD); and the World Bank.

B. BUSINESS CASE FOR IMPLEMENTING SDMX STANDARDS AND GUIDELINES

This Section of the Starter Kit presents key elements of the business case that national agencies need to develop prior to SDMX implementation. The presentation of a solid business case to senior management is a prerequisite for the successful implementation of SDMX within an organisation. The intended audience for this Section of the current document are staff in national agencies responsible for the initial preparation of the business case. This often entails the collaboration between a number of units within a national statistical agency such as those responsible for IT, data dissemination, coordination as well as statistical subject matter experts.

Whilst the case for implementing SDMX to facilitate data exchange at national and international levels has been elaborated in considerable detail at various Expert Meetings and Global Conferences over the last 10 years, further discussion of the benefits for national agencies is essential. SDMX implementation requires considerable use of resources, expertise and effort at the national level and these will only be forthcoming if a substantive business case of the potential benefits can be made for SDMX implementation, particularly in the context of both broader national strategic goals and alternative dissemination standards. In other words, SDMX implementation by national agencies must be considered within the light of a broader range of strategic, corporate, institutional and statistical initiatives such as the modernisation of statistical processing systems, currently underway in most statistical agencies around the globe.

Many of these corporate initiatives are outlined in the *Implementation Plan for the Regional Programme for the Improvement of Economic Statistics in Asia and the Pacific*[5] endorsed by the ESCAP Committee on Statistics in December 2012. Planned improvements from the Implementation Plan covered a number of strategic corporate issues such as improvements in statistical advocacy, coordination, infrastructures, and skills. Such issues are included in the corporate plans of most national statistical organisations around the world.

1. The challenge

The dissemination of data to users is taking place in an increasingly more heterogeneous dissemination environment where there is less focus on paper publications. Almost all agencies are in the process of modernising their statistical processing and dissemination infrastructures to improve their efficiency and to make use of emerging technologies such as the internet and personal hand held devices, mobile phones, tablets, etc., for the dissemination of the most up-to-date statistical outputs (in tabular and graphic forms of presentation) to users. The ultimate objective is to improve the relevance, timeliness and other elements of quality of the statistical outputs they disseminate to government and non-government users.

Many statistical agencies are taking a more integrated approach to addressing this challenge by attempting to realign their business model away from being mere producers of data, by changing both the products they produce and the processes used to produce these products. This strategic vision of the High Level Group for Strategic Directions in Business Architecture in Business Statistics (HLG-BAS) was endorsed by the Conference of European Statisticians (CES) in June 2011. The realignment process is embedded in many of the strategic planning processes currently being implemented by national statistical agencies around the globe – refer 3 below.

[5] Refer http://www.unescap.org/sites/default/files/CST3-CRP1E.pdf

2. High level rationale for utilising standards such as SDMX and Data Documentation Initiative (DDI)

The document, *Business case for harnessing DDI and SDMX as standards for expressing statistical information in support of production of official statistics*[6], provides a high level rationale for achieving this strategic vision through utilising standards such as SDMX and the Data Documentation Initiative (DDI) to support the production of official statistics. The document uses the analogy of the construction industry where every component used is standardised in some way to produce a very heterogeneous range of outputs. The analogy refers to the "industrialisation and standardisation[7]" of processes used to produce official statistics. The focus of such standardisation underway in many national statistical agencies is in regard to processes within their own agency to eliminate domain-specific systems that are not reusable or extendable. Such standardisation of processes is sometimes within the context of industrialisation and standardisation taking place internationally.

The business case document then articulates a set of essential and highly desirable characteristics of an industry standard for representing statistical information and then evaluates SDMX and DDI within the context of those standards. The main essential characteristics entail whether the standard is:

- comprehensive and coherent (e.g. can relate to both micro and aggregate data);
- human interpretable and machine actionable;
- readily rendered, on a consistent basis, into different forms; and
- readily extensible to support local requirements.

The highly desirable characteristics are whether the standard harnesses existing standards and capabilities (e.g. tools and applications) developed to support existing standards.

3. Specific cases for realignment of business objectives and processes: Analysis of existing national strategic plans

More specific cases for the realignment of business objectives and processes by national statistical organisations around the world are illustrated in many of the NDSDs and strategic plans located in the Knowledge Base on the Paris21 website[8]. The development by UN ESCAP of the *Regional Programme for the Improvement of Economic Statistics in Asia and the Pacific* in 2010 and the subsequent *Implementation Plan* for this programme in 2012 illustrates the issues and areas of realignment faced by national agencies in all regions.

An analysis of NSDSs was undertaken by UN ESCAP in the development of the *Regional Programme* and the subsequent *Implementation Plan* using the Paris21 Knowledge Base, supplemented with information obtained from national statistical office websites. Additional information was obtained by means of questionnaires sent to countries in the region and at several meetings and workshops of national statistical agencies across the region and international organisations convened in Bangkok.

This work identified four main common areas of need for improvement and, implementation strategies across Asia and the Pacific region by the year 2020. These areas are also very relevant in

[6] Refer http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=63930579

[7] More specifically, the processes involved in the "industrialisation" of statistics are the: use of common tools; common processes; and common methodologies. Industrialisation recognises that all statistics are produced in a similar way and that no statistical domain is special. Finally, the adoption of these common elements provides increased flexibility to adapt to new sources and to produce new outputs [Vale, Steven, UN ECE, March 2013, *The Data Deluge: What Does it Mean for Official Statistics* – refer footnote 9 below]

[8] Refer http://www.paris21.org/Knowledge-Base

the development of a national business case for SDMX implementation. To avoid the replication of this earlier work only the main themes are outlined and discussed below. These are:

Advocacy	Increase political support and the appreciation of the importance of investing in the improvement of national capacity to produce a core set of economic statistics.
Coordination	Improve coordination of statistical activities across the national statistical system (NSS) and among development partners.
Infrastructure	Improve national statistical infrastructure for the production and dissemination of a core set of economic indictors (infrastructure specifically mentioned in the Implementation Plan includes: business registers; quality assessment frameworks; metadata repositories; and the adoption of modern statistical data editing techniques).
Skills	Improve human resource skills of NSS staff essential for the collection, compilation and dissemination of the core set of economic statistics.

The more specific relevance of each of these themes to the development of a business case for SDMX implementation is summarised below in the following table.

Table 1: Relevance of SDMX and related standards to NSS priority needs

Theme	SDMX relevance
Advocacy	National statistical agencies in many developing countries report that ongoing collection, compilation and dissemination of statistics, especially major surveys and censuses, is heavily dependent on external technical assistance / resources. There is therefore a need to improve advocacy within the national statistical systems of those countries to ensure the provision of adequate funding by government and the long-term sustainability of the statistical system. Such funding is only likely to be forthcoming if statistical agencies can demonstrate the relevance of their outputs to the major problems / issues facing the country. There is a need to reduce the emphasis on paper publications and to focus on the release of information to a range of users with different levels of expertise via online databases that are updated in real-time and accessible across the NSS. Agencies need to tailor data to the different needs / requirements / requests of individual users – from analysts, media (timeliness issue), to the general public. In this context statistical agencies need tools that are flexible and can produce a range of outputs such as tables, graphical presentation on a range of different access platforms such as hand held media. Agencies need to ensure that online databases are not only replicas of paper publications. Finally, many NSDSs identified the need to develop new statistical outputs including small area data and micro-data whilst at the same time ensuring the confidentiality of data providers. There is a need to implement tools to efficiently ensure confidentiality. Related to this is a requirement for a policy / strategy for micro-data.
Coordination	Almost all national statistical organisations reported the need to improve coordination which is seen as a key issue, both within the organisation and across the national statistical system. The latter is particularly important in decentralised NSSs, with respect to both the production and the dissemination of statistics. Producer agencies and users report the problem of multiple data sources that are

Theme	SDMX relevance
	inconsistent and lack timeliness. There is also the need to minimise the risk of different agencies disseminating similar or identical statistical series that are different. In NSDSs this need is often stated in terms of the need to firm up the coordinating role of the NSO through changes in legislation, etc. There is also often an absence of guidelines for setting national standards across the statistical collection / compilation / dissemination cycle. Whilst SDMX can provide the technical tools to alleviate these problems there is a need first for agencies within the NSS to reach agreement on key issues such as responsibilities, organisation of data flows and on the overall systems architecture that best suits their needs. SDMX and related standards could be used to rationalise and improve the data flows between data producers and disseminators within the NSS – refer Step II below.
Infrastructure	All NSOs report the need to improve statistical infrastructures including registers, metadata repositories, databases, data dissemination tools and facilities. Many stated the need to identify and adopt international statistical standards across the NSS, in particular, classifications. Classifications need to be modified to ensure relevance for national environment but at the same time maintaining international comparability of key aggregates. Statistical integration necessitates the use of standard concepts and code lists for different statistical domain collections both within NSO and across the NSS – elimination of domain-specific systems that are not extendable or reusable. NSOs are often seen as playing a leading role in this. Many countries mentioned the need to improve methodological transparency to both national and international users through the ongoing provision of structured metadata that is closely linked to data.
Skills	All NSOs highlighted the importance of developing the skill capacity of staff across the NSS to undertake the above initiatives. The absence of such skills is seen as a major barrier to change and improvement. For SDMX implementation there is a need for staff in IT, dissemination and statistical domains to acquire an initial basic and later an in depth understanding of the standard and related guidelines and the various implementation tools for development and ongoing maintenance. The required skills are discussed further in Step III.

SDMX and related standards should therefore be seen as tools that could be utilised by national statistical agencies to address these issues, but only following the development of a business case that reviews SDMX as well as other dissemination standards such as DDI. SDMX can only work effectively if these other issues are addressed as well. These aspects are discussed is greater detail below in Step II of the structured SDMX implementation process - Consideration of a range of issues by countries prior to embarking on SDMX implementation.

4. Elements of a business case for an SDMX implementation

The following is a list of potential arguments that could be included in a business case to justify the implementation of SDMX by a national agency. These arguments need to be considered in the context of broader corporate strategic objectives such as those outlined above. Obviously, the institutional environment, etc., of individual organisations vary considerably and the issues presented below should be viewed more as a potential "shopping list" which would need to be tailored to the needs and circumstances of a specific agency in the preparation of the business case. The two main elements of such a business case comprise a summary of the: problems faced; and how SDMX can help[9].

[9] Source: *Checklist for the Implementation of SDMX in a Statistical Domain*, SDMX, 2014

a. Problems faced: analysis of current situation

- Data are collected in various ways (surveys, web queries, data files, metadata text, emails, etc.).
- Data collection is a labour intensive, manual operation.
- Possibility of errors and inconsistencies.
- Statistics production needs to turn into an "industry" (through extended use of common and standardised processes, transforming raw data into statistical products according to generic and commonly accepted information concept). Each statistical organisation is a factory of statistical information. Together, these statistical organisations form the "official information" industry. As with any established industry, the production of official statistical information should have its own industrial standards.

There is vastly increasing volumes of data now available (e.g. transactional data from the internet and other sources as an element of the "data deluge")[10].

- This environment:
    - limit the case for, and public acceptance of, the expensive and intrusive traditional approach to collecting data via surveys;
    - creates many new opportunities for "value adding" to this raw data through statistical processing and creation of new statistical outputs (other organisations may move to this role if producers of official statistics do not);
    - requires an organisation which is more “agile” in regard to sourcing data, processing data and producing new statistical outputs. In most cases the required business transformation cannot be supported in a sustainable manner by existing statistical production processes and the IT infrastructure underlying existing processes.
- Data collection is on the increase as a result of the introduction of new activities and initiatives at both the statistical corporate and broader government levels.
- *[For decentralised statistical systems*] Many organisations collect and disseminate similar or the same data. This poses an increasing burden for data providers and increases the risk of the dissemination of inconsistent data across the national statistical system.

b. How SDMX can help

- SDMX provides:
    - standardisation endorsed by ISO 17369:2013;
    - a common language for the exchange of statistical data and metadata;
    - a common model for statistical data in an environment where there is a wide range of different data models and transmission formats used for the exchange of data and metadata;
    - a standardised way of organising and exchanging data, enabling interoperable implementations within and between systems concerned with the exchange, reporting and dissemination of statistical data and related metadata;
    - a proven governance model to manage the life-cycle of a domain’s SDMX artefacts;

[10] Steven Vale (UN ECE) in his March 2013 presentation, *The Data Deluge: What Does it Mean for Official Statistics* attempts to quantify the “data deluge” and outlines the implications for official statistics in terms of changing roles / functions etc – refer http://www1.unece.org/stat/platform/pages/viewpage.action?pageId=77170614

- guidelines on how to represent data and metadata, e.g. the creation of data structures, an SDMX project approach checklist;
- standard code lists and concepts to help harmonize the exchange of data;
- a standard for building SDMX registries, therefore enabling tools to easily connect to multiple registries and allowing federation of content;
- a standard programming model for retrieving data from web services.

- SDMX facilitates the development of standard tools that can be used by a wide range of national and international organisations, such the SDMX Reference Infrastructure (SDMX-RI) developed by Eurostat, the IMF's SDDS Plus, SDMX Converter and DevInfo – refer Step I.3 in Section C below.

- SDMX is an industry standard (ISO 17369: 2013) and as such can help ensure that separate practical implementations (e.g. IT modules/applications supporting statistical production) are harmonised in terms of inputs and outputs, can interoperate with each other and could successfully be developed collaboratively and then shared among agencies.

- Applying industry standards helps translate, on a consistent basis, shared concepts into standardised practical implementations.

- SDMX can reduce reporting burden:

  - the reporter can automate publication of SDMX via a web server;
  - the data message content can be pre-validated by SDMX data structures;
  - data can be queried (or "pulled") by collecting agencies. This avoids the creation of data packages by the reporter for each data flow (the "push" method);
  - thanks to data sharing between international organisations, data will be exchanged only with one international organisation (IO) instead of bilateral relations with each IO involving double or triple work on both ends;
  - data sharing will also help to improve the quality and consistency of official statistics disseminated by the IOs involved (dissemination of consistent data sets and area aggregates across IOs).

- SDMX helps the standardisation of statistical data and metadata:

  - use of Cross Domain Concepts and Content-Oriented Guidelines;
  - shared artefacts can be accessed from a registry;
  - once defined DSDs and MSDs can be used throughout an organisation (and across agencies)[11].

- SDMX helps reduce development and maintenance costs:

  - shared toolbox of SDMX software (i.e. SDMX Reader, Hub, etc.);
  - open source approach;
  - no licensing costs.

- SDMX helps to reduce data errors:

[11] The document *Modelling a Statistical Domain for Data Exchange in SDMX*, published on the SDMX.org website in March outlines recommendations on how to determine the number of DSDs for a subject-matter domain. It follows the approach that a decision on the number of DSDs needed for describing a statistical domain should come after the discussion on the data model and the architectural approach to be followed in the data collection exercise.

- through the use of agreed structures for transmissions;
- much time saved in data conversion/mapping/validation. A standard mapping structure is part of SDMX.

- SDMX provides a faster and more reliable data and metadata sharing:
  - processes can be automated;
  - easier to validate content.

- SDMX helps to improve the harmonisation of statistical business processes:
  - Can be used throughout the Statistical Information System (data production, data exchange, data dissemination)

- SDMX helps standardising IT applications and infrastructure
  - SDMX functionality systematically built into statistical applications.

- International community can share experiences and software:
  - sharing and reuse of software and know-how among organisations;
  - open source culture;
  - staff exchanges, secondments

- SDMX and its sponsoring organisations provide capacity building and training.

Some of the information derived from the consideration of a range of issues (institutional, statistical, IT, skills-related) prior to embarking on the implementation of SDMX technical standards and content-oriented guidelines (COGs), refer Section C, Step II below, could be used in the preparation of the business case for SDMX implementation, including the justification for the development of a small scale pilot project.

5. Objectives of existing SDMX implementations

The initial motivation for the development of the SDMX initiative came from international organisations to improve the efficiency of collecting aggregate data and metadata from their member countries. This arose from a need to minimise the reporting burden placed on national agencies which often flowed from duplication of effort to meet frequently overlapping requests from different international agencies.

As will be demonstrated below in Annex 1 (Examples of international agency and national agency SDMX implementations), there are a wide range of different SDMX implementations by agencies at the national level which have been developed to meet different priorities and needs. These range from:

- improving the efficiency of aggregated data exchange to international organisations (e.g. KODAPS implementation by the Korean Statistical Information Service);
- extending these improvements to encompass the dissemination of more disaggregated data to all users, including national users as well as data flows between other agencies within the national statistical system (NSS);
- integration with other standards such as DDI to improve end-to-end statistical processes including data collection, storage, and dissemination (e.g. by the Australian Bureau of Statistics (ABS)).

More specifically, SDMX standards and guidelines may be used by national agencies:

- to collect highly disaggregated data from national data providers and / or to improve the efficiency of data exchange between different agencies within the NSS. Legacy data flow systems within individual NSSs often duplicate each other, use different types of exchange (emailed Excel files, web-linked Excel files, PDFs, etc) and manual systems. These may result in multiple data sources for the same series within the NSS with the potential for different series / timeframes from each source. The use of SDMX standards and guidelines enables automation of the data exchange process between regular users / data providers and the possibility of real-time updates; and

- for subsequent data dissemination and reporting, for example, to a number of international organisations, and / or national users. Most countries currently meet their obligations to provide data (and metadata) to a large number of global and regional international agencies[12] via multiple and resource consuming data flows. Use of SDMX standards enables the development of an NSS hub that is updated automatically, from where international agencies can obtain the latest data using either push or pull-modes;

Agencies need to decide whether to implement SDMX for one or both of these potential uses.

[12] For example, countries in South-East Asia provide economic statistics data to the: ASEAN Secretariat, Asian Development Bank (ADB); Bank for International Settlements (BIS); Executives' Meeting of East Asia-Pacific Central Banks (EMEAP); International Monetary Fund (IMF); Organisation for Economic Cooperation and Development (OECD); South East Asian Central Banks (SEACEN); United Nations Statistical Division (UNSD), United Nations Conference on Trade and Development (UNCTAD); World Bank.

C. STRUCTURED PROCESS FOR SDMX IMPLEMENTATION BY NATIONAL AGENCIES

This Section of the Starter Kit describes the five practical implementation steps described in Section A that could be used by national statistical agencies contemplating the implementation of the SDMX standards and guidelines. As mentioned in Section A above, these steps comprise:

I. Acquiring a basic understanding of key SDMX artefacts
II. Addressing issues to be considered by countries prior to embarking on SDMX implementation
III. Identification of national SDMX capacity development needs and capacity development delivery
IV. Determine which set of SDMX implementation tools to use
V. Participation in SDMX networks

The information provided below within each of these steps complements that provided in the *Checklist for the Implementation of SDMX in a Statistical Domain*[13] and *Modelling a Statistical Domain for Data Exchange in SDMX* documents developed by the SDMX Statistical and Technical Working Groups in 2014 and 2015. A comparison of the implementation steps outlined in the Starter Kit and the two other documents referred to previously is provided in the following diagram.

The Checklist document is aimed at national statistical agencies new to SDMX and describes the steps to be addressed when implementing an SDMX project for a statistical domain. It includes the general aspects of these issues and refers users to more detailed technical documents for more information. The starting point is acquiring at least a basic understanding of SDMX artefacts and considering key issues well before SDMX implementation commences. The modelling document on the other hand is a slightly more technical discussion of the general principles on how to design and create SDMX artefacts in a statistical domain, followed by a step-by-step approach based on the SDMX information model. Finally, the checklist document outlines general project management steps to be taken expressed in the specific context SDMX to properly implement the standards and guidelines in a statistical domain.

