The 2018 SDMX Experts Workshop

OECD Premises, Boulogne-Billancourt, France

1 SUMMARY

The 9th SDMX Expert Workshop was held at the OECD in Boulogne-Billancourt, France, from 22 to 25 October 2018. The event gathered over sixty representatives from international organisations, including all SDMX Sponsor agencies, national institutions from four continents and the private sector. Experts were invited to discuss the evolution of the SDMX standard in order to respond to issues in existing data exchange and sharing scenarios, and to discuss strategic issues related to SDMX implementation and capacity building including a "training the trainers" session.

The workshop had a forward-looking agenda, with each session structured around presentations and extended, interactive discussion from the floor, including:

- extending the applicability of SDMX to use cases not envisaged in the original design, such as census, administrative data and microdata in general;
- handling more varied data types such as geospatial data objects (e.g. points, paths, shapes);
- covering larger business workflows, providing process control, and extended content validation (through VTL);
- increasing the suitability of SDMX for datawarehouseing and modelling needs beyond exchange scenarios;
- responding to productivity and maintainability gaps

Progress was also made in the creation of a common SDMX capacity building strategy by identifying a number of skills and user profiles, which will be used to better address training needs. The discussion included the experiences from providers and receivers of training activities. The existing training materials have been linked to the identified capabilities and tools covered in the presentations. The file is now available on sdmx.org.

The discussions resulted in follow-up actions for the SDMX statistical and technical working groups, primarily to begin the process of defining the scope and work programme for "SDMX 3.0", the next version of the standard, which will be a key strategical driver for the future SDMX roadmap.

2 SESSION MINUTES AND ACTIONS

2.1 USING SDMX FOR MULTI-DIMENSIONAL MODELLING

Chair: Gyorgy Gyomai (OECD)

2.1.1 Summary

This session tackled some of the recurrent issues in converting time-series or tabulated statistical content into multi-dimensional data models. An abstraction of three concept groupings (object-property-unit of measure) was proposed as a starting point for multidimensional data modelling. The presentation comprised of showing the utility of this extra abstraction layer, describing the boundary issues around these concept groups/superconcepts, and then discussing at some length "the curse of the indicator" (i.e. handling/refining the main composite, non-atomic concept of simple data models) and the modelling units of measure.

Besides taking an in-depth look at cubes design, the session also exposed some modelling and design issues that appear in the wider context of a data-warehouse, which hosts data for a multitude of statistical domains. This primarily focused on concept schemes organisation, and optimal levels of separation between statistical domains.

In the resulting discussion, participants made the following points:

- Modelling decisions should also be related to exchange and implementation contexts such as whether the model is intended for reporting or dissemination, micro, or macrodata;
- The optimal concept scheme organization may be different for referential metadata (more likely to be centralized/cascading) and data (domain based/overlapping)

2.1.2 Actions

- Include aspects of microdata modelling and usage contexts in the next version of the SDMX modelling guidelines;
- Verify that the forthcoming inclusion of multiple measures in SDMX 3.0 helps address the modelling issues outlined in this session; and while they make local/tabular models easier to represent, they do not become an obstacle for harmonization and connectivity on a larger scale.

2.2 USE OF SDMX FOR MICRODATA AND/OR IN SUPPORT OF SURVEYS AND REGISTER-BASED DATA

Chair: Alessandro Bonara (ECB)

2.2.1 Summary

This session gave an overview of the benefits and challenges of using the SDMX information model for modelling microdata, surveys and register-based data and contrasted it with another key standard in this domain (DDI). Bryan Fitzpatrick from Rapanea Consulting Limited provided an overview of the capability of DDI as a tool for microdata and compared it with SDMX. He emphasized that SDMX has no mechanism with which one can link multiple tables into a complex structure (hence it may not be as performant within the production chain), nonetheless, he argued that SDMX can be used for dissemination of micro data in 'simpler', single table structures. It was also debated that in order to evolve SDMX for the use of

microdata one should look at how DDI-structures can translate into SDMX at the frontier where microdata is ready for dissemination or aggregation.

In his contribution titled "Why use SDMX for microdata?" Edgardo Greising (ILO) presented the experience in dealing with micro datasets (like household surveys or censuses). He explained that in the context of microdata, the differentiation between dimension, attribute and especially measure is of a greater importance, and the multimeasure representation has a stronger appeal than in macrodata contexts. Given that VTL already supports tabular/multimeasure objects, the communication surface between SDMX and VTL can be made wider, improving versatility. Ultimately, the challenges of putting large volumes of microdata in certain SDMX formats ("verbosity" of XML for microdata) have to be addressed.

The last presentation of this session came from Meike Scherer (Deutsche Bundesbank), presenting an SDMX based 'House of Microdata' as an organization-wide central micro-database. She explained how the Bundesbank uses SDMX for standardization and harmonization in the House of Microdata, and showed how they achieve data integration.

Discussions concluded with a recommendation to work further on the SDMX standard for exchange to improve its fitness for granular data and metadata. A lesser explored set of SDMX objects, i.e. uncoded dimensions/attributes will have to be assessed in the microdata context as one of the key requirements in this context is to maintain identifiers for members of a volatile statistical populations, where coded lists are not practical.