Diagram 1: Comparison between the higher level implementation steps outlined in the Starter Kit, SDMX Modelling and SDMX Checklist documents

Starter Kit	Modelling	Checklist
I. Acquiring basic understanding of SDMX	4.1 Agree on exchange needs	1. Initiate
II. Addressing issues before implementation	4.2 Define the concept scheme	2. Plan and organise
III. Identification of capacity development needs	4.3 Code the concept scheme	3. Design
IV. Which implementation tools to use	4.4 Define a DSD matrix	4. Develop and test
V. Participation in SDMX networks	4.5 Optimise the DSD matrix	5. Implement
	4.6 Create SDMX artefacts	6. Evaluate

[13] Refer SDMX 2014, *Checklist for the Implementation of SDMX in a Statistical Domain*, p. 7 [Not yet publicly available]

STEP I: ACQUIRE A BASIC UNDERSTANDING OF KEY SDMX ARTEFACTS

A key prerequisite for national agencies is to gain a basic understanding of the key SDMX artefacts briefly summarised below. These comprise:

1. SDMX technical standards and content-oriented guidelines (COGs)
2. Data Structure Definitions (DSDs) and Metadata Structure Definitions (MSDs)
3. SDMX implementation tools

The Starter Kit does not attempt to describe these artefacts in any detail but provides links to where further readily available information may be accessed. Further discussion on how these artefacts relate to the SDMX Information Model is provided in Section C, Step III below.

Rather than starting from a zero understanding of SDMX, it would be efficient if staff in national agencies involved in initial discussions with SDMX experts in international organisations or national agencies in other countries acquired beforehand at least a basic knowledge of these artefacts. In addition, it would be useful for these staff to have an understanding of multi-dimensional modelling, and the different types of metadata. Such knowledge also provides the context for subsequent steps outlined below. Experience to date with national implementation of SDMX standards is that a high degree of understanding of SDMX artefacts, and proficiency in the use of available SDMX tools is essential.

Once an organisation has acquired a basic understanding of both SDMX standards and the available SDMX tools through a review of the literature referenced below, the most effective way of gaining more in depth knowledge, and to demonstrate the benefits implementation can bring, is to implement a small scale pilot study. This may involve using either one of the existing global DSDs listed in Annex 2 below or less complex ECOFIN / MDG DSDs, and available suites of implementation tools described below[14]. Such a pilot project should have a clear set of objectives and be limited in scope (in terms of statistical domain coverage) and timeframe. Such a pilot could also provide insights into the issues to be considered by countries prior to full-scale SDMX implementation – refer Step II below.

Issues to be considered and practical advice for countries mounting a pilot project are also outlined in Step II.

1. SDMX STANDARDS AND GUIDELINES

The SDMX standards and guidelines that have been developed by the seven SDMX sponsoring agencies include[15]:

- SDMX Technical Standard 2.1 issued in 2011 and consolidated in 2013. This was published by the International Organization for Standardization (ISO) as International Standard (IS) 17369 in January 2013. This ensures that SDMX technical standards build on other recognized standards and provide the basis for interoperability with them. The components of the current SDMX Technical Standard are:

    - Section 1 – Framework
    - Section 2 – Information Model

[14] As mentioned below in the discussion on DSDs, global DSDs are quite complex (for a number of reasons (including the large number of dimensions that require mapping) and may not be a good choice for a pilot study. MDGs and ECOFIN DSDs may be a more appropriate choice for an initial pilot study as they are much easier to understand.

[15] Source: SDMX, 2014, *SDMX Content-Oriented Guidelines; Part 1 – Introduction*, p.5

- Sections 3A and 3B – SDMX_ML
- Section 4 – SDMX-EDI
- Section 5 – Registry Specification
- Section 6 – Technical Notes
- Section 7 – Web Services Guidelines

- Statistical guidelines covering: harmonisation of concepts and terminology (Content-Oriented Guidelines), and the development of guidelines for the creation and maintenance of SDMX artefacts such as code lists, data structure definitions, etc.

  Content-Oriented Guidelines (COG) include the:

  - Cross domain concepts
  - Cross domain code lists
  - Statistical subject matter domains
  - SDMX Glossary

Available resources

All of the SDMX standards and guidelines are readily available on the SDMX website[16]. Other key readable documents and resources currently available on the SDMX website include:

- *SDMX Content-Oriented Guidelines – Part 1: Introduction* [First released 2009; last revision March 2014]
- *SDMX User Guide* – Provides guidance to Version 2.1 of the SDMX Technical Standards released in 2011.
- SDMX Tutorials developed by Eurostat. A suite of tutorial videos and documentation includes: SDMX in Action; the European Statistical System (ESS) Metadata System; Learning SDMX: Tutorials[17].

Other documents planned over the longer term which will provide suitable background information for countries implementing SDMX standards include Guidelines for the Design of SDMX Metadata Structure Definitions.

2. DATA STRUCTURE DEFINITIONS AND METADATA STRUCTURE DEFINITIONS

Data Structure Definitions (DSDs) and Metadata Structure Definitions (MSDs) describe how datasets and metadata sets are organised, and define the mechanism for referencing those datasets and metadata sets described by structural metadata[18]. DSDs and MSDs are key elements of SDMX standards and if the objective of reducing the reporting burden of national agencies is to be achieved,

[16] Refer http://sdmx.org/ - A succinct description of the COGs and their use are available in the following presentations: http://www.oecd.org/std/47790060.pdf and http://www.oecd.org/std/47790138.pdf

[17] Refer https://webgate.ec.europa.eu/fpfis/mwikis/sdmx/index.php/Main_Page

[18] "Statistical data are described by a set of metadata values taken from specific concepts. These concepts (referred to as structural metadata) act as identifiers and descriptors of the data, e.g. country, time, frequency, scale. This system of concepts identifying and describing the data can be referred to as the named dimensions of a multi-dimensional cube of data. Reference metadata on the other hand are additional explanatory text which explains the concepts used for the data, methods and processes used for the generation of the data and which describe quality aspects of the data." - Eurostat 2010, *SDMX Self-Learning Package, Student Book No. 2 – The SDMX Information Model*, p 4, Eurostat, Luxembourg.

their development should be undertaken cooperatively between international organisations (in cooperation with national agencies).

a. Data Structure Definitions

The SDMX sponsoring agencies and other international organisations have developed DSDs for a limited number of the statistical domains described in the COG's *Statistical Subject Matter Domains*, including national accounts and balance of payments (BoP) domains. These are readily available in the *SDMX Global Registry* which may be accessed via the SDMX website[19]. Listed below in Annex 2 are the DSDs that have been prepared to date or are in the process of being developed [20], together with the name of the maintenance agency[21].

Global, local and shared DSDs

A distinction can be drawn between DSDs that are global in nature, whose intended use extends beyond the corporate requirements of the maintenance agency, and those that that are primarily local or shared.

A DSD is considered as global if it meets one of the two criteria below:

- It is designed as a standard data structure for global use (i.e. having a very wide geographical coverage or cross-domain nature), with more than one SDMX sponsor organisation represented in the ownership group and one of the members of the ownership group acting as maintenance agency on behalf of the ownership group;

- DSDs labelled as "global" by the SDMX sponsors considering the recognised expertise in the domain concerned of one of the organisations represented in the ownership group and the potential usefulness of the artefact for the whole SDMX community. In this case the DSD will have to meet strict criteria of versioning, governance, maintenance, adoption and endorsement.

As mentioned above, examples of global DSDs include those developed for national accounts[22], balance of payments[23] and merchandise trade. One of the key features of Global DSDs is their use of identical concepts, concept mnemonic and concept description for common dimensions.

Local DSDs are DSDs developed for the specific needs of one organisation only. An example is a structure for use in internal production processes. Shared DSDs refer to any DSD used by two or more statistical organisations and which does not meet the criteria outlined above for being a Global DSD.

---

[19] For a more direct link to the *SDMX Global Registry* refer https://registry.sdmx.org/home.html

[20] Refer https://registry.sdmx.org/

[21] The following link explains the creation of DSDs - Explanation of creation: http://www.oecd.org/std/47790070.pdf

[22] The European System of National and Regional Accounts transmission programme and its questionnaires are a well-established collection system for national accounts data, which has been used to transmit data across Europe (and beyond) since the mid-1990s. National accounts data based on ESA 2010 and the new global national accounts DSD has been transmitted by EU Member states and Candidate countries to Eurostat using SDMX standards since September 2014 [*ESA2010 Transmission Programme: What's new as of the second half 2014?* Eurostat – refer: http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.20/2013/ws-Istanbul/Session_2_-_ISI_STS40_ESA_2010_TP.pdf].

[23] An issue that complicates the use of recently developed global DSDs for national accounts and balance of payments was their development using the most recent international standards for both statistical domains and industrial classification, namely SNA 2008, BPM6 and ISIC Rev. 4 respectively. Although the global DSDs may be used in national proof of concept SDMX pilot studies involving the main national accounts and balance of payment aggregates by countries that have not yet implemented the latest international standards, at this stage SDMX may not be used for the dissemination of detailed institutional and sector aggregates. The SDMX sponsoring agencies are currently investigating options that will enable the use of the full utilisation of global DSDs by countries that have not yet implemented the latest international standard for national accounts and balance of payments, though it is unlikely that this will be implemented prior to April 2015.

An example of a shared DSD includes the one developed for the OECD's short-term economic indicators (STEIs). Some of the indicators included in the STEI DSD duplicate those in the two global DSDs referred to above.

Ideally, DSDs should be developed that support the reporting needs of a number of international agencies and offer greater flexibility for data dissemination. This flexibility is often provided through the inclusion of a larger number of dimensions in the DSD which facilitates the dissemination of more disaggregated data, though perhaps at the expense of complicating the mapping process and the need for even greater cooperation between IT and statistical domain specialists[24]. The paper, *Modelling a Statistical Domain for Data Exchange in SDMX* from the SDMX working groups[25] outlines a small number of options regarding the number of DSDs required for a statistical domain being implemented – refer *Available DSD resources* below.

Single domain and multi-domain DSDs

Another distinction in the types of DSDs that are currently available (or being developed) is between those that cover single (though broad in scope) statistical domains outlined in the list of *Statistical subject matter domains* referred to above. Such DSDs comprise those that have been developed for national accounts, balance of payments, foreign direct investment, merchandise trade, education, etc. Multi-domain DSDs on the other hand are intended to promote the exchange / dissemination of data flows encompassing more than one subject domain and have been developed for initiatives such as the IMF's SDDS Plus, the OECD's Short-term Economic Statistics, and UNSD's UN CountryData.

As can be seen in Annex 2, to date most global DSDs have been developed for single statistical domains. However, multi-domain DSDs will become more common in the future as the number of SDMX implementations by national agencies and international organisations expands.

Mapping of national concepts to DSD Dimensions

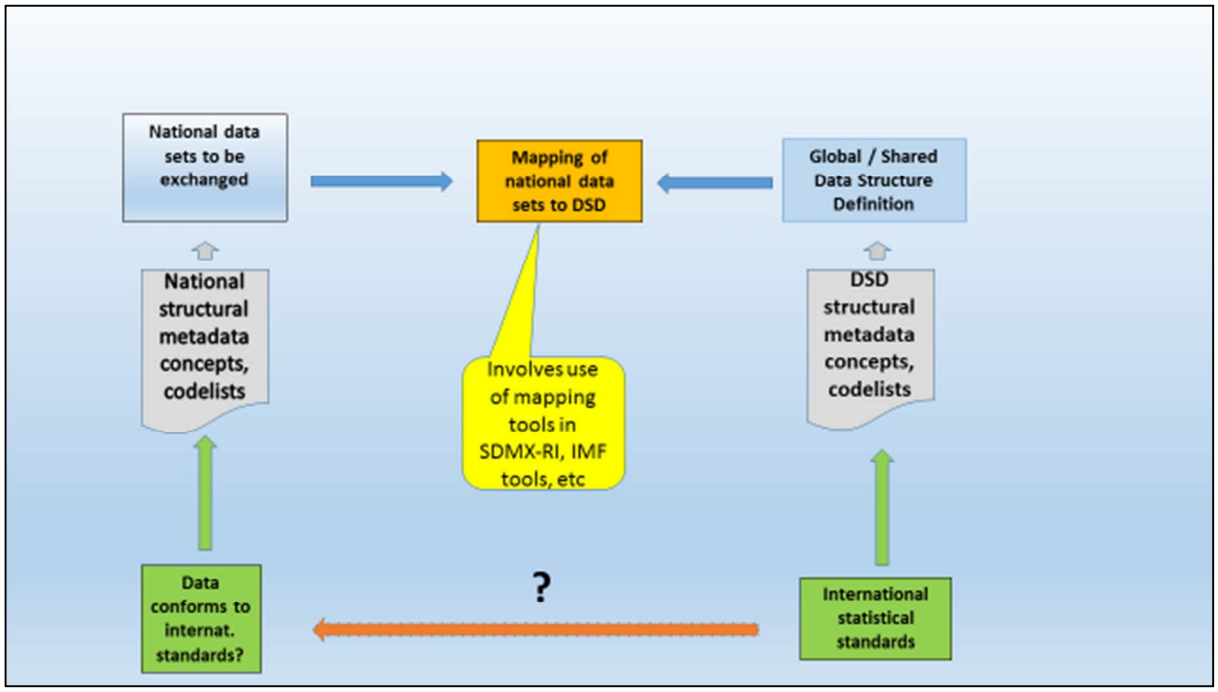
The accurate mapping of national concepts to the DSD dimensions is an essential prerequisite for successful SDMX implementation. Even using the mapping tools available in SDMX implementation developed by the IMF and Eurostat, this process is perhaps the most labour intensive aspect of implementation and one that requires the close cooperation between IT specialists within the agency with their detailed knowledge of the database environment and the statistical domain subject matter experts with their understanding of the content / concepts of the structural metadata embedded in the DSD dimensions. Subject matter specialists can determine whether a national concept is sufficiently similar for it to be mapped to the DSD dimension. Such knowledge is particularly important in the mapping of code lists where there may not be a one-to-one concordance between the DSD code list and the national code list. As will be elaborated further, the mapping process that utilises mapping tools available in SDMX implementation software such as SDMX-RI is not always a straight forward mechanical process. However, the SDMX information model includes a mapping structure for this purpose called the "Structure Set" which can be stored in SDMX registries and reused.

[24] Though the large number of domains in global DSDs facilitates their use for the dissemination of detailed statistical aggregates, it is not necessary for all dimensions to be used in a national implementation. The *Guidelines for the Design of SDMX Data Structure Definitions* published in the SDMX website allows the selection of a subset of DSD dimensions – refer Section 4.2, Number and Relations of DSDs and Section 6.2, Defining Modified DSDs.
DSD guidelines (p. 18) allow for the use of multiple, interconnected or satellite DSDs and / or the specification of "constraints at the data flow definition or data provision agreement level. In this instance DSDs are not even needed, they are "virtual" in this case. The different data flow definitions and/or data provision agreements all refer to the same master DSD but with different sets of constraints. Another possibility is the definition of satellite DSDs that all refer to the same master concept scheme and master code lists but differ in terms of constraints."

[25] Expected to be made available in the first quarter of 2015.

The mapping of national data and metadata structures to DSDs is illustrated in the following overview diagram that illustrates the key objects and mapping processes.

Diagram 2: Overview of mapping processes and objects



Available DSD resources

- *Guidelines for the Design of SDMX Data Structure Definitions* [Version 1 published June 2013]
- *Governance of Commonly Used SDMX Metadata Artefacts* [Version 1.2 published March 2014]. Outlines governance of developing and maintaining DSDs, MSDs, code lists, etc.
- *Mapping of SDMX Cross-Domain Concepts* to metadata frameworks at international organisations [January 2009]. Mapping of SDMX concepts has been undertaken by a number of the SDMX sponsoring agencies namely in the context of: IMF-Data Quality Assessment Framework; and Eurostat-SDMX Metadata Structure – see below.
- *Guidelines for the Creation and Management of SDMX Code Lists* [Version 1 December 2013, revised version expected early 2015]
- *Modelling a Statistical Domain for Data Exchange in SDMX* [published March 2015] – provides recommendations on how to determine the number of DSDs for a subject-matter domain, following the approach that the decision on the number of DSDs needed for describing

a statistical domain should come after - and as a consequence of - the discussion on the data model and the architectural approach to be followed in the data collection exercise. The document is targeted towards domain experts and data modelling experts involved in the design of SDMX artefacts for representing a statistical domain.

b. Metadata Structure Definitions

The provision of methodological information or metadata with statistics that outlines concepts, definitions and describes methods used in collection, compilation, transformation, revision practices and dissemination of statistics, etc., is an essential function of all agencies disseminating statistics at both national and international levels. The availability of adequate metadata enables the assessment of the quality of the indicators disseminated by countries around the globe, in particular, by those of developing countries[26]. SDMX standards and guidelines necessitate the structuring of and facilitate the efficient exchange of metadata. However, there are a number of metadata-related issues that need to be considered by countries prior to SDMX implementation, especially in regard to the use of Metadata Structure Definitions (MSDs) – refer Step II below.

MSDs are structural metadata that describe the structure of metadata sets. An MSD defines the concepts and their hierarchy which comprise the metadata to be reported, the types of objects to which the metadata relate, and the means for identifying those objects. An effective MSD will allow metadata aware systems to understand the areas in which additional metadata may be available, and to create efficient systems for collecting and presenting the reference metadata[27].

To date, no global MSDs have been developed and the use of SDMX standards for metadata exchange have only been undertaken by Eurostat in the context of its corporate ESS Metadata Handler and by UNSD for Millennium Development Goals (MDGs). The IMF's Data Quality Assessment Framework for the SDDS and GDDS requires subscribing countries to provide metadata on the basis of a structured metadata template. Although the metadata items included in the Eurostat and IMF initiatives vary there is considerable overlap and each are based on the same metadata concepts and definitions.

Although the need for the linkage of data and metadata is inescapable, national agencies need to be pragmatic in determining the relative priority regarding the use of SDMX MSDs for the actual transmission of metadata to users. This pragmatism stems from the fact that changes in statistical methodology described in the metadata are often less frequent than the daily, weekly, monthly, quarterly and annual data dissemination cycles which benefit from the use of efficient transmission standards. Of greater importance is the actual existence of reference metadata for all statistical domains disseminated by the agency that describes concepts, collection methodology and techniques used in compilation and dissemination that allows the user to gain an understanding of the relevance and quality of the data in the context of its intended use[28].

---

[26] For a more detailed description of the need for metadata and metadata standards refer: OECD (2007): *Data and Metadata Reporting and Presentation Handbook*, OECD, Paris. Also available from http://www.oecd.org/std/37671574.pdf

[27] Eurostat, *SDMX Self-Learning Package, Student Book No. 5, Metadata Structure Definition*, p. 5, Eurostat, Luxembourg

[28] A distinction may be made between descriptive metadata that describes things such as scope and coverage, concepts, compilation methodologies and qualitative metadata that provides information on the quality of the data, strengths and limitations, etc. Such metadata are often expressed in the context of a quality framework that encompasses quality dimensions such as accuracy, comparability, coherence, timeliness, etc. Statements of the quality of the data may be given in the form of Quality Declarations such as those prepared on the basis of Eurostat's *Quality Assurance Framework* or the IMF's *Data Quality Assurance Framework*.

In many national statistical agencies reference metadata, where it exists, is frequently either incomplete or in the form of large amounts of unstructured text in static MS-WORD or PDF files, which makes its access by users more complicated. In addition to the availability of such metadata, organisations need to consider how it could be structured, in the absence of a global metadata template standard perhaps using one of the metadata templates developed by either Eurostat or the IMF (refer Annexes 3 and 4).

i. Eurostat's ESS Metadata Handler (ESS MH)

Eurostat's European Statistical System (ESS) Metadata Handler comprises IT applications for managing the production, exchange and dissemination of European statistical system metadata. It accommodates the Euro-SDMX Metadata Structure (ESMS)[29] and the ESS Standard for Quality Reports Structure (ESQRS)[30] for the production, exchange and dissemination of the national and Eurostat metadata files.

The ESS MH stimulates the use of the SDMX based ESS metadata standards within the ESS and the further integration of the ESS metadata systems. The tool enables the on-line production of the national ESMS and ESQRS files and their exchange within the ESS.

Euro-SDMX Metadata Structure (ESMS)

The ESMS facilitates the documentation of methodologies, quality and the statistical production processes in general. It uses 21 high-level concepts, with a limited breakdown of sub-items, strictly derived from the list of cross domain concepts in the SDMX Content Oriented Guidelines – refer Annex 3 for a listing of the Euro-SDMX Metadata Structure concepts. When fully implemented, there will be more than 2 000 Eurostat and national ESMS files produced and disseminated[31].

ESS Standard for Quality Reports Structure (ESQRS)

The ESQRS provides the concept base for the ongoing and systematic monitoring and reporting of the quality of statistics produced within the European Statistical System. This will enable producers to monitor the quality of the statistics by concentrating on the main quality concepts, namely: relevance; accuracy; timeliness and punctuality; accessibility and clarity; comparability; and coherence.

The ESQRS uses eight main concepts taken from the SDMX cross-domain concepts plus more detailed sub-concepts measuring data quality. It is a metadata structure which is used across all statistical domains[32].

Single Integrated Metadata Structure (SIMS)

In 2009 Eurostat created the Single Integrated Metadata Structure (SIMS)[33] which integrated and harmonised the two existing metadata and quality reporting structures described above, the ESMS and the ESQRS. The SIMS is presented below in Annex 5. The working group responsible for the development of SIMS recommended that all statistical processes in the ESS should at least have a

[29] For further information about ESMS and the structure of the standard refer http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/Annexes/ESMS_Structure.xls. For ESMS files produced so refer http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/metadata/metadata_structure

[30] For further information refer http://ec.europa.eu/eurostat/ramon/statmanuals/files/ESQR_final.pdf

[31] Refer http://www.oecd.org/std/47624538.pdf

[32] Source: Eurostat

[33] The 2009 edition of the *Technical Manual of the Single Integrated Metadata Structure* was developed by the Task Force on Quality Reporting, a sub-group of the Working Group on Quality in Statistics on the recommendation of the High-Level Task Force Sponsorship on Quality, in close cooperation with the ESS Metadata Working Group. The manual will be revised in 2015. For 2009 version refer http://ec.europa.eu/eurostat/documents/64157/4373903/03-Single-Integrated-Metadata-Structure-and-its-Technical-Manual.pdf/6013a162-e8e2-4a8a-8219-83e3318cbb39

basic quality report in the form of short or user quality report (ESMS). If the specific needs and/or the context of the statistical process require more detailed information on the different quality aspects, then the use of the detailed or producer quality report (as ESQRS) is recommended. The main advantage of SIMS is that it provides the conceptual framework and complete inventory for all quality and metadata concepts which will be stored in the same database by the use of the ESS Metadata Handler and can therefore be re-used for other metadata and quality reporting. The database is also accessible for Member States. Creation and exchange of reports will be quick and automated, based on the pre-defined report structures which are automatically retrievable from the system [SIMS Technical Manual, p. 10].

ii. IMF's Data Quality Assessment Framework (DQAF) / African Development Bank (AfDB) Open Data Platform (ODP)

Structured metadata on the basis of templates are provided by around 180 countries subscribing to the IMF's SDDS and GDDS that facilitate assessments of data quality across a range of statistical domains – refer Annex 4 for DQAF metadata items. The metadata concepts used in the templates are mostly consistent with those used by Eurostat, OECD, etc[34].

At the moment, the ODP does not include a metadata component which is something the IMF and the AfDB will look into at a later phase. Immediate plans are to provide countries with a data dissemination platform that meet the "Open Data" criteria[35].

iii. Millennium Development Goal (MDG) metadata

The Millennium Development Goal (MDG) Indicators website[36] disseminates official data, definitions, methodologies and sources for more than 60 indicators compiled by specialist UN agencies to measure progress towards the Millennium Development Goals. The data and analyses are the work of the Inter-agency and Expert Group (IAEG) on MDG Indicators, coordinated by the United Nations Statistics Division (UNSD).

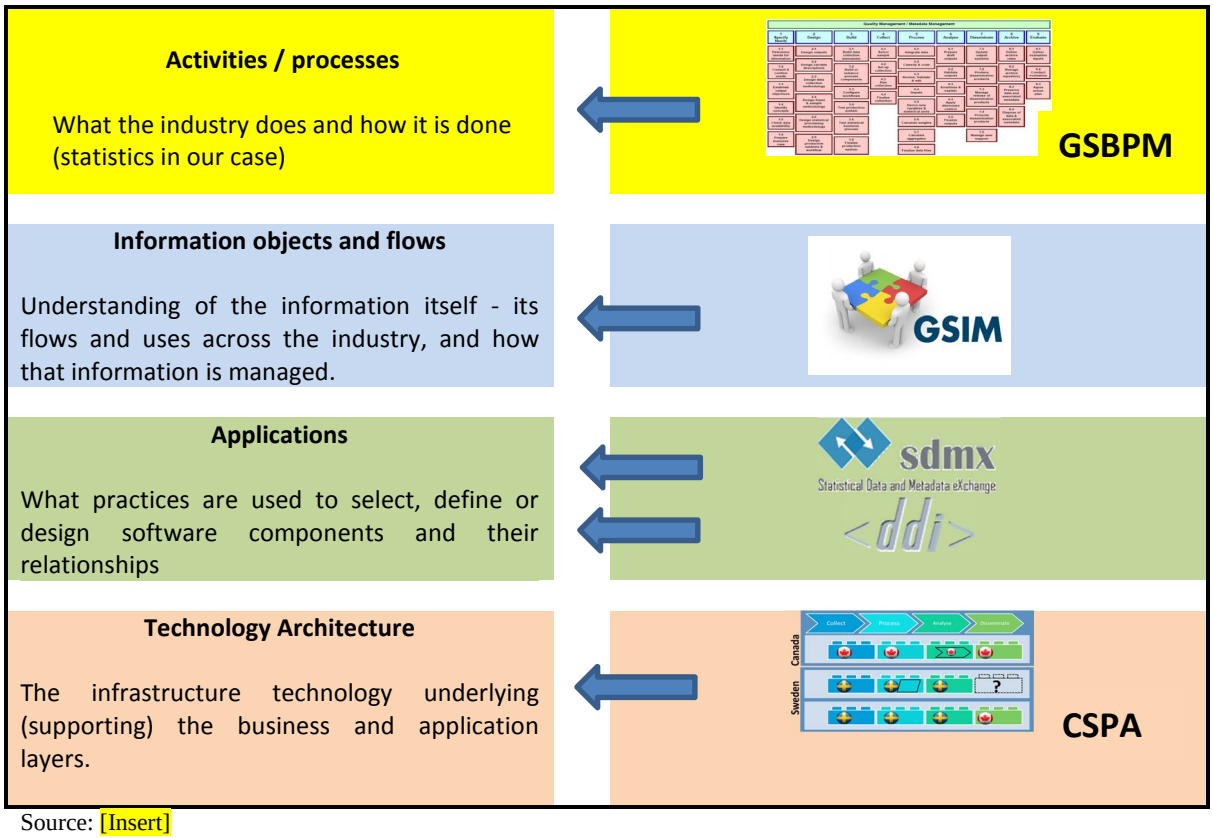
The metadata on the UNSD MDG website are structured as follows: definition; method of computation; comments and limitations; sources of discrepancy between global and national figures; process of obtain data; treatment of missing values; regional and global estimates; and expected time of release. SDMX metadata standards have been used to disseminate this metadata on the UNSD website[37].

3. OTHER RELEVANT STANDARDS

SDMX, with its focus on the exchange of aggregated data, is but one of a range of inter-related standards. These other standards and information models comprise the: Data Documentation Initiative (DDI); Generic Statistical Business Process Model (GSBPM); General Statistical Information Model (GSIM), and Common Statistical Production Architecture (CSPA). The relationship between these standards is illustrated in the following diagram.

[34] Correspondence tables between the metadata templates developed by Eurostat, the IMF and the OECD can be found on the SDMX website at http://sdmx.org/?page_id=11
[35] Source: IMF
[36] Refer http://unstats.un.org/unsd/mdg/
[37] As discussed below, country-level metadata are reported by countries participating in the CountryData project.

Diagram 3: Relationship between SDMX and other relevant standards



Source: [Insert]

a. Data Documentation Initiative (DDI)

DDI focuses on the description of microdata[38] and metadata associated with it. DDI and SDMX should be viewed as being complementary rather than competing standards. The focus of SDMX is to facilitate the automated exchange and processing of data, particularly macrodata, and metadata

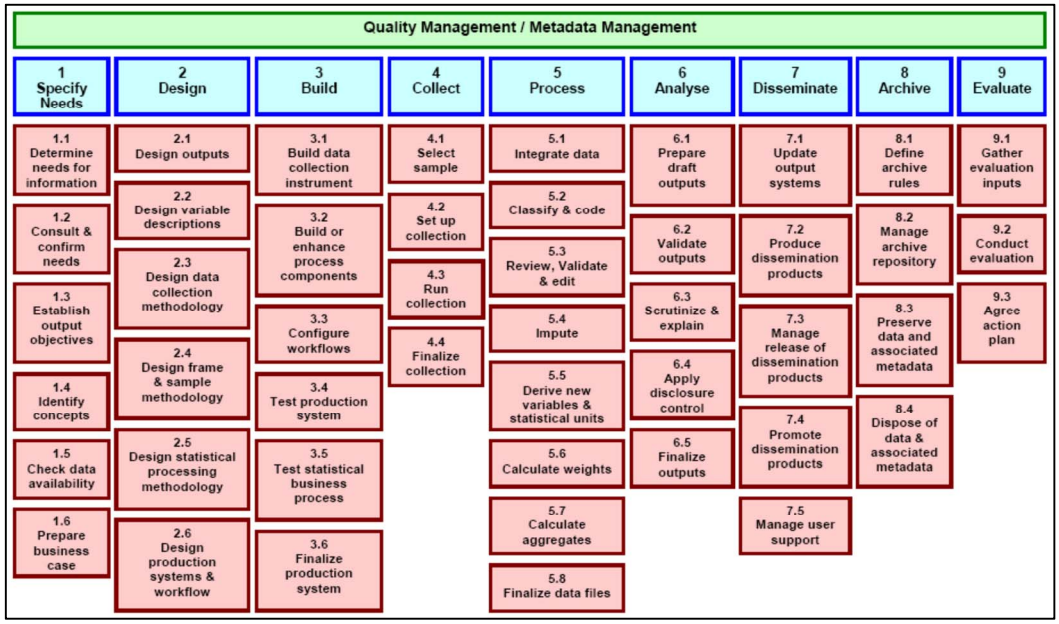
[38] Refer http://www.ddialliance.org/what

between organisations. On the other hand DDI is a specification for capturing metadata about social science data over the whole data lifecycle[39].

b. Generic Statistical Business Process Model (GSBPM)[40]

GSBPM is a framework describing the statistical production process in terms of standard components (phases and sub-processes). It is intended to apply to all activities undertaken by producers of official statistics.

Diagram 4: Generic Statistical Business Process Model (GSBPM)



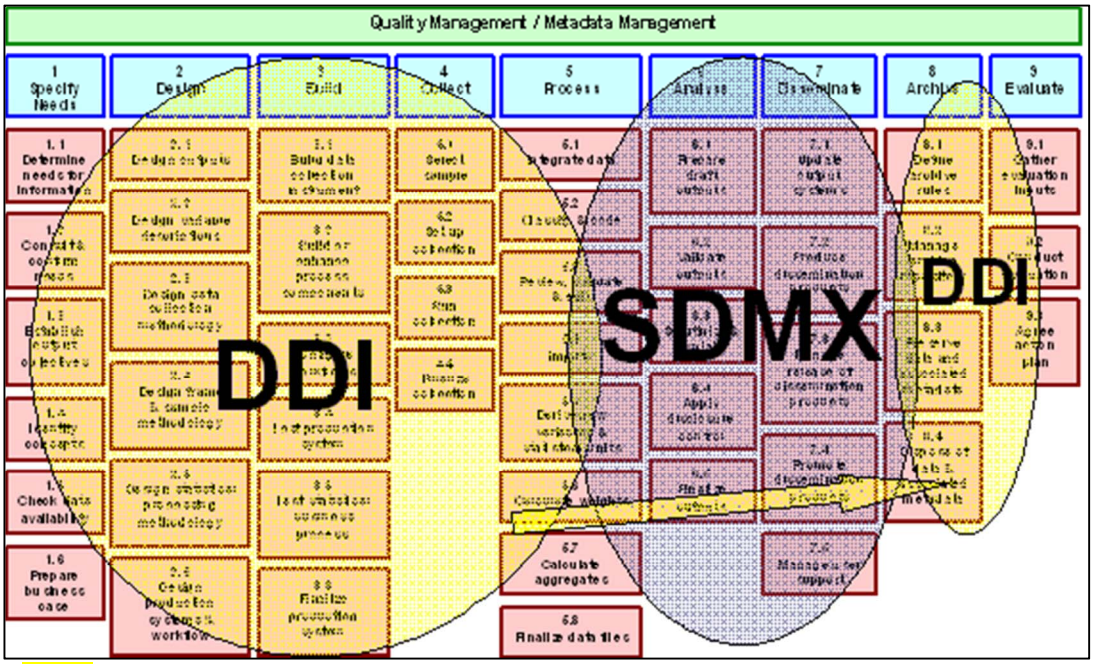
Source: UNECE, 2013, *Generic Statistical Business Process Model (GSBPM)*, UNECE, Geneva

The relationship between SDMX and GSBPM and DDI is illustrated in the following diagram.

Diagram 5: Relationship between SDMX, GSBPM and DDI

[39] Source: Gregory, A., and Heus, P, *DDI and SDMX: Complementary, Not Competing, Standards*. Detailed commentary on the differences between the two standards is beyond the scope of the current document. Those requiring further information are referred to Gregory and Heus, available at http://www.opendatafoundation.org/papers/DDI_and_SDMX.pdf

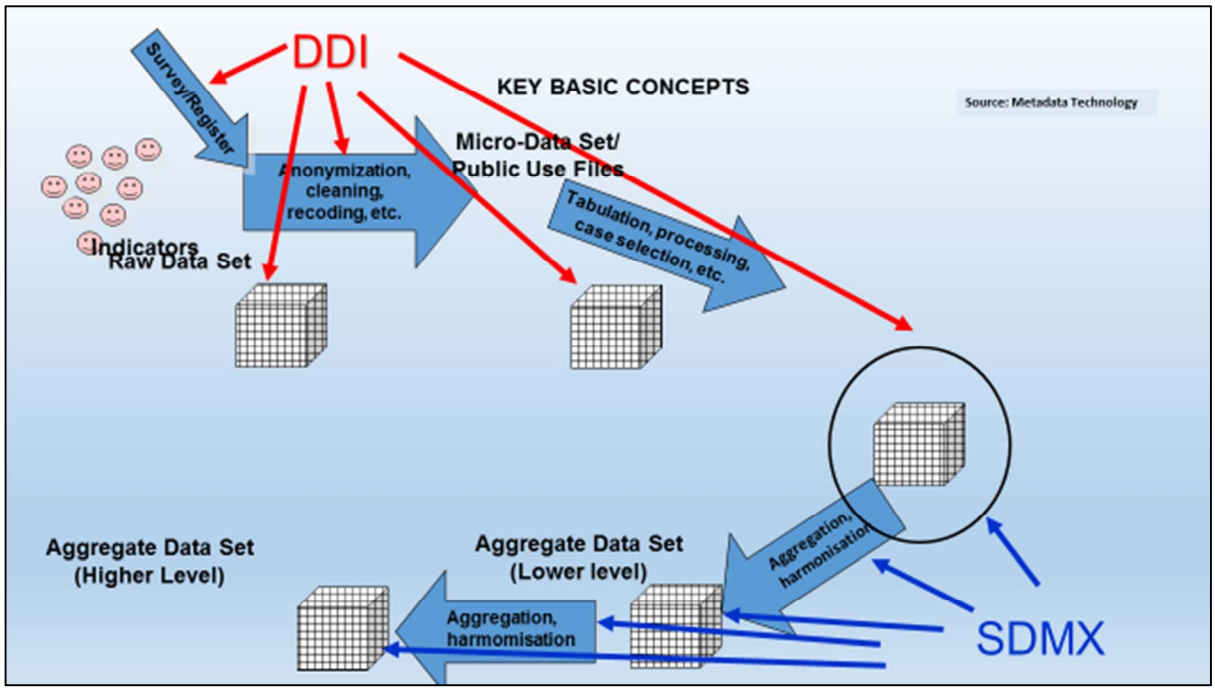
[40] Refer UNECE, 2013, *Generic Statistical Business Process Model (GSBPM)*, UNECE, Geneva – available at http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0



Source: [Insert]

The complementary relationship between DDI and SDMX is further illustrated in the following Diagram.

Diagram 6: Complementary relationship between the uses of DDI and SDMX



c. Generic Statistical Information Model (GSIM)

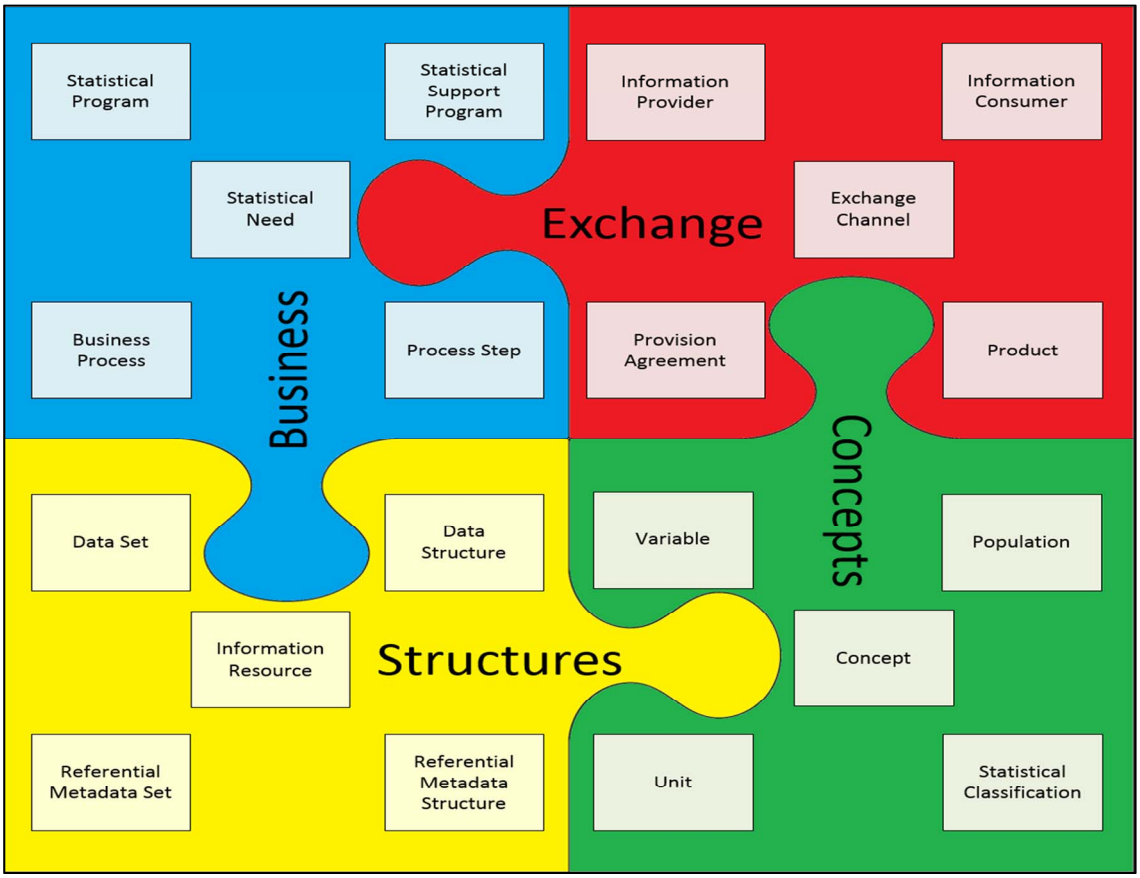
GSIM is a reference framework that provides a set of standardised descriptions of information objects

which are the inputs and outputs in the design and production of statistics. GSIM helps to explain significant relationships among the entities involved in statistical production, and can be used to guide the development and use of consistent implementation standards or specifications. GSIM provides the information object framework supporting all statistical production processes such as those described in the Generic Statistical Business Process Model (GSBPM), giving the information objects agreed names, defining them, specifying their essential properties, and indicating their relationships with other information objects.

Examples of such objects include data and metadata (such as classifications) as well as the rules and parameters needed for production processes to run (for example, data editing rules). GSIM identifies around 110 information objects, which are grouped into four top-level groups. The following diagram provides a simplified view of the information objects identified in GSIM and gives examples of the objects that are in each of the four top-level groups[41].

Diagram 7: Generic Statistical Information Model (GSIM)

[41] Refer UNECE, 2013, *Generic Statistical Information Model (GSIM): Communication Paper for a General Statistical Audience*, Version 1.1, December, 2013, UNECE, Geneva – available at http://www1.unece.org/stat/platform/display/gsim/GSIM+Communication+Paper



Source: UNECE, 2013, *Generic Statistical Information Model (GSIM): Communication Paper for a General Statistical Audience*, Version 1.1, December, 2013, UNECE, Geneva, p. 5

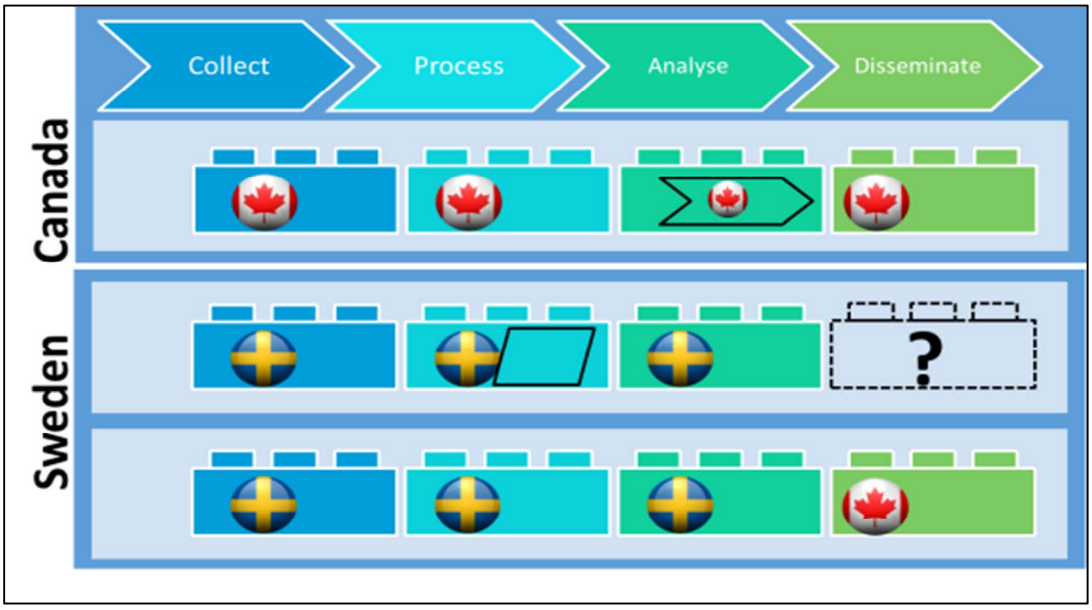
d. Common Statistical Production Architecture[42] (CSPA)

CSPA outlines a generic architecture for statistical production. It describes the mechanisms to design, build and share components with well-defined functionality that can be readily integrated into multiple processes both within an organisation and between organisations. By adopting this common reference architecture, it will be easier for each organization to standardize and combine the components of statistical production, regardless of where the statistical services are built. As shown in the following diagram, Sweden could reuse a statistical service from Canada because they both use the same component. The statistical services that are shared or reused across statistical organizations might be new statistical services that are built to comply with CSPA or legacy/existing tools which comply with the architecture. This is shown in the diagram by the shapes inside the building blocks.

The use of SDMX (and DDI) is a core recommendation in the CSPA framework.

Diagram 8 Application of the Common Statistical Production Architecture (CSPA)

[42] Refer UNECE, 2015, *Common Statistical Production Architecture (CSPA)*, UNECE, Geneva – available at http://www1.unece.org/stat/platform/display/CSPA/CSPA+v1.1



Source: UNECE, 2015, *Common Statistical Production Architecture (CSPA)*, UNECE, Geneva

In conclusion, a key element in the adoption of all of the standards described briefly above, as well as SDMX, is the adoption of common concepts and terminology by all of the implementing national statistical agencies and international organisations.

4. SDMX IMPLEMENTATION TOOLS

The following information was gleamed from papers and presentations given at meetings of the SDMX Global Conference, the SDMX Expert Group, the SDMX website, websites of SDMX sponsoring agencies and recent discussions with OECD, IMF, AfDB, Eurostat, UNSD and national statistical organisations (NSOs).

Options for national agencies implementing SDMX are to use tools that have been developed by international organisations or, to develop their own applications – refer Annex 1 below. Over the last five years or so the SDMX sponsoring agencies have developed a range of tools that incorporate the SDMX Technical Standards and Content-Oriented Guidelines (COGs). The tools have been developed to facilitate a number of specific implementation initiatives undertaken by each agency (see below) and provide facilities such as web services, registries and mapping.

There are currently a number of SDMX implementation tools that national statistical agencies could use or adapt[43]. These implementation tools comprise:

- SDMX-RI which has been used extensively by EU Member countries for the exchange of data and metadata. SDMX-RI has also been adopted by other international organisations such as the OECD.

- SDMX Converter an application tool for converting between SDMX-ML and other formats.

- IMF's SDDS Plus which has used SDMX for the exchange of statistics in the National Summary Data Page (NSDP). It has also been adopted by the African Development Bank (AfDB) for their

[43] The different tools, demonstrations on their use, etc, are available in the SDMX Tools Repository also located on the SDMX website at http://www.oecd.org/std/47610077.pdf.

Open Data Platform (ODP). The use of the tools by developing countries in Africa could be relevant for those in other regions around the world.

- DevInfo - a database system for organizing, storing and presenting data in a uniform way to facilitate data sharing at the country level across government departments, UN agencies and development partners. DevInfo also has features that produce tables, graphs and maps for inclusion in reports, presentations and advocacy materials. In 2004, DevInfo was endorsed by the United Nations Development Group (UNDG) to assist countries in monitoring achievement of the Millennium Development Goals (MDGs)[44].

- SDMX Global Registry (local implementation or cloud solution) a central repository for global SDMX artefacts.

- Tools developed by non-government organisations such as Fusion Products developed by Metadata Technology with part-funding by the SDMX sponsor organisations and the SuperSTAR suite of software developed by Space-Time Research.

The interoperability of these implementations and whether it would be possible to use a combination of some of the tools developed is still being discussed. Theoretically, the inclusion of the COGs in the tools and the use of global DSDs should make them interoperable. In addition, adoption by a growing number of organisations of the SDMXSource[45] open source components will guarantee software interoperability.

As mentioned above, a number of national statistical agencies (such as the Australian Bureau of Statistics, the Swiss Federal Statistics Office and Mexico's INEGI) have opted to develop their own implementation tools.

Criteria that countries could use for selecting the implementation tools that best suits their needs are described in structured implementation Step IV below.

a. SDMX Reference Infrastructure (SDMX-RI)

Eurostat's universal framework for modern data provision, SDMX-RI, provides a set of building blocks allowing a statistical office to expose data to the external world based on access rights. SDMX-RI was developed by Eurostat to facilitate the efficient exchange of data and metadata by European Union Member states. A common characteristic of the IT environment of many of these countries is the storage of data and metadata on well-developed database systems that are also used as platforms for the dissemination of (sometimes) highly disaggregated data to both government and non-government users. However, many EU Member statistical agencies still store at least some of their key statistics in Excel files.

The SDMX-RI building blocks are designed to provide data and structural metadata based on mappings to each organization's dissemination data warehouse. The infrastructure uses SDMX standards and was developed to simplify the exchange of data and metadata. It provides standard software and components, allowing individual statistical organisations to interact and exchange their data using the same software and methodology. SDMX-RI's modular approach enables users to use either part or the entire infrastructure.

The SDMX-RI building blocks[46] comprise:

[44] Source: About DevInfo website - http://www.devinfo.org/libraries/aspx/AboutDevInfo.aspx?T=ADI&PN=diorg/di_about.html
[45] Refer www.sdmxsource.org
[46] Refer http://www.oecd.org/std/47727536.pdf and https://webgate.ec.europa.eu/fpfis/mwikis/sdmx/index.php/SDMX_Reference_Infrastructure_SDMX-RI

- Web Service - a service that accepts both SDMX Query messages for data and structural metadata.
- Web Client - a web application that allows users to create basic SDMX queries to be used for retrieving data from a Mapping Store database and dissemination databases using a SDMX-RI Web Service instance.
- Mapping Assistant - provides mapping information between SDMX artefacts and the local data storage scheme. Data reporting organizations will have to map the concepts in their local databases to the concepts and code lists of the DSD.
- Test Client - meant to test the SDMX-RI building blocks and the Web Service endpoint[47]. The test client has also been used to create files and push data to file repositories. For example, FTP servers, mail boxes, web site repositories.

As mentioned above, Eurostat has also developed a number of training videos, work books and self-assessment resources to assist the implementation of the tools in their SDMX-RI.

In addition, Eurostat provides training packages. Eurostat has already helped present SDMX to developing countries in Latin America, in partnership with UN ECLAC. Eurostat offers regular training on SDMX in Luxembourg and Europe and from time to time provides capacity development at the margins of meetings such as the SDMX Expert Group meeting held in Korea in October 2014 and the SDMX Global Conference held in Bangkok in September 2015.

The UNdata Application Programming Interface (API) provides dynamic, programmatic access to data within the UNdata platform[48]. Developers can use the API to dynamically query UNdata to obtain the latest data and display the result on a web page, download to local storage for further processing, etc. UNdata API is powered by Eurostat's SDMX-RI. The API is implemented as a SOAP Web Service that can be used to query the datamarts using the SDMX standard. The UNdata API currently contains six datasets including MDG data, CountryData and UNESCO data. UNSD plans to gradually extend coverage to include other UN databases.

b. SDMX Converter

The SDMX Converter is an open source application tool for converting between SDMX-ML and other formats. The Converter can be used by NSOs who store aggregated data for the statistical domain(s) being implemented in Excel. In such instances countries need to run Converter software which links data stored in national Excel spreadsheets to a standard Excel format template specified by Eurostat. However, the downside of this process is that mapping between the two sets of Excel formats will need to be done manually and not via the SDMX-RI Mapping Assistant. The Converter may be used without SDMX-RI. Other currently supported formats for the SDMX Converter include: GESMES/TS; GESMES/2.1; GESMES/DSIS; CSV; FLR; DSPL.

c. IMF's SDDS Plus

SDDS Plus[49] is a new tier in the IMF's Data Standards Initiatives. SDMX is the dissemination format selected for SDDS Plus, replacing HTML. Adherence to SDDS Plus is open to all SDDS subscribers, especially those with systemically important financial sectors. SDDS Plus has nine additional data categories to the SDDS and longer time series – requirement is for five years of data for all prescribed components.

[47] Source: Eurostat
[48] Refer: http://data.un.org/Host.aspx?Content=API
[49] Link to Powerpoint overview of SDDS Plus: http://www.oecd.org/std/SDMX%202013%20Session%203.11%20-%20The%20benefits%20of%20SDMX%20for%20SDDS%20Plus.pdf

The nine additional data categories, which an adhering country commits to fully observe by the end of 2019, cover: sectoral balance sheets; general government operations and general government gross debt; the other financial corporations' survey, financial soundness indicators, and debt securities; and participation in the Coordinated Portfolio Investment Survey, the Coordinated Direct Investment Survey, and the Currency Composition of Official Foreign Exchange Reserves exercises. The National Summary Data Page (NSDP) in HTML is replaced by a series of SDMX-ML files or an SDMX web service call (one per data category). This eliminates the need for a coordinating agency to centralize the publication of the NSDP data. Contributing agencies can post/register their SDMX files directly. Countries continue disseminating data in "human readable" formats on their national sites (following existing practices, formats, "look and feel").

The IMF provides the following tools to support implementation[50]:

- Cloud based SDMX converter: for countries without an existing SDMX infrastructure the IMF SDMX converter can map Excel, CSV or GESMES-TS to SDMX-ML
- IMF SDMX Registry and Registration Service: to inform of the availability of new data files
- Draft Technical Implementation Guide: provides guidelines for preparing the NSDP using SDMX[51]
- IMF developed mapping of countries' data to DSD to assist implementation

In addition to the NSDP-DSD, the IMF will support any "DSD for Global Use" developed under the auspices of SDMX Sponsors. Countries are free to decide which DSD (NSDP-DSD; Global DSDs) they wish to use for each data category.

The IMF is also investigating the possibility of providing a data hub that will centralize and facilitate access to data provided by SDDS Plus adherents. The successful implementation of the SDDS Plus dissemination model would pave the way for enhancing data access for the 180 economies in SDDS/GDDS.

The AfDB's Open Data Platform is adopting SDMX and the NSDP-DSD for the dissemination of economic and financial statistics by countries in Africa – refer below.

IMF – African Development Bank (AfDB) Open Data Platform (ODP)

The ODP is a cloud-based platform. Its development was sponsored by the African Development Bank who make it available to data producing agencies in Africa. It does not require any infrastructure investment on the part of the user agency.

The IMF is currently setting up each agency in a country (NSO, central bank, ministry of finance) with its own instance of the ODP, which the national agency fully governs (the Fund provides a suggested governance model). It is very straightforward to set up a dataset in the ODP and to load data into it via Excel or CSV. The data can be loaded using local codes and the mapping to SDMX codes (the ECOFIN DSD, the same used for SDDS Plus) takes place automatically in the platform. The data then become available to users via the user interface (which provides dashboard and charting functionality) and supports machine-to-machine exchange using the ODP SDMX web service. Such facilities provide the opportunity to take full advantage of SDMX implementation and to improve the efficiency of data dissemination at the national level[52].

[50] The IMF service uses tools developed by Metadata Technology – SDMX Registry, Cloud-based services for validation and mapping.
[51] Refer http://www.imf.org/external/pubs/ft/sdds/guide/plus/2014/sddsplus14.pdf
[52] Link to YouTube playlist of six videos describing the Open Data Platform and the data dissemination tools it provides - http://www.youtube.com/watch?v=xl9XvkKhf38&list=PL63A62B5F4AE68645

The data can also be downloaded in various formats (Excel, CSV, pdf, SDMX) via the user interface. The IMF has already started using the ODP SDMX web service as its data transmission (pull mode) system for ODP-using countries.

Countries in the African region have expressed very strong interest for all international organizations to adopt the OPD and its SDMX-based transmission of data for collecting information from countries. This will require international organizations to collaborate closely in agreeing on global DSDs for an expanding number of statistical domains. These DSDs would be utilised in the development of ODPs (using national codes) and mapped to a standard set of SDMX codes. The ECOFIN DSD is suitable for economic and financial statistics, but other DSD(s) would be needed for socio-demographic data. The IMF has started working with ILO and FAO (and AfDB) in that direction and would welcome working with regional partners in other parts of the globe.

The ECOFIN DSD is a "dissemination" DSD and, as per the SDMX Standards "Technical Notes" for public viewing of data. It contains a small number of dimensions (i.e. four). Based on their experience so far with ODP countries, the Fund sees significant advantages in using a "simple" DSD, and avoiding the multiplicity of dimensions such as the potential 26 dimensions in the national accounts global DSD. Many dimensions lead to difficulty in the mapping process and generate data outputs that are difficult to visualize on the web or use with mobile apps. The IMF sees the latter as an important future "client" of the ODP[53].

Domain coverage

There is no limit to the potential domain coverage of the ODP. When the Fund helps countries with populating data in the ODP instance of a data producing agency, they set in that ODP instance the mapping between the national codes and the ECOFIN DSD, so that the web service can output the data according to the ECOFIN DSD. The SDMX web service of the ODP could also output the data in SDMX format on the basis of the national codes, although the resulting SDMX-ML data files would not be linked to a "formal" DSD. The ECOFIN DSD is available on the IMF registry[54] and is used by the IMF's data collection system to check the conformance of the data file to the ECOFIN DSD.

This implies that the ADB-ESCAP initiative could develop a DSD and load data into the ODP according to this specific DSD (or include in the ODP the mapping between the national codes and the initiative's DSD) – the web service could then be set to output the data according to the project's DSD.

However, the objective of the data sharing platform offered by the ODP is to work collaboratively with international organizations in harmonizing data reporting templates and DSDs, so that countries don't have to load multiple files and formats in ODP to meet data requirements of individual international organisations that could instead be harmonized across collecting agencies. So, to the extent that ADB and ESCAP are interested in economic and financial statistics, the IMF suggests working together to harmonize the respective data requirements and the IMF would be willing to extend the ECOFIN DSD if this could lead to a reduction in the reporting burden of countries. For data in the socio-economic domains, the IMF also encourage international organizations that have similar data requirements to work together to develop simple DSD(s) on the basis of harmonized data collection templates[55].

The AfDB reported two main challenges in the implementation of the ODP across countries in the region.

[53] Refer to *SDMX Implementation Model* paper described above in the discussion on Data Structure Definitions
[54] Available at http://dsbb.imf.org/Pages/SDDS/GetPage.aspx?pageName=SDMXDataStructuresforNSDP
[55] Source: IMF

- To get countries to upload and update their data sets on the ODP.

- Each country has its own indicators, definitions, classifications, which may depart from international standards. Hence the importance of mapping with the international DSDs. The AfDB has started this work with the IMF, where the data are being uploaded in the ODP, mapped to the IMF codes and DSDs for macrodata. It is still being piloted in three countries and will be extended to others once the process has been finalized.

As mentioned above, a priority is to have the international community agree on a common set of DSDs which regional institutions could then use.

Mapping process for the IMF – *AfDB initiative*[56]

For the IMF – AfDB ODP initiative mapping is undertaken at the time of the creation of a dataset into the ODP. For example, when establishing a BoP dataset in a specific country's ODP:

1. it is set with a set of basic structural metadata (e.g. country, time, frequency, scale) that are the same across all datasets;
2. then, an "Indicator" code is introduced for all the BoP concepts, based on the national codes;
3. then, an "alias" ECOFIN code for these indicators is attached to each national code, allowing data users to "talk" the same language across all ODP instances available on the African continent. The codes differentiate between methodologies to a certain extent (e.g., BPM5, BPM6).

The consistency of the mapping phase is ensured via IMF technical assistance missions that work with national agencies in establishing the mapping. For example, when working with labour market indicators, the IMF carefully reviews the concepts underlying national data to ensure that they are mapped to the concepts of the ECOFIN DSD. To support a better understanding of the country's needs, reference is made to the SDDS/GDDS metadata prepared by the country, which cover a large portion of the datasets of the ECOFIN DSDs.

For "simple" database environments, e.g. Excel-based data compilation systems, data can be loaded into the ODP directly using Excel (with some structure, so that the "Indicator" code is available). When working with database applications, the Fund's approach is to export data (for example, from an Oracle database) into csv formats, which are straightforward to load into ODP.

d. DevInfo

DevInfo is a UN endorsed data dissemination application currently utilised by a large number of countries to monitor socio-economic data. As mentioned above, DevInfo was endorsed by the United Nations Development Group (UNDG) for the reporting and dissemination of the Millennium Development Goals (MDGs).

DevInfo[57] is a database system that provides tools for organising, storing and presenting data in a uniform way to facilitate data sharing at national and international levels. DevInfo uses SDMX to facilitate data and metadata exchange. DevInfo also has facilities to produce tables, graphs and maps. DevInfo provides tools to create a standardised SDMX-compliant DSD and to create an SDMX data exchange registry.

[56] Source: IMF

[57] The current version, DevInfo 7.0, was launched in October 2012.

UNSD's use of DevInfo for the MDG and CountryData projects

The United Kingdom Department for International Development (DFID) funded a project on the exchange of development indicators produced by countries, including their MDG equivalents.

As mentioned above, the MDG DSD was developed and approved by the Interagency and Expert Group on MDGs Indicators (IAEG) for the dissemination of official MDGs compiled by specialist UN agencies. CountryData on the other hand is a superset of the MDG DSD with additional indicators as requested by project countries. Whilst the DFID project retained the dimensionality of the MDG DSD, the code lists were extended. As a result, the CountryData DSD is fully compatible with the MDG DSD.

The MDG DSD has been used in a limited-scale exchange of MDG indicators between UNSD and other agencies, while CountryData DSD has been used more widely to exchange development indicators between UNSD and a total of 15 countries. It is also used to disseminate indicators collected from a total of about 100 countries.

The pilot project[58] involved eleven countries around the globe[59] and was designed to give the participating countries their first experience with SDMX using well-known and easy-to-use platform and tools. Countries were able to decide on their preferred platform, and 9 of the 11 countries opted for DevInfo[60.] Working with the project, the DevInfo Support Group developed a set of tools that facilitate data exchange, such as the SDMX Mapping tool and SDMX Registry. The DevInfo SDMX Mapping Tool was the interface that simplified mapping between the DevInfo database structures and the CountryData/MDG DSD. The DevInfo SDMX Registry, with its SOAP-based API, supported SDMX subscriptions and notifications and was integrated with DevInfo Version 7.0.

UNSD's *CountryData*[61], which is the central repository for the project, is part of the *UNdata* platform[62]. *CountryData* enables countries to share official national development data with the global user community. National statistical systems have identified priority development indicators based on their specific country context, and these data are now available via *CountryData*. *CountryData* automatically receives and published data from countries' registries. Countries control what is published and data are published without any intervention by UNSD. The structured reference metadata disseminated on *CountryData* facilitates understanding of national MDG data and any sources of discrepancies.

ILO's use of DevInfo

As outlined below in Annex 1, the ILO is currently engaged in a pilot project involving ten countries with DevInfo to generate an ILOSTAT template for DevInfo 7 that would allow countries to define a local repository for ILOSTAT indicators comprising both data and metadata and then transmit the information to the ILO in SDMX. This would realise the following benefits:

- countries would be able to load data into a DevInfo database and then DevInfo will produce the SDMX files according to the ILO DSDs;
- the national repository would be the DevInfo database, and the "mapping" to ILO DSDs will be intrinsic to it.

[58] Source: UNSD Powerpoint presentation – available at http://unstats.un.org/unsd/statcom/statcom_2013/seminars/SDMX/UNSD%20Presentation.pdf
[59] Namely: Burundi, Cambodia, Ghana, Laos, Liberia, Morocco, Palestine, Rwanda, Thailand, Uganda, Vietnam
[60] Thailand developed data exchange based on its data exchange and dissemination system, StatXChange, and uses Metadata Technology's Fusion Registry. Morocco developed SDMX exchange using its own data dissemination database, BDS.
[61] Refer: http://data.un.org/countrydata
[62] Refer http://data.un.org/

Countries would also be able to use all DevInfo features to disseminate the information in their websites.

e. The SDMX Global Registry

The SDMX Global Registry[63], launched in May 2014, is intended to provide a central repository for all global artefacts (DSDs, MSDs, code lists, mapping sets, etc). In the long run the Registry will provide a place for public metadata; provide links to all local registries, and most important, provide a central point of access for all organisations wishing to use SDMX.

The SDMX Global Registry also provides two protocols for applications to query for SDMX content. These are RESTful Web Service and SOAP Web Service.

The SDMX Global registry was initially developed to support the technical world-wide implementation of the national accounts SNA 2008 and balance of payments BPM6 standards. It will eventually be expanded to provide a space for additional global and shared DSDs and their related objects as they become available to assist the SDMX implementation of other statistical domains. The SDMX sponsoring organisations have also formulated guidelines and quality requirements to govern the content of the new registry[64].

A number of the SDMX sponsoring agencies also have their own registries to facilitate their implementation projects. These include the BIS, ECB, Eurostat, IMF, and the United Nations.

f. Implementation tools developed by other organisations

A number of other organisations have also developed suites of software to assist the implementation of SDMX. The following list is by no means exhaustive and other organisations may be included in future revisions of the current document.

Metadata Technology

Metadata Technology has developed a range of free tools whose development and maintenance is funded by some of the SDMX sponsor organisations[65]. These include:

- *Fusion Registry:* SDMX Registry – a structural metadata respository that supports all of the structure artefacts that are in the SDMX standard. The registry is accessible using the SDMX web services for submission, query, and retrieval. It also has a GUI for metadata authoring and visualisation. The Fusion Registry is the registry used for the Global Registry service as well as registry services used by many other organisations.

- *Fusion Transformer*: a command line based application (i.e. invoked from and integrated into user systems) that validates and transforms all syntax and versions of SDMX data and structural metadata.

- *Fusion Weaver*: a GUI-based desktop application that performs the same functions as the Fusion Transformer.

- *SdmxSource*: implements all aspects of the SDMX Information Model - www.sdmxsource.org

[63] Refer https://registry.sdmx.org/home.html
[64] Refer *SDMX Global Registry Content Policy*, March 2015. Available at SDMX.org, tab Guidelines, Other Guidelines
[65] Refer http://www.sdmxfusion.com/products.html#Registry

Metadata Technology also developed the SDMX Common Component Architecture and an open source implementation of this in Java. This architecture has been adopted by Eurostat in its SDMX-RI and the Java implementation is used by the SDMX-RI components.

Other products developed by Metadata Technology (some of which are pay-for products) are:

- *Fusion Security:* Controls access rights and privileges for Fusion products such as Fusion Registry and Fusion Matrix.

- *Fusion Matrix:* a multi-functional application supporting multiple uses cases for the dissemination of data and associated metadata. This includes a database (if the data are not already stored in a database), metadata repository for quality metadata, SDMX web services and a programmers interface for developing bespoke data discovery and data visualisation front ends.

- *Fusion Browser:* developed in collaboration with the Plural Array Foundation. This is an Excel add-in that enables a user to discover, query for, and visualise data exposed by SDMX web services directly from within Excel.

- *Fusion Audit:* provides an insight into both live and historical activity for the Fusion Registry, Fusion Matrix and Fusion Security.

Space-Time Research

The SuperSTAR suite of software products developed by Space-Time Research provides tools for fast, confidentialised data tabulation, analysis, visualisation, reporting and dissemination of large amounts of organisational data in one integrated environment. Users may manipulate and analyse data at any level of aggregation, from unit record or micro data to aggregated cubes of data. SuperCROSS supports data loading and saving in SDMX format[66].

g. Concluding remarks on existing SDMX implementation tools

Options for a country deciding on which SDMX implementation tools to use include either selecting one or a mixture of those described above, or developing their own applications. It is unlikely that there will be a one-size-fits-all SDMX solution to the diverse needs of all countries around the world. For example, countries with more developed and comprehensive (in terms of domain coverage) database environments, where there is a need to disseminate data to a wide range of national and international users, might opt for greater flexibility and use global DSDs such as those recently developed for national accounts, BoP and merchandise trade. On the other hand, countries with Excel-based data storage systems might opt for an ODP / ECOFIN DSD approach. Within these two broad approaches there are a wide number of variants and legacy systems.

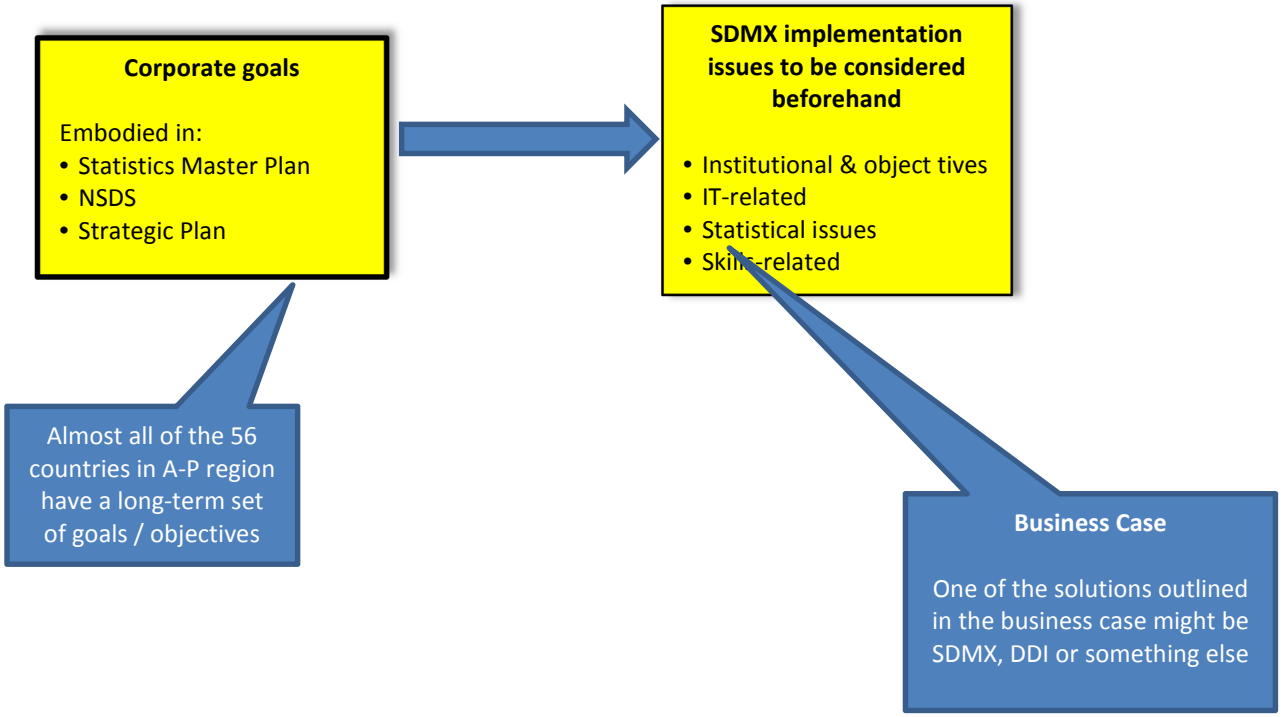
66 Refer http://spacetimeresearch.com/

STEP II: ADDRESSING ISSUES TO BE CONSIDERED BY COUNTRIES PRIOR TO EMBARKING ON SDMX IMPLEMENTATION

The process of mapping national datasets to DSDs and MSDs is key to the implementation of SDMX standards and guidelines at the country level, and one that requires considerable planning and exchange of information and cooperation between national statistical agencies and international organisations involved in the process.

It should be emphasised that a wide range of implementation issues need to be considered, and in some instances addressed, beforehand by national agencies across the NSS in countries contemplating SDMX implementation. These are related to and often derived from the key strategic NSDS goals and objectives such as those outlined in Table 1 above. In relation to SDMX implementation they cut across a number of topics including: institutional, IT, statistical, as well as skills-related issues[67]. As illustrated in the following diagram such implementation issues need to be related back to corporate goals in the form of a business case (refer Section B above) justifying the implementation of SDMX or other data / metadata exchange standard.

Diagram 9: Relationship between corporate goals and SDMX implementation issues



Key decisions on the range of issues outlined in the previous paragraph (in particular those involving institutional and statistical related issues) are required to identify the potential improvements in efficiencies in data and metadata reporting and dissemination / reporting that could be gained through SDMX implementation. This will assist the development of a sound business case that provides justification for the expenditure of the resources required for the initial implementation of SDMX and subsequent ongoing maintenance.

More importantly, these decisions will also provide technical basis for actual SDMX implantation, by ensuring that appropriate choices are made with respect to IT-related issues involved in actual SDMX

[67] Source: Eurostat, *SDMX Self-Learning Package, Student Book No. 7 - SDMX Architecture Using the Pull Method for Data Sharing*, pp 8-9, Eurostat, Luxembourg

implementation, such as the selection of appropriate data structure definitions, data and metadata exchange scenarios and selection of implementation tools best suited to the needs of the national organisation. Statistical domain experts must play a key role in these early decisions. A common mistake is for countries to decide on these and other IT-related issues before key issues, which are often institutional and statistical, are first given due consideration. SDMX technical standards and guidelines are primarily tools designed to improve the efficiency of what are basically statistical processes, rather than an end in themselves.

1. Institutional issues and objectives

- Linkage of SDMX implementation to existing strategic planning processes such as NSDS, Statistical Master Plan. Implementation should not be considered in isolation to existing corporate / NSS institutional and technical (statistical, IT, infrastructure) planning processes. It is an element of modernisation of statistical infrastructures, etc. Within the context of an existing strategic planning process a better scenario would be to formulate the business case / need for change and then identify which standards / tools to use such as SDMX, DDI, etc[68].

- What do national agencies want to achieve with SDMX implementation? What are their objectives? These, together with the identification of the statistical domains encompassed in the SDMX implementation (see below), are key areas that requires careful detailed consideration and thought. For example, are they interested in using SDMX standards:

  - to collect highly disaggregated data from national data providers, and /or;
  - for subsequent data dissemination and reporting to end users, for example to a number of international organisations, and / or national users.

  Agencies need to decide whether to implement SDMX for one, or both of these data flows. Additional factors that need to be taken into account include whether the data flows are regular and ongoing (where the timeliness of data access may be a critical issue), or designed to meet only periodic, irregular or ad hoc requests for data and metadata. Consideration of these issues at an early stage will assist in making subsequent decisions as to the appropriate SDMX data and metadata exchange scenarios countries may wish to implement and the choice of the appropriate platforms on which data and metadata may be stored and shared.

- Who would be the main drivers for SDMX implementation within national agencies – IT, dissemination and marketing, coordination or statistical subject matter areas. In decentralised NSSs which agency would take the lead / coordination role in SDMX implementation across the NSS? Options for data flows using SDMX standards within decentralised NSSs include:

  - Ministries and other data producers use SDMX standards for transferring data and metadata to an NSS hub maintained by a coordinating national agency which then disseminate outputs to national users and international organisations.

  - Ministries and other data producers use current data and metadata reporting arrangements to an NSS hub which then uses SDMX standards to disseminate outputs to national users and international organisations.

  - Each data producer within the NSS use SDMX standards to disseminate data and metadata to national users and international organisations.

[68] The SDMX sponsoring agencies are currently developing the *Checklist for the Implementation of SDMX in a Statistical Domain* which is intended to be a checklist of all the steps to be addressed when implementing an SDMX project [SDMX 2014, *Checklist for the Implementation of SDMX in a Statistical Domain*, p. 7, SDMX [draft version]]

The lead unit within an NSO and / or lead agency within the NSS would also take responsibility for organising a proof of concept pilot study that would demonstrate the benefits of SDMX implementation. This would help provide advocacy to attract high level attention, particularly at the national level for exchange of data between government agencies.

The establishment of effective coordination between statistical agencies within the NSS is a particularly critical issue where SDMX implementation is designed to improve efficiencies in the exchange of data (such as administrative data and perhaps related metadata) between one or more ministries and the NSO that may be used as input series to the compilation of official statistics such as national accounts. Involving other agencies within the NSS will require negotiation of many of the institutional, statistical, IT and resource related issues outlined in this section of the current document.

- Which statistical domains are to be involved in the SDMX implementation? For example, will initial SDMX implementation focus on one domain (for example, national accounts) or a number of statistical domains such as those encompassed in a range of short-term economic indicators. Countries may narrow their focus initially and decide on the initial implementation of one statistical domain using processes and SDMX artefacts that can subsequently be broadened (scaled up) to encompass additional domains (and perhaps more detailed data disaggregations) once experience on the use of SDMX technical standards and content guidelines has been gained.

  The strategic planning process could also envisage the statistical agency moving away from the paradigm of thinking in terms of specific publications (especially paper publications) and to reorientate the agencies business model from being mere producers / compilers of statistics to something else. This "something else" may include more dynamic exchanges of data built on automated pull/notification processes.

  For decentralised national statistical systems (NSSs) there may be issues relating to coordination between statistical agencies, especially when SDMX implementation is intended to cover a wide range of statistical domains or where data collection, compilation and dissemination are undertaken by different agencies.

2. Statistical issues

Statistical issues largely centre on:

- The conformity of the statistical indicators to be encompassed in the SDMX implementation, to international statistical standards as specified by international organisations such as IMF, OECD, UNSD and supranational bodies such as Eurostat. Conformity relates to national concepts / definitions, structural metadata and related code lists to international statistical standards.

- Identification of the appropriate Data Structure Definition (DSD) to be used in the SDMX implementation to which the national concepts, etc., will be mapped. Options here include the use of:

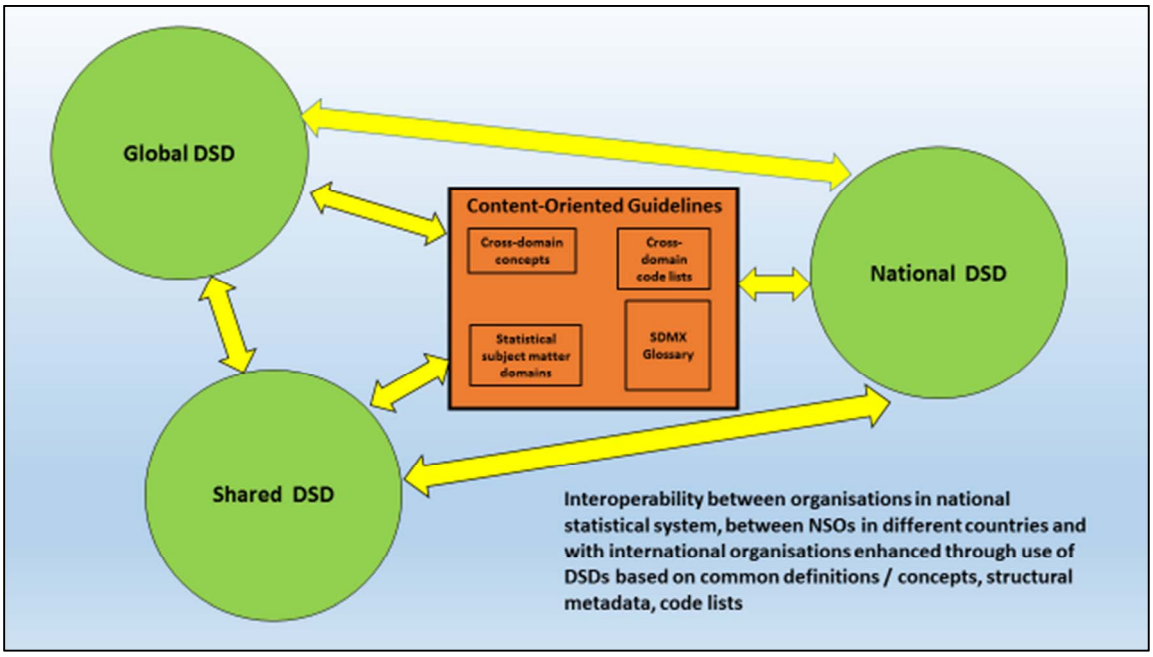
  - Global DSDs that have been developed for national accounts, balance of payments, foreign direct investment, government finance statistics, international merchandise trade, etc.

  - Shared DSDs that have been developed for OECD's short-term economic statistics, IMF's ECOFIN, UNSD's CountryData.

  - The development of national DSDs.

Where possible, countries are strongly encouraged to use either a global DSD or shared DSDs that maximise the possibilities for exchanging / reporting data with statistical agencies in other countries or with international organisations through the use of common structural metadata and code lists derived from the SDMX Content-Oriented Guidelines for both dimensions and attributes that comprise the DSD as illustrated in the following diagram.

Diagram 10: Interoperability of DSDs enhanced through use of concepts, etc., in the SDMX Content-Oriented Guidelines



- Consideration of a range of metadata-related issues on the existing structural and reference metadata environment for the statistical domains covered in the SDMX implementation initiative. Relevant issues include:

  - what metadata currently exists;
  - whether it is structured and uses a standard metadata template that covers a number of domains – refer below;
  - whether the national structural metadata concepts are consistent with those used in the SDMX Content-Oriented Guidelines, i.e. for the Cross Domain Concepts located in the SDMX Glossary;
  - where it is stored. In many countries metadata are frequently stored and disseminated in Word files or via PDFs.

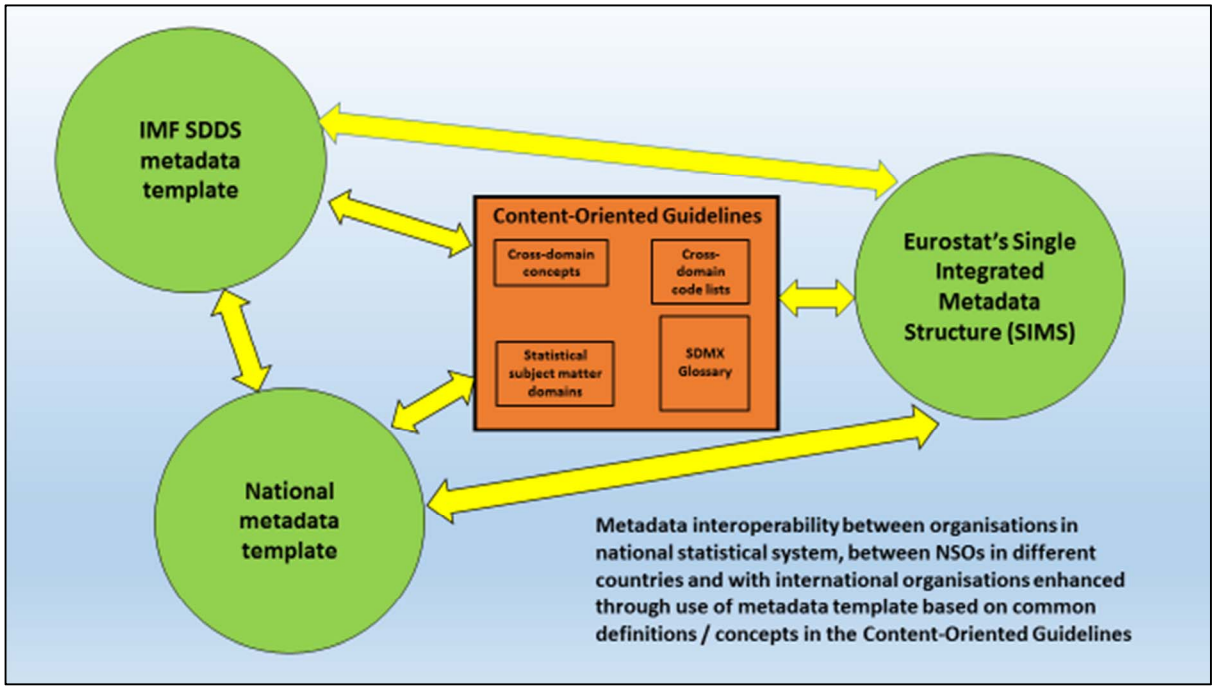
- The use of a standard metadata template containing standard metadata items that could be used for all statistical subject matter domains. Options for such a template are the use of:

  - the template used for the IMF's Special Data Dissemination Standard (SDDS);
  - Eurostat's Single Integrated Metadata Structure; or
  - a national metadata template.

Again, as in the case for DSDs, where possible, countries are strongly encouraged to use either an existing international metadata template or a national template based on structural and reference

metadata concepts derived from the SDMX Content-Oriented Guidelines. The use of such common concepts will also maximise the possibilities for the efficient exchanging / reporting of metadata with statistical agencies in other countries or with international organisations, as illustrated in the following diagram.

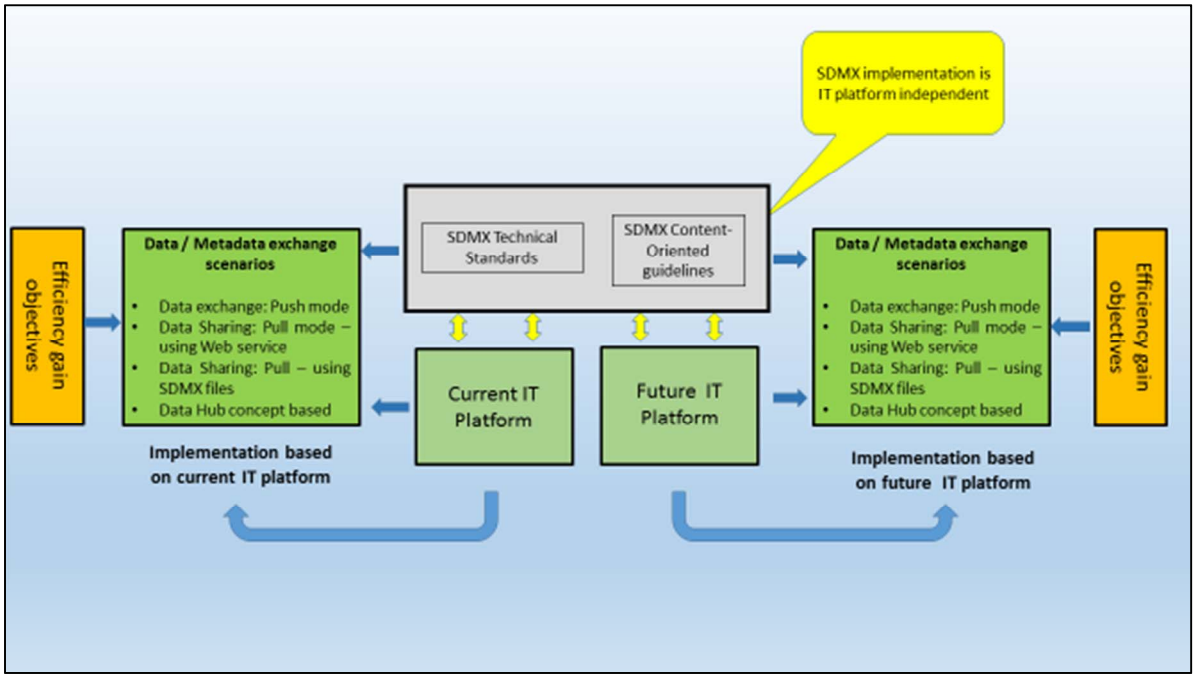
Diagram 11: Metadata interoperability enhanced through use of concepts, etc., in the SDMX Content-Oriented Guidelines



3. IT-related issues

When considering IT-related issues a clear distinction needs to be made between issues associated with SDMX Technical Standards and those associated with the IT platforms, which primarily concern the hardware (servers) where data are physically stored and the storage medium used by the agency, such as Excel spreadsheets, databases and data warehouses. SDMX implementation is to a large extent platform independent. As shown in the following Diagram, the expected efficiency gain objectives determine which of a range of data / metadata exchange scenarios the agency may select for SDMX implementation. Such considerations may be based either on the IT platforms the organisation currently has in place or on planned future platforms, resulting from, for example, the migration of data from Excel spreadsheets into a data warehouse.

Diagram 12: Relationship between SDMX Technical Standards and IT platform



The key and most time consuming process for SDMX implementation is the harmonisation of structural metadata that will facilitate the exchange of data. This is achieved by defining the necessary structural metadata (ideally in an MSD) that will describe data. Normally, data are already described by local concepts and code lists in their databases. The first step is to assemble information on the national database environment for each of the statistical domains being implemented. This will facilitate the mapping of local structural metadata used in the national system against those provided in the DSD.

In this context, information is required on:

- IT-related issues relevant to each of the statistical domains being considered for SDMX implementation, including:

  - the database environment where data are stored, e.g. in databases, data warehouses or Excel, CSV files. National statistical agencies frequently use more than one storage medium for the range of statistical outputs they disseminate. This multiplicity of database / storage environments extends even further where data flows across different organisations in the NSS are taken into consideration;
  - the national structural metadata and code lists used to define the data to be exchanged;
  - what data are currently reported to international organisations, who does this and how are data transmitted, i.e. each data flow needs to be identified, together with the level of statistical disaggregation provided to each organisation.

- Which structural metadata concepts (dimensions) are used to describe the datasets and where these metadata are currently stored. Data are very often described using different concepts and code lists by individual data providers at the national level and data collectors at the international level.

Generally, the data and structural metadata are stored in databases or in files (e.g. Excel, CSV, etc). These two cases could lead to different SDMX implementation approaches, such as:

- data and structural metadata continue to be stored in Excel or CSV files and the only need is to translate those files into SDMX-ML datafiles to be pulled by the data collector;
- data and structural metadata are already stored in a database and it is necessary to develop new or adapt existing suitable software interfaces to make the system software compliant;
- a separate special-purpose database is established to store the data and metadata.

Factors to be considered when deciding on which of the SDMX implementation tools reviewed in Section C.4 above are outlined in Step IV of the structured SDMX implementation process described below.

4. Skills-related issues

Skills-related issues centre on the preparation of an assessment of national agency financial, staff resource and capacity and skill development requirements for SDMX implementation. For example, technical assistance may be required from international agencies for initial access to implementation tools such as SDMX-RI – refer Step III below. Resources will only be forthcoming if a proper business case for SDMX implementation is developed in the context of corporate strategic plans and the budget preparation cycle.

Preparation for a small-scale pilot SDMX implementation

This sub-section provides practical advice to agencies preparing a small-scale pilot SDMX implementation. As mentioned above, such an exercise is the best means for an agency to acquire an understanding of the SDMX technical standards and guidelines and SDMX implementation tools, and to develop the capacity of staff within the organisation who may be involved in subsequent full-scale implementation.

The issues to be considered by agencies mounting such an exercise are a microcosm of those outlined above in the context of full-scale SDMX implementation, though on a narrower scale. Practical tips / advice for developing a pilot exercise based on the experience of a number of countries who have been there before are provided below:

Practical tips

1. Never underestimate the amount of time and resources required to prepare for, conduct and evaluate the results of the pilot exercise. The amount of time, etc., required varies inversely with the ambitiousness of the pilot with respect to the domain to be included; the nature of the SDMX exchange process to be tested; the range of different parts of the organisation who will be involved in the pilot (IT, statistical subject matter; training; dissemination; coordination; etc); and the number of other agencies within the NSS who will be involved in the pilot project.

2. Do not be too ambitious in the scale of the project with respect to the issues outlined above. A common mistake is for pilot projects to be too ambitious and as a result either become an end in itself, or a dead end. If there is a sense that the pilot is becoming too ambitious and complex then an early decision needs to be made to scale it back.

3. Formulate a clear set of objectives for the pilot project at an early stage. These could include: firming up estimates of the efficiencies that may be achieved by full-scale SDMX implementation. Such information will help develop a sound business case justifying the expenditure of resources for such implementation; developing the technical knowledge and organisational capacity (in the context of SDMX) of staff who may be involved in the full-scale

implementation of SDMX further down the track; to identify the objectives of full-scale implementation in relation to the nature of the data / metadata exchanges it will encompass; to prepare more accurate estimates of the resources that will be required for full-scale implementation and the time-scale for implementation; or to determine to most appropriate means of implementing full-scale implementation, e.g. incrementally, etc.

4. The aim of the pilot project may be to identify potential efficiency improvements in the transmission of data (and metadata) between agencies within the NSS, say for example, data that may be used by the NSO as input series for the compilation of official statistics such as national accounts. In this case it would be appropriate to involve only one other agency / ministry, etc., in the pilot project. The reason for narrowing the number of other agencies involved in the pilot project is that the time required for negotiations with other agencies on the institutional, statistical, IT and resource related issues outlined above, and the amount coordination effort required, should not be underestimated. The experience gained in such negotiations and coordination mechanisms with a single other agency within the NSS can then be applied when other agencies are involved in subsequent, post-pilot SDMX implementation.

5. Identify where the skills required for conducting the pilot project are to be obtained. Obviously, staff within the organisation will undertake the bulk of the work involved in conducting the pilot. Options include: from within the organisation; other agencies within the country; statistical agencies in other countries; or from international / supranational organisations.

6. Ensure that the pilot project outcomes can be linked to full-scale implementation. The ideal outcome would be that full-scale implementation merely requires scaling-up the modalities of the pilot project, for example, to encompass other statistical domains, data flows, exchanges with other organisations in the national statistical system. However, in reality, the outcomes of the pilot may not be so simple and alternative parameters may need to be considered.

7. Undertake an objective analysis of the outcomes of the pilot, the lessons learnt, mistakes, strengths, etc. These will be used to plan and prepare for full-scale implementation.

Issues to be considered when preparing for a pilot project

As mentioned above, the issues to be considered when preparing for an SDMX pilot project are a microcosm of those required for full-scale SDMX implementation. Even at the risk of duplicating the information provided above it is worthwhile restating these. They may be broken down into priority issues which need to be considered at a very early stage, and other issues.

A. Priority

1. Which statistical domain is to be included in the proposed pilot implementation? Factors to be considered here include the existence of a relevant global or shared DSD for the domain being considered. As mentioned above in the discussion on DSDs, global DSDs are quite complex (for a number of reasons including the large number of dimensions that require mapping) and may not be a good choice for a pilot study. CountryData and ECOFIN DSDs may be a more appropriate choice for an initial pilot study as they are much easier to understand.

2. How would SDMX be used in the case of the statistical domain involved in the pilot project? For example, would it be used for:

    - data and metadata reporting to an international organisations such as the IMF, UNSD, etc;
    - improving the efficiency of data and metadata exchange between statistical agencies within the national statistical system;

- aggregate data and its related metadata dissemination to external users via the organisation's website;
- improving the storage of structural and reference metadata?

3. Which unit / area within the organisation would be the primary contact and main driver for the pilot implementation project - e.g. IT unit; training area; subject matter area; data dissemination or coordination units?

B. Other issues to be considered

1. Institutional issues and pilot study objectives

Issues here include linkage of the SDMX pilot project to the organisation's strategic planning processes / objectives such as those outlined in an NSDS, statistics master plan, etc. Ideally, SDMX implementation should be considered within the framework of existing corporate / national statistical system (NSS) institutional and technical (statistical, IT, infrastructure) strategic planning processes. It is an element of modernisation of statistical infrastructures, etc.

2. Statistical issues

Statistical issues largely centre on the conformity of the indicators to be included in the SDMX pilot project to international statistical standards as specified by international organisations (for example, in the case of national accounts this standard would be SNA 2008). Initial metadata-related issues to be considered include the existing metadata environment for the statistical domain to be covered in the SDMX pilot project – such as:

- what metadata currently exists for the statistical domain in the organisation;
- whether the metadata is structured and uses a standard metadata template that covers a number of statistical domains;
- whether the national structural metadata concepts are consistent with those used in the SDMX Content-Oriented Guidelines;
- where the metadata are stored? In many countries metadata are frequently stored and disseminated in Word files or via PDFs.

3. IT-related issues

The key and most time consuming process for SDMX implementation is the harmonisation of structural metadata that will facilitate the exchange of data. This is achieved by defining the necessary structural metadata that will describe the data being exchanged in the pilot project. Normally, data are already described by local concepts and code lists in national databases or on Excel spreadsheets. The first step therefore is to assemble information on the database / data storage environment in the organisation for the statistical domain being considered for the SDMX pilot study. For example:

- The database environment where data are stored within the organisation, e.g. in databases, data warehouses or Excel, CSV files. National statistical agencies frequently use more than one storage medium for the range of statistical outputs they disseminate. This multiplicity of database / storage environments extends even further where data flows across different organisations in the NSS are taken into consideration.

- The national structural metadata and code lists used to define the data to be exchanged.

- What data / metadata are currently reported between agencies for the statistical domain envisaged for the pilot study, who does this reporting from the organisation and how are data transmitted, i.e. each data flow for the (part) statistical domain being implemented for the pilot project needs to be identified, together with the level of statistical disaggregation provided?

- Which structural metadata concepts are used to describe the datasets and where are these metadata currently stored? Data may be described using concepts and code lists that differ from current international recommendations and guidelines. Data and structural metadata are stored either in databases or in files (e.g. Excel, CSV, etc).

The final IT-related task is to identify the SDMX implementation tools that will be used for the pilot project. These are described in Step I.3 above. Issues to be considered in selecting the tools most appropriate for the organisation are outlined in Step IV below.

4. Skills-related issues

Skills-related issues centre on the identification of the organisation's capacity and skill development requirements for SDMX implementation. Capacity development may be required for staff from a range of different areas within the organisation: IT area, statistical domain specialists, and staff in data dissemination and coordination units, all of whom may need to be involved in the development and implementation of the SDMX pilot project. For example, technical assistance may be required for initial access to implementation tools to be used for the pilot study and beyond, such as Eurostat's SDMX-RI.

STEP III: IDENTIFICATION OF NATIONAL SDMX CAPACITY DEVELOPMENT NEEDS AND CAPACITY DEVELOPMENT DELIVERY

National agency consideration of the issues outlined in Step II above will assist in the identification of specific national SDMX capacity development needs on IT, statistical, institutional and metadata-related issues, which in turn helps identify capacity development delivery.

Capacity development required for SDMX implementation falls into three broad categories which comprise developing an: understanding of all objects in the SDMX information model; specifically the SDMX information model Concepts and Concept Scheme; and developing the capacity to access, install and utilise SDMX implementation tools.

1. Develop an understanding of all objects in the SDMX Information Model

The SDMX Information Model (SDMX-IM) describes a set of formal objects which are the building blocks of a standard view of the data and metadata exchange process. The SDMX-IM includes the following SDMX objects or in SDMX terminology "artefacts":

- Concepts and Concept Scheme
- Code lists
- Data structure definition (DSD)
- Data set
- Metadata structure definitions (MSD)
- Metadata set
- Dataflow and Metadataflow definition
- Data provider

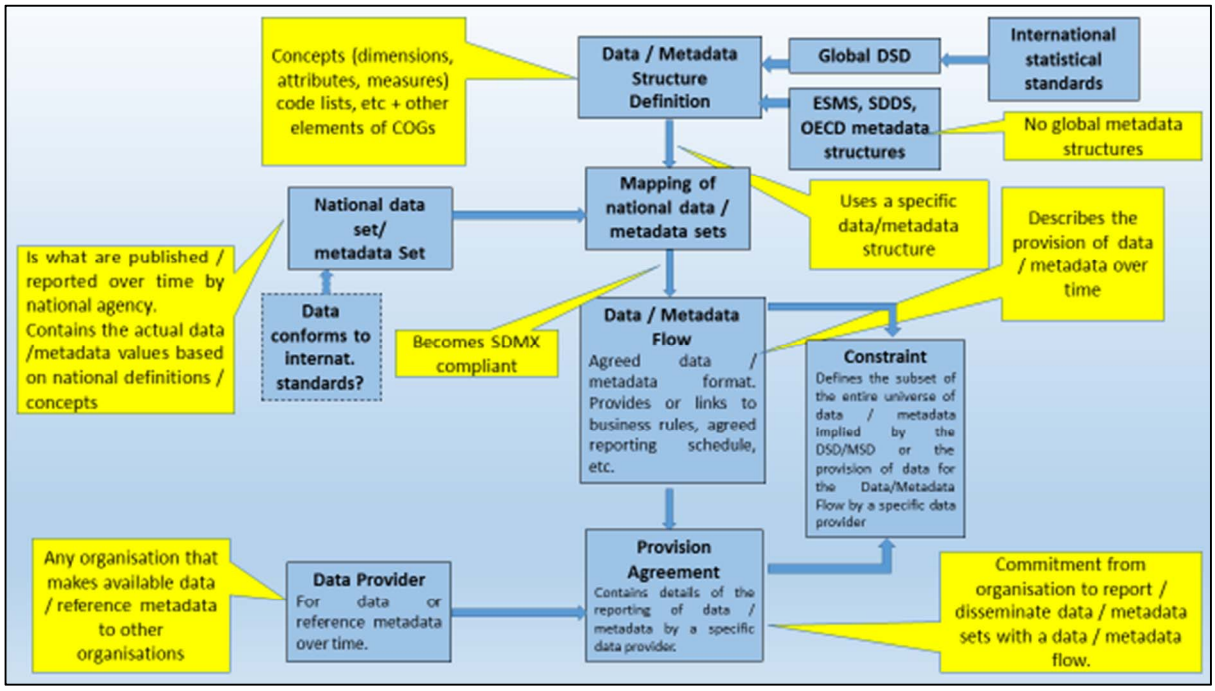
The concepts used in DSDs and MSDs must have a specific set of properties according to the SDMX-IM. These properties include the: description of the concept (i.e. the definition); identifier of the concept; code-list if the concept is coded; and the maintenance agency which maintains the concept for use within the DSD and MSD.

A relatively non-technical description of each of these objects is provided in the student book for Eurostat's *SDMX Self-Learning Package No. 2: The SDMX Information Model*[69]. The target audience in a national statistical agency for developing a detailed understanding of all the objects in the information would be IT or data dissemination specialists.

The relationship between the main artefacts in the SDMX-IM is shown in the following Diagram.

[69] Refer https://webgate.ec.europa.eu/fpfis/mwikis/sdmx/index.php/Main_Page

Diagram 13: Relationship between the main components of the SDMX Information Model (SDMX-IM)



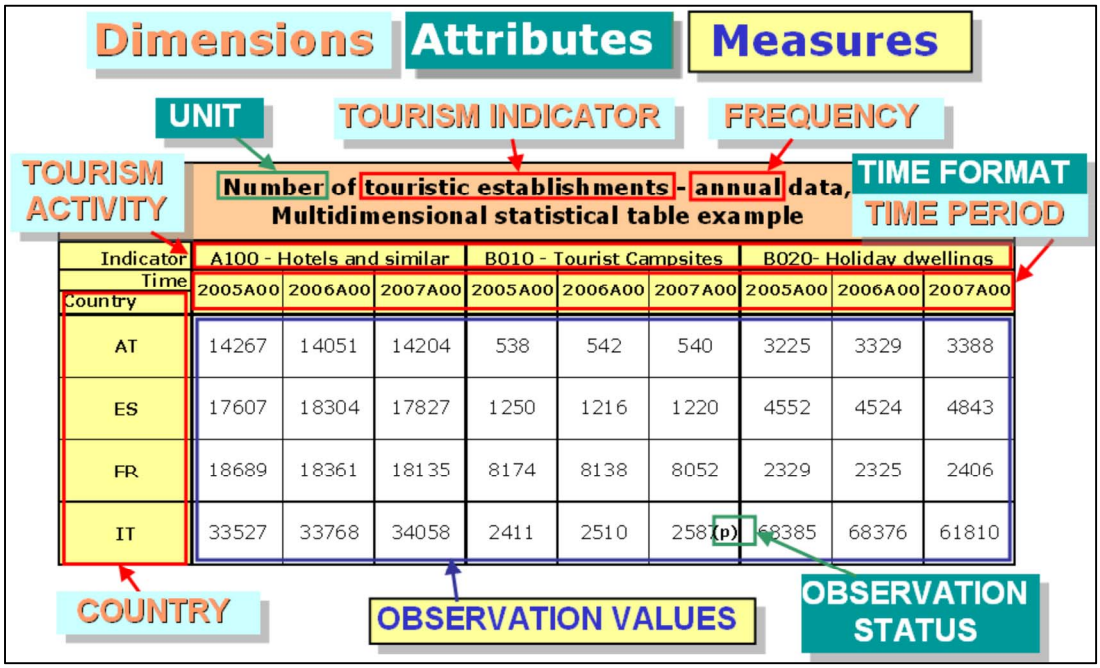
2. Develop an understanding of SDMX Concepts and Concept Scheme

Whilst it would be essential for IT / data dissemination specialists to have such an understanding of all the objects in the SDMX-IM it is also essential for statistical domain specialists in a national agency to have a similar depth of knowledge of at least the elements in the first object, Concepts and Concept Scheme, i.e. dimensions, attributes and measures. SDMX implementations are more likely to succeed or extend beyond the pilot testing stage if knowledge of SDMX concepts / concept schemes, etc., extends beyond the IT departments of national statistical agencies.

For example, the involvement of domain specialists with their detailed knowledge of the specific underlying concepts embedded in the multidimensional statistical tables subject to exchange / dissemination is essential for the accurate reformatting of these tables into SDMX concept structures. Examples of such multidimensional tables are provided in the following diagrams.

Finally, it is essential for countries to avoid having knowledge of SDMX restricted to only one or two individuals in the national statistical agency.

Diagram 14: Elements of the SDMX Concept Scheme in a multidimensional statistical table



Source: Student book for Eurostat's SDMX Self-Learning Package No. 2: The SDMX Information Model

Such a Concept Scheme outlined in the above example of a multidimensional table may be reformatted into an SDMX Concept Structure as shown in the following Diagram.

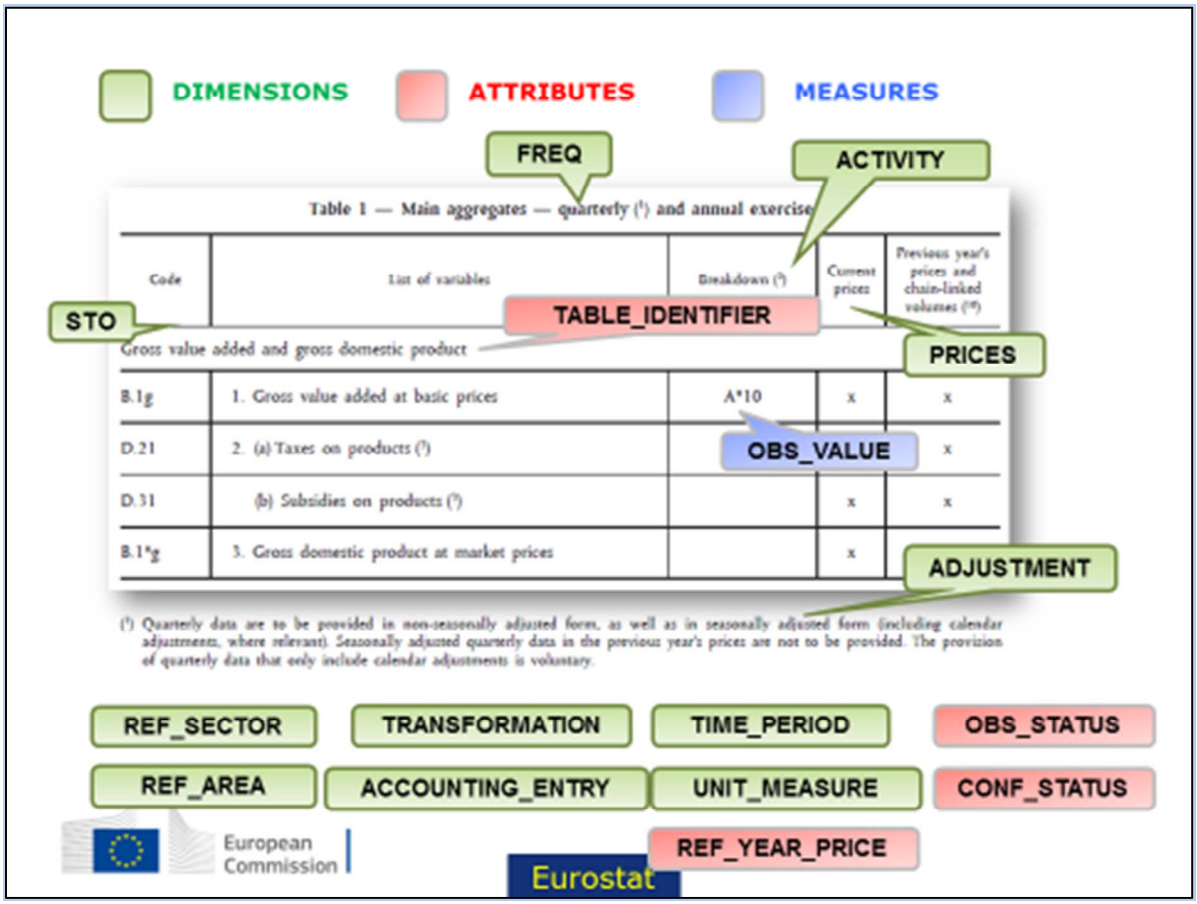
Diagram 15: SDMX Concept Structure derived from national dataset

Concept structure for the Multidimensional table example

Key	Concept ID	Concept Name	Attachment level	Usage status	Code List ID	Code List Name
DIMENSIONS						
1	FREQ	Frequency		Mandatory	CL_FREQ	Frequency code list
2	COUNTRY	Tourism Country		Mandatory	CL_COUNTRY	Country code list
3	INDIC_TO	Tourism Indicator		Mandatory	CL_TOUR_INDICAT	Tourism Indicator code list
4	ACTIVITY_TO	Tourism Activity		Mandatory	CL_TOUR_ACTIVITY	Tourism Activity code list
	TIME_PERIOD	Time period		Mandatory		
MEASURES						
	OBS_VALUE	Observation value		Conditional		
ATTRIBUTES						
	OBS_STATUS	Status of the observation	Observation	Conditional	CL_OBS_STATUS	Observation status code list
	UNIT	Unit	Series	Mandatory	CL_UNIT	Unit code list
	TIME_FORMAT	Time format	Series	Mandatory	CL_TIME_FORMAT	Time format code list

Similarly, a data structure may be derived from the national accounts table in the following diagram.

Diagram 16: Deriving an SDMX Concept Scheme from a national accounts statistical table



Source: D. Suranyi, presentation, *Introduction to SDMX*, given at an ESTP Workshop on SDMX in National Accounts, Luxembourg, 4-6 March 2015

As in the previous example, an SDMX Concept Structure may be similarly derived from this table.

Diagram 17: Deriving an SDMX Concept Structure from a national data table

Data Srucutre Definition	
ID	NA_MAIN
Agency	ESTAT
Version	1.0
Name	Structure for NA Main tables

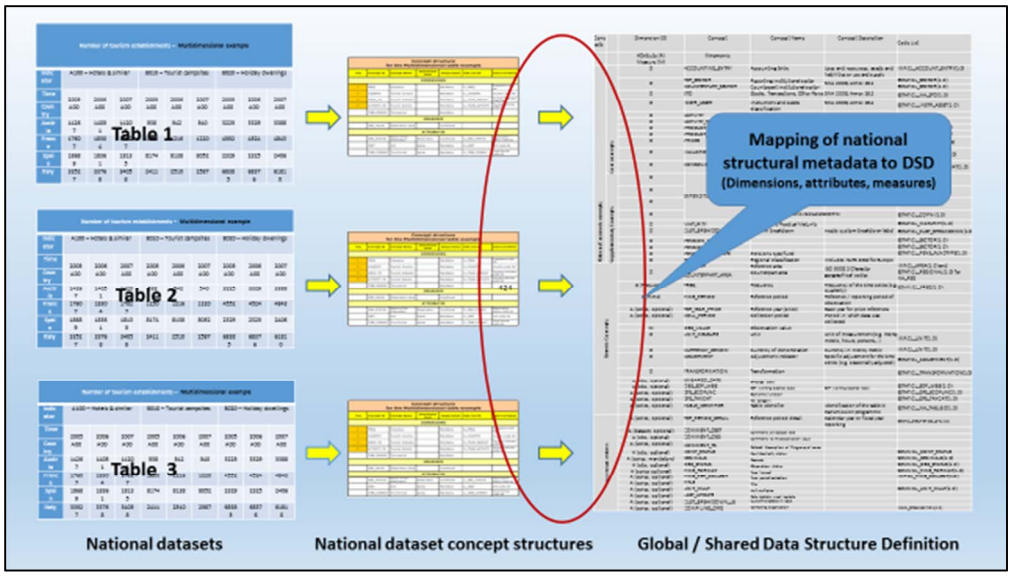
Concept ID	Concept Name	Code list ID	Role	Attachement level	Code list ID
FREQ	Frequency	CL_FREQ	Dimension		CL_FREQ
REF_AREA	Reference area	CL_AREA	Dimension		CL_AREA
ACCOUNTING_ENTRY	Accounting Entry	CL_ACCOUNTING_ENTRY	Dimension		CL_ACCOUNTING_ENTRY
ACTIVITY	Activity classification	CL_ACTIVITY	Dimension		CL_ACTIVITY
ADJUSTMENT	Adjustment indicator	CL_ADJUSTMENT	Dimension		CL_ADJUSTMENT
PRICES	Prices	CL_NA_PRICES	Dimension		CL_NA_PRICES
REF_SECTOR	Reporting institutional sector	CL_SECTOR	Dimension		CL_SECTOR
STO	Stocks, Transactions, Other Flows	CL_NA_STO	Dimension		CL_NA_STO
TRANSFORMATION	Transformation	CL_TRANSFORMATION	Dimension		CL_TRANSFORMATION
UNIT_MEASURE	Unit and derived data coding	CL_UNIT	Dimension		CL_UNIT
TIME_PERIOD	Reference period		Dimension		
OBS_VALUE	Observation value		Measure		
OBS_STATUS	Observation status	CL_OBS_STATUS	Attribute	observation	CL_OBS_STATUS
CONF_STATUS	Confidentiality status	CL_CONF_STATUS	Attribute	observation	CL_CONF_STATUS
REF_YEAR_PRICE	Reference year (price)		Attribute	series	
TABLE_IDENTIFIER	Table identifier	CL_NA_TABLEID	Attribute	series	CL_NA_TABLEID

European Commission

Eurostat

The relationship between these national datasets, the SDMX concept structure derived from those national datasets and the mapping of these elements to the global or shared DSD is illustrated in the following diagram.

Diagram 16: Relationship between national datasets, SDMX concept structures and the mapping process



3. Develop the capacity to access, install and utilise SDMX implementation tools

Developing the capacity of national agency staff to access, install, utilise and maintain the implementation tools selected by the country for its SDMX implementation is crucial for achieving the long-term goals and objectives of the project. Experience to date has shown that utilisation of the tools described in Step I.3 above is not straight-forward and requires a high level of expertise and knowledge of both the SDMX Information Model and the extensive range of options, etc., each suite of tools provides. Such expertise currently resides with international organisations, in particular, those that developed the implementation tools, with a relatively small number of national statistical agencies – refer Annex 1 below, and with a small number of private organisations that have developed applications that assist the implementation of SDMX – refer Step I.3.f above.

4. Capacity development delivery

As mentioned previously, capacity development and the provision of technical assistance by international agencies may be required in the area of initial access to and use of SDMX implementation tools[70]. This may also take the form of demonstrating the use of such software and its component tools to NSO / NSS staff. Following initial implementation, it is expected that staff in the national lead agency (such as the NSO) would take responsibility for subsequent capacity development across the NSS as required. The lead agency would also take responsibility for national ongoing maintenance of the implementation software as well as updates and other software changes developed by the source international agency.

The last issue to be considered is responsibility for capacity development delivery. Although the national agency has primary responsibility for skill development, actual delivery may require the provision of technical assistance by national partner agencies and/or by international organisations such as ESCAP, IMF, etc.

The process for the identification of capacity development needs requires the completion of a matrix such as the one outlined in the following table. The issues listed in the first column of the table were derived from the NSS priority development needs outlined in Table 1 in the discussion on the development of a business case for SDMX implementation outlined in Section B above.

Table 2: Capacity development matrix

Issue	Specific SDMX capacity development need	National capacity development target(s)	Capacity development delivery
Advocacy			
Coordination			
Infrastructure	Understanding of all objects in the SDMX Information Model	IT, data dissemination specialists	
	Understanding of SDMX Concepts and Concept Scheme	Statistical domain specialists	
	Capacity to access, install, utilise and maintain SDMX implementation tools	IT, data dissemination specialists	

[70] Experience to date with SDMX implementation using SDMX-RI by EU Member states, the IMF's Open Data Platform in African countries and the MDG DevInfo pilot project has shown that such support by the international lead agency (Eurostat, IMF and UNSD) has been essential.

STEP IV: DETERMINING WHICH SET OF SDMX IMPLEMENTATION TOOLS TO USE

Given the heterogeneity of statistical agencies in countries around the world in terms of the statistical, institutional and IT environments identified in Step II above, it is unlikely any one of the existing sets of SDMX tools will be appropriate to the needs of all agencies. National agencies may opt to use one or more of these existiIn ng tools or develop their own. Unfortunately, there is very little information currently available that compares the advantages / disadvantages of each of the implementation tools reviewed. In this context, issues that agencies need to consider when reviewing the suitability of existing tools include:

- stage / status of their development. The tool(s) being considered should be "operational" and not be at the pilot or development stage;

- ability of tools to link to a heterogeneous range of national database environments, in particular, those in developing and small-sized NSSs. Many statistical agencies have a number of different databases for the various statistical domains they compile and disseminate. These can range from data warehouses to Excel spreadsheets;

- cost and/or adherence to the open source principles of the adopting agency;

- availability of both technical and non-technical documentation outlining functionality and use of the tools. Documentation should also outline in detail the prerequisites for the use of mapping tool. The availability of these resources is a major factor determining whether or not SDMX implementation can be undertaken with or without extensive technical support from international organisations responsible for the development of the implementation tool;

- whether or not the implementation tool can be used in the IT environment currently in place in the implementing agency. In particular, the storage medium where the data sets in the statistical domain(s) being implemented reside, such as Excel spreadsheets;

- availability of training courses and online tutorials;

.

- suitability of IT architecture. For example: if the tool is cloud-based are there technical or political constraints; if the agency uses a service-oriented architecture (SOA), is the tool adaptable?

Downloading and use of selected SDMX implementation tool

Following the selection of the SDMX implementation tool appropriate to the needs of the national agency the tool would be downloaded. The download may be the precursor to full-scale / ongoing SDMX implementation for one or more statistical domains, or merely a proof of concept pilot exercise involving a statistical domain using one of the global DSDs listed in Annex 2 below to validate the utility of the tools or to demonstrate the potential benefits of SDMX to stakeholders within the national agency, such as senior management.

Links to the SDMX implementation tools described in Step I.3.f above are provided in the following table.

Table 3: Access to SDMX implementation tools

Implementation tool	Location

Implementation tool	Location
SDMX Converter	https://webgate.ec.europa.eu/fpfis/mwikis/sdmx/index.php/Main_Page
SDMX-RI[71]	https://webgate.ec.europa.eu/fpfis/mwikis/sdmx/index.php/Main_Page
IMF SDDS Plus	http://www.imf.org/external/pubs/ft/sdds/guide/plus/2014/sddsplus14.pdf
United Nations DevInfo	http://www.devinfo.org/libraries/aspx/home.asp
Tools developed by:	
- Metadata Technology	http://www.sdmxfusion.com/products.html#Registry
- Space Time Research	http://spacetimeresearch.com/
-	

[71] A Powerpoint presentation "How to set up an SDMX architecture for reporting and dissemination", was given by Francesco Rizzo (ISTAT) in the SDMX training course held in Korea following the 2014 meeting of the SDMX Expert Group in October 2014, refer http://www.oecd.org/std/SDMX-2014-F-Rizzo-ISTAT-2.pdf

STEP V: PARTICIPATION IN SDMX NETWORKS

As can be seen from the above, SDMX implementation is a long and technically complex process and sharing the difficulties is certainly a good way to alleviate the burden; furthermore it can improve cross-border harmonisation. Countries with similar characteristics (in terms of economic development, size, socio-economic situation, geographic location and with a common need to report to the same international organisations, etc.) might join efforts to achieve economies of scale. Such cooperation may take various forms: formal committees, ad hoc task forces, informal teleconferences, etc. Some countries may already be collaborating on other issues; the mandate of these groups could perhaps be extended to cover SDMX implementation. Joining efforts could certainly be a strong argument to obtain support from international organisations as this would mean economies of scale for them as well.

A key objective of the joint ADB-ESCAP SDMX initiative referred to in Section A above is the promotion of SDMX networks that could be used by countries in the Asia-Pacific region implementing SDMX standards to share their experiences and knowledge. Obviously, such networks could also be used by countries in other regions. In order to avoid the creation of yet another collaboration tool the intention is to link into the existing global tools and networks, in particular, those that can be accessed from the redesigned SDMX.org website.

Summary information for implementations by international organisations and examples of selected recent implementations by national agencies in countries around the world is provided in Annex 1 below. These implementations have been developed using the tools described above or other SDMX tools. The initiatives utilise SDMX technical standards and guidelines to facilitate the broad and narrow SDMX uses outlined above. The implementations summarised in Annex 1 are restricted to those that are "operational", or will become operational in the near future, and which have progressed beyond the pilot or feasibility stage of development. Links are provided to more detailed information where available, from sources such as the SDMX.org website.

Some of these implementations may be used by other agencies contemplating SDMX implementation as proof of concept to demonstrate the benefits of broader SDMX implementation, particularly for the benefit of senior management of the agency.

The extent of involvement by national agencies in global or regional networks is a continuum depending on national circumstances and resource availability, and may take the form of:[72]

- collaboration with more advanced systems in Europe and elsewhere that may have experience in SDMX implementation. There are a number of mechanisms for such collaboration. For examples, countries could apply to be part of the SDMX Technical Working Group or the Statistical Working Group[73];

- establishing links to other sub-regional organisations (e.g. ASEAN, SPC, SAARC) using these forums for advocating for the priorities identified and for obtaining the support from heads of NSOs as well as finding linkages to their work programmes;

- merely providing summary information on their SDMX implementation, perhaps with links to more detailed documentation;

- contributing to experience and data sharing platforms and discussion forums;

[72] Refer draft report from 2014 MSIS meeting, paras. 15-16, 18 and 30-32.

[73] National agencies wishing to do so should contact secretariat@smdx.org for further information.

- providing technical assistance / advice by email;
- providing technical assistance / advice through visits of experts to implementation countries and / or staff exchanges;

- participation in virtual expert group meetings on specific topics by teleconference;

- contributing to experience and data sharing platforms and discussion forums;

Finally, UNESCAP or other regional bodies may organise focused and technical workshops on priority issues. Such workshops may be organised at the margins of other meetings taking place in the region such as MSIS, SDMX Expert Group meetings or SDMX Global Conferences.

ANNEXES

Annex 1: Examples of international and national agency SDMX implementations[74]

[To be expanded as more information is obtained]

Country	Agency	Usage	Detail	Contact
International agencies				
NA	OECD	Collection of short-term economic statistics (STES)	Developed a STES DSD. Used SDMX-RI implementation tools. OECD member countries were invited to participate in this exercise in 2011 and more than 20 countries are now participating in this exercise.	David Barraclough
NA	ILO	Collection and dissemination of labour statistics	The ILO is developing a number of SDMX implementations involving labour statistics: . Automation of data reporting. The Organisation has implemented an SDMX API to access any information in the ILOSTAT database and also to collect data received in SDMX files / messages provided they are based on DSDs developed by the ILO. The ILO has prepared detailed documentation on the system[75]. . For data interchange with other international agencies the ILO has commenced a pilot project with Eurostat to receive in SDMX-ML the annual data EU Member states report to Eurostat in lieu of the ILO's yearly indicators collection for those countries. This interchange currently takes place using Excel worksheets. From the ILO, the SDMX API has all that is needed. It generates the DSD for each of the indicators to be reported and is also able to upload an SDMX dataset provided it is based on one of their DSD. Eurostat is evaluating how to generate the files, most likely with the SDMX-RI suite of tools. . The ILO is working with INEGI (Mexico) to enable their annual reporting to be done through SDMX instead of the traditional Excel questionnaire. The ILO merely provided	Edgardo Greising greasing@ilo.org

[74] This Annex may be removed from subsequent versions of the current document and relocated on the web where it can be more readily updated as further information becomes available.

[75] *ILOSTAT SDMX Gateway: Dissemination API Reference Guide*, Department of Statistics, ILO, February 2014 – [not yet available online]

Country	Agency	Usage	Detail	Contact
			their DSDs and INEGI is working in producing the datasets from their data warehouse [tools they are using are not known]. ILO experience is that most developing countries do not store their time series in data warehouse. There is a tendency for these countries to disseminate data in Excel, PDF or HTML. Also, metadata are not always available with the data collected. To overcome this the ILO is currently working on a project with DevInfo to generate an ILOSTAT template for DevInfo 7 that would allow countries to define a local repository for ILOSTAT indicators comprising both data and metadata and then transmit the information to the ILO in SDMX. The intention is to test the DevInfo implementation in ten countries around the globe. The ILO is currently in the process of clearing the "administrative" steps towards issuing the contract to make the adaptation described above. The idea is that countries would be able to feed a DevInfo database with labour indicator's data and metadata and then DevInfo will produce the SDMX files according to the ILO DSDs. That means that the national repository will be the DevInfo database, and the "mapping" to ILO DSDs will be intrinsic to it. So no external mapping is needed. One advantage is that the country will be able to use all DevInfo features to disseminate the information in their websites. [Source: ILO]	
NA	UNSD	Millennium Development Goal (MDG) reporting and dissemination	UNSD implemented a pilot project using the SDMX standards for data and metadata exchange using DevInfo database technology to exchange data between national and global levels. Pilot involved eleven countries – see below.	Abdulla Gozalov gozalov@un.org
		UNdata API	The UNdata API is powered by Eurostat's SDMX-RI. The API is implemented as a SOAP Web Service that can be used to query the datamarts using the SDMX standard. UNSD intend to gradually move all of its databases onto the API.	
NA	World Bank	Collecting and maintaining external debt data - Debtor Reporting System (DRS)	Developed jointly with the Commonwealth Secretariat (COMSEC) and UNCTAD since the target countries already utilised their debt systems. The new system is designed to streamline the reporting to DRS from borrowing countries. The initiative developed its own implementation tools. DSD dimensions such as COUNTRY, CURRENCY, etc., were derived from the SDMX Content-Oriented	

Country	Agency	Usage	Detail	Contact
			Guidelines since they are used in every topic. The initiative also had to develop their own dimensions to explain DRS's loan-by-loan data, such as CREDITOR TYPE, DEBTOR TYPE, PURPOSE of the loan, etc. Countries are currently migrating their data into the latest version of the CS-DRMS at which time they will be able to report to the DRS using SDMX standards. [Source: World Bank]	
NA	UNESCO	Uses SDMX to facilitate the exchange of statistics across a number of domains with other international organisations.	UNESCO has implemented numerous data exchanges with other international organisations in the Output/Dissemination workflow. This has all been done using customised DSDs (e.g. for ICT and Culture) for the specific exchange as opposed to global DSDs. The latter are being developed for R&D and education with the OECD and Eurostat and could be approved in 2015. There are significant opportunities to improve data quality and process efficiency related to collecting and sharing data with OECD and Eurostat. This is a key focus for UNESCO with this initiative. Providing a means for countries to provide data via SDMX is a secondary objective. The principle method that UNESCO will be deploying for these projects is SDMX-injected Excel questionnaires. These questionnaires will be passed through the Eurostat Excel-to-SDMX converter to produce SDMX-ML files. The efficiency gains in the three organisations will require putting in place data exchange hubs (2015 hopefully but could slip to 2016) which can be used to share data with each other. Once these hubs are in place, extending the solution outward so that countries can submit data to one of these hubs will be a rather straightforward task. Having the capability of countries providing SDMX-ML files directly will be initiated by UNESCO with a small group of African countries in 2015. [Source: UNESCO]	Brian Buffett - b.buffett@unesco.org
National agencies				
Australia	Australian Bureau of Statistics (ABS)		Much of ABS key infrastructure was built to suit a particular need, i.e. to progress data through the various stages of the statistical cycle from survey conception to publication[76]. Due to the nature of development processes and the progressive selection of multiple IT products (Lotus Notes, Blaise, Oracle, Microsoft, SAS, etc.) over an extended period of time, most ABS systems are “stove piped” around statistical operations, e.g. household verses economic versus census collections. Many also use technologies that	

[76] For further information refer http://www.abs.gov.au/websitedbs/d3310114.nsf/4a256353001af3ed4b2562bb00121564/7057eb9a73a186c4ca2576c00018b2ba!OpenDocument

Country	Agency	Usage	Detail	Contact
			are now a generation out of date. Relational databases, SQL queries, cathedral system developments and proactively minimising storage requirements (and hence the quantity of metadata kept) to minimise costs are very much the norm. At an operational level the systems have been appended and adapted over the years to meet emerging needs on an ad-hoc basis as opposed to an overall vision of system integration. The result has been islands of coherence amidst an ocean of information where metadata cannot flow freely within the ABS or outside into the broader statistical community. The ABS Information Management Transformation Program (IMTP) is an enterprise level re-engineering program that aims to support a: • Client environment where statistics are readily available and can be easily integrated with data from other sources. • Statistical production environment that is highly productive and satisfying for staff to use. • Statistical development environment that is flexible; • Systems environment that is built around standard models and supports shared collaborative development, including internationally. SDMX will be one of the tools to support such end-to-end integrated statistical production processes. Because many ABS statistical activities involve micro-data, SDMX will need to work with other standards such as DDI-L which are relevant to earlier phases of the UNECE Generic Statistical Business Process Model (GSBPM).	
Brazil	IBGE			
Cambodia		Pilot country for UNSD MDG project	See above	
Italy	Istat		Use SDMX to disseminate data from a single web application in a distributed environment where data are not stored centrally.	
Korea, Republic of	Statistics Korea	Data provision to international organisations and an NSS data	Developed the Korean Data Provision System (KODAPS) to provide data to international organizations – OECD, United Nations and other international agencies.. KODAPS uses DSDs provided by I/Os. Functionality included in KODAPs includes mapping.	

Country	Agency	Usage	Detail	Contact
		dissemination system.	KOSIS is the national statistics database developed by Statistics Korea. Statistics Korea and is the gateway for official statistics produced in the country's NSS as well as international and North Korean statistics. The database contains data produced by 120 statistical agencies covering over 500 statistical subjects. Using SDMX standards Statistics Korea has developed the KOSIS Data Sharing Service (Open-API) which are open to the public so that the public and private developers can build their own services in the statistical databases in the KOSIS. The Open-API allows KOSIS users to automatically retrieve and utilize KOSIS data (e.g. vis mobile app.). Information available via the Open-API includes: lists of statistics available; metadata (structural and reference); as well as statistical tables. The aim is to provide a centralised data service system which overlays the country's decentralized statistical production system.	
Lao PDR		Pilot country for UNSD MDG project	See above	
Mexico	INEGI		SDMX implementations are a mix of projects and continuous processes. Available flows on INEGI's SDMX section located on the organisations website include: Short-Term Economic Indicators (STEI); Short-Term Economic Statistics (STES); Quarterly GDP (using global national accounts DSD); Goods Commercial Trade of Mexico (using UNSD's DSD to provide information to COMTRADE); Millennium Development Goal Indicators; Infra-Annual Labour Indicators (quarterly, monthly) pilot test of new dataflows for ILO.	
Netherlands	Statistics Netherlands		Do not use SDMX for internal purposes, only for exchange	
New Zealand	Statistics New Zealand			
Norway	Statistics Norway			
Russian Federation	Rosstat			
Switzerland	FSO		The Swiss Federal Statistics Office opted not to use SDMX-RI and has developed their own implementation tools.	
Thailand		Pilot country for UNSD MDG project	See above. The platform used by Thailand was StatXChange rather than DevInfo. This platform has been used for the last five years for exchange of data between line ministries.	
Vietnam		Pilot country for UNSD MDG project	See above	

Annex 2: DSDs that have been developed or are in the process of being developed[77]

[To be revised / expanded as additional information is obtained]

DSD Title	Maintenance agency	Nature of DSD	Domain specific / multi-domain	Current status	Data coverage	No. of dimensions
Balance of Payments (BOP)	IMF	global	Domain specific	Operational	Balance of payments, external reserves, international investment position (IIP), co-ordinated portfolio investment survey (CPIS), co-ordinated direct investment survey (CDIS)	16
National Accounts (NA) [30 September 2013]	Eurostat	global	Domain specific	Operational		26
Foreign Direct Investment (FDI)	OECD	global	Domain specific	Operational		19
Government Finance Statistics (GFS)	IMF	global	Domain specific	Operational		13
MDGs	UNSD	global	Multi-domain	Operational		12
Debt Reporting by Developing Countries	World Bank	global	Domain specific	Operational	DSD developed jointly with COMSEC and UNCTAD. External debt and selected foreign assets from creditor, debtor and market sources and institutions	17
R&D statistics	UNESCO*	global	Domain specific	Under development	Concept scheme covers government budget appropriations or outlays for research and development (GBAORD) and R&D statistics. DSD is being developed in cooperation with Eurostat and the OECD.	
Education	UNESCO*	global	Domain specific	Under development	Concept scheme created covers the whole of ISCED 2011. DSD is being developed in cooperation with Eurostat and the OECD. The maintenance agency is still to be identified.	
International merchandise trade	UNSD	global	Domain specific	Under development	Concept scheme covers the whole domain, 42 concepts in total. DSD developed in cooperation with the OECD, United Nations	26

[77] This table may be removed in subsequent versions of the current document and placed on the web where it can be updated on a more regular basis as further information on DSD development becomes available.

DSD Title	Maintenance agency	Nature of DSD	Domain specific / multi-domain	Current status	Data coverage	No. of dimensions
					and Eurostat. The maintenance agency is still to be identified. Expected to be finalised in 2015.	
ECOFIN	IMF	shared	Multi-domain	Operational	Used for SDDS Plus	6
STS (Short-term Statistics)	Eurostat	shared	Multi-domain	Operational		8
Short-term Economic Indicators	OECD	shared	Multi-domain	Operational	Comprises two DSDs for: short-term indicators (prices, real indicators, etc); infra-annual labour indicators	
UN CountryData	UNSD	local	Multi-domain	Operational	The CountryData DSD is based on the MDG DSD, uses the same dimensional structure (dimensions/attributes). Some codelists have been extended to support non-MDG development indicators in the project.	9
Fisheries statistics	Eurostat	local	Multi-domain	Operational	Includes catch, landings and aquaculture statistics. DSDs are available on the Euro SDMX Registry.	
Culture statistics	UNESCO	local	Multi-domain	Under development		
ICT	UNESCO	local	Domain specific	Under development		
Communications	UNESCO	local	Domain specific	Under development		
National Statistics Data Page for SDDS Plus	IMF	shared	Multi-domain			
Labour force statistics	ILO	local	Multi-domain	Operational	Comprises two separate DSDs	
Labour force survey	Eurostat	shared	Multi-domain			8
Labour cost index	Eurostat	shared	Domain specific			7
Consumer price index	Eurostat	shared	Domain specific			6
Fisheries catch statistics	FAO					
Health statistics	WHO					
Waste statistics						
Air transport statistics						
Pesticide statistics						
Job vacancy statistics						

Source: SDMX website, international agency websites. * Maintenance agency still to be identified.

Annex 3: EURO-SDMX Metadata Structure Concepts
(Release 3, March 2009)

In addition to the concept name, for each concept Release 3 documentation also includes: concept codes, more detailed description, representation, and ESS guidelines.

1. Contact	1.1 Contact organisation
	1.2 Contact organisation unit
	1.3 Contact name
	1.4 Contact person function
	1.5 Contact mail address
	1.6 Contact email address
	1.7 Contact phone number
	1.8 Contact fax number
2. Metadata update	2.1 Metadata last certified
	2.2 Metadata last posted
	2.3Metadata last update
3. Statistical presentation	3.1 Data description
	3.2 Classification system
	3.3 Sector coverage
	3.4 Statistical concepts and definitions
	3.5 Statistical unit
	3.6 Statistical population
	3.7 Reference area
	3.8 Time coverage
	3.9 Base period
4. Unit of measure	
5. Reference period	
6. Institutional mandate	6.1 Legal acts and other agreements
	6.2 Data sharing
7. Confidentiality	7.1 Confidentiality - policy
	7.2 Confidentiality - treatment
8. Release policy	8.1 Release calendar
	8.2 Release calendar access
	8.3 User access
9. Frequency of dissemination	
10. Dissemination format	10.1 News release
	10.2 Publications
	10.3 On-line database
	10.4 Micro-data access
	10.5 Other
11. Accessibility of documentation	11.1 Documentation on methodology
	11.2 Quality documentation
12. Quality management	12.1 Quality assurance
	12.2 Quality assessment
13. Relevance	13.1 User needs
	13.2 User satisfaction
	13.3 Completeness
14. Accuracy and reliability	14.1 Overall accuracy
	14.2 Sampling error
	14.3 Non-sampling error
15. Timeliness and punctuality	15.1 Timeliness
	15.2 Punctuality
16. Comparability	16.1 Comparability - geographical

	16.2 Comparability - over time
17. Coherence	17.1 Coherence – cross domain
	17.2 Coherence - internal
18. Cost and burden	
19. Data revision	19.1 Data revision - policy
	19.2 Data revision - practice
20. Statistical processing	20.1 Source data
	20.2 Frequency of data collection
	20.3 Data collection
	20.4 Data validation
	20.5 Data compilation
	20.6 Adjustment
21. Comment	

Source: Eurostat – refer http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/metadata/metadata_structure

Annex 4: IMF Data Quality Assessment Framework Metadata Items

H 0.1 National descriptor	
H 0.2 Metadata update date	
H 0.3 SM Update date	
H 0.4 Certificate date	
H 0.7 Data category notes	
0.1 Legal and institutional environment	0.1.1 Responsibility for collecting, processing and disseminating statistics
	0.1.2 Data sharing and coordination among data producing agencies
	0.1.3 Confidentiality of individual reporters' data
	0.1.4 Ensuring statistical reporting
0.2 Resources	0.2.1 Staff, facilities, computing resources, and financing
	0.2.2 Ensuring efficient use of resources
0.3 Relevance	0.3.1 Monitoring user requirements
0.4 Quality management	0.4.1 Quality policy
	0.4.2 Quality monitoring
	0.4.3 Quality planning
1.1 Professionalism	1.1.1 Impartiality of statistics
	1.1.2 Selection of sources, methodology, and modes of dissemination
	1.1.3 Commenting on erroneous interpretation and misuse of statistics
1.2 Transparency	1.2.1 Disclosure of terms and conditions for statistical collection, processing and dissemination
	1.2.2 Internal governmental access to statistics prior to release
	1.2.3 Attribution of statistical products
	1.2.4 Advance notice of major changes in methodology, source data and statistical techniques
1.3 Ethical standards	1.3.1 Guidelines for staff behaviour
2.1 Concepts and definitions	
2.2 Scope	2.2.1 Scope of the data
2.3 Classification / sectorization	2.3.1 Classification / sectorization
2.4 Basis for recording	2.4.1 Valuation
	2.4.2 Recording basis
	2.4.3 Grossing / netting procedures
3.1 Source data	3.1.1 Source data collection programs
	3.1.2 Source data definitions, scope, classifications, valuation, and time of recording
	3.1.3 Source data timeliness
3.2 Assessment of data source	3.2.1 Data source assessment
3.3 Statistical techniques	3.3.1 Source data statistical techniques
	3.3.2 Other statistical procedures
3.4 Data validation	3.4.1 Validation of intermediate results
	3.4.2 Assessment of intermediate data
	3.4.3 Assessment of discrepancies and other problems in statistical outputs
3.5 Revision studies	3.5.1 Revision studies and analyses
4.1 Periodicity and timeliness	4.1.1 Periodicity
	4.1.2 Timeliness
4.2 Consistency	4.2.1 Internal consistency
	4.2.2 Temporal consistency
	4.2.3 Intersectoral and cross-domain consistency
4.3 Revision	4.3.1 Revision schedule

	4.3.2 Identification of preliminary and / or revised data
	4.3.3 Dissemination of revision studies
5.1 Data accessibility	5.1.1 Statistical presentation
	5.1.2 Dissemination media and format
	5.1.3 Advance release calendar
	5.1.4 Simultaneous release
	5.1.5 Dissemination on request
5.2 Metadata accessibility	5.2.1 Dissemination of documentation on concepts, scope, classifications, basis of recording, data sources and statistical techniques
	5.2.2 Disseminated levels of detail
5.3 Assistance to users	5.3.1 Dissemination of information on contact points
	5.3.2 Availability of documents and services catalogues

Source: SDMX website (2009)

Annex 5: Creation of the Single Integrated Metadata Structure from the ESMS and ESQRS

[To be inserted from Eurostat website]