2.2.2 Actions
- Incorporate the points on microdata validation, multiple measures, and large data volumes in the analysis on SDMX 3.0;

2.3 HOW CAN WE BETTER USE THE EXTENDED SDMX INFORMATION MODEL? E.G. STRUCTURESETS, CONSTRAINTS, ANNOTATIONS
Chair: David Barraclough (OECD)

2.3.1 Summary
This session proposed changes to the SDMX information model around mapping, processing, presenting, and validating data. It provided an opportunity for participants to debate the proposals, propose further use cases, and discuss tool and system implementation pathways. The session featured presentations on:

- improvements to the StructureSet artefact for mapping;
- multiple measures in SDMX to allow a more efficient representation of registers and microdata;
- allowing multiple flags on an observation;
- using annotations in a standard way for presentation metadata and processing;
- improving the use of SDMX constraints for validation and managing structural metadata.

The StructureSets presentation explained the problems to create mappings that involve complex relationships. A new artefact is proposed called a KeyValueMap (accepting partial keys as well) which will make creating SDMX mappings more user-friendly and compact.

The Multiple Measures presentation described the problems when trying to model microdata using SDMX, especially when there are measures of different types; the mono-measure model in SDMX is rather non-intuitive for the end user, and hence the proposal to allow the use of SDMX for a more intuitive modelling and processing of microdata and register-based data. The SDMX TWG propose to allow several measures of different types, and attach attributes to the different measures. It was noted however that the introduction of multiple measures increases the ways to model data (especially for macrodata), therefore, in order to avoid confusion, tight guidelines and good mapping tools are needed to connect multi and mono-measure representations.

The Annotations presentation outlined the opportunities to use SDMX to drive statistical processing and data presentation; for example: on which axis to present a concept in a chart, how to display a data flow, does the data or artefact have reference metadata associated with it. Annotations have already been used for such cases but their use is not standardized, therefore shared tools cannot be easily developed.

The multiple flags presentation described the current problem in SDMX where a data point cannot have multiple values for the same attribute, e.g. OBS_STATUS cannot be marked as Estimated and Break at the same time. The SDMX TWG propose a change to the data message schema that enables multiple values for the same attribute; the values are separated by a space. Comments on the proposal were that this will only work for coded attributes, not for free-text; and that new annotations may be a more elegant/efficient mechanism to represent multiple attributes.

The constraints presentation focused on two changes to make constraining and subsetting datasets more efficient and user-friendly. The first change is to allow wildcards in expressions, which will shorten the time it takes to create the constraint list, and make it more manageable. The second change is to make each of the nine constraint types a separate maintainable artefact; this will make constraints much easier to query and implement compared to the current situation where the whole constraint superset has to be queried and parsed. It was also raised that currently constraints could only be applied to dimensions and not attributes, which limited their usability. It was proposed to extend the applicability of constraints to attributes in addition to dimensions, which was met with agreement. The next steps will be to consider including these changes to constraints in the scope of SDMX 3.0 and to elaborate the proposal.

2.3.2 Actions
- Improve StructureSets in SDMX 3.0;
- Include multiple measures representation in SDMX 3.0 and provide modelling guidelines;
- Enumerate a standard vocabulary for SDMX annotations;
- The SDMX TWG should reflect on the multiple flags use case and solution in the analysis of SDMX 3.0;
- Improvements to SDMX constraints will be included in the analysis on SDMX 3.0.

2.4 HOW TO REPRESENT TEMPORAL CHANGES OF COMPILATION METHODOLOGY, METADATA OR OBSERVATIONS IN AN SDMX MODELLING AND EXCHANGE CONTEXT

Chair: Heinrich Ehrmann (BIS)

2.4.1 Summary

After the introduction of the business case, the presentation laid out the main modelling choices offered by SDMX to handle changes in reporting methodology over time, change in reporting units, time-dependent validity of codes and relationships or revisions to reported observations. These options were complemented by a presentation of the ECB's approach of maintaining "vintages" of observations/time-series, and structural artefacts, and how this concept, albeit outside the realm of the SDMX standard, can be used to soften the impact of methodological changes on the management of statistical data.

The key options outlined were a) keep all data and changes under the original identifier (series-code), possibly by using the pre-break concept and complementing the data with reference metadata, or b) create a new identifier (series-code) under which the data, following the new methodology, will be reported and stored. (This can be achieved by pre-empting such changes in a foreseen dimension and corresponding code-values, or else by a "*differentiator dimension*" (not linked to a statistical concept) that is purely used to create parallel series for the different methodologies. The choice between these options typically depends on whether back-casting data for the new methodology is an option or not, if reference metadata can be provided, and if a distinction can be made between reporting and dissemination DSDs. Discussion points included:

- Should we enhance the API and standard to support querying for specific vintages? With regards to frequent changes of versions of artefacts, it was suggested to modify SDMX' method of referencing artefacts by way of introducing "semantic versioning", an approach that would allow minor changes to code-lists without updating affected references;
- A suggestion to introduce proper time dependency for codes in code-lists and code-hierarchies to obsolete the creation of new code-list versions for most practical cases.

2.4.2 Actions

- The SDMX-SWG will complete a new guideline on these questions and publish it for review;
- Consideration of the discussion points in the scope of SDMX 3.0, such as expressing lifetimes for codes.

2.5 HOW TO MAKE PROGRESS WITH SDMX REFERENTIAL METADATA EXCHANGE

Chair: Andreas Hake (IMF)

2.5.1 Summary

The session aimed at stimulating the discussion on ways to kick-start the SDMX-based referential metadata exchanges, which significantly lag behind the progress made with data exchanges. Eurostat and Metadata Technology shared their experience with referential metadata in the SDMX format and their specific use cases. David Barraclough (chair of the SDMX Statistical Working Group) presented the work on a content-oriented guideline for standardizing reference metadata reporting in SDMX.

The group acknowledged the issues presented and other experts shared similar experiences. Challenges include the lack of widely agreed Metadata Structure Definitions (MSDs) as an equivalent to the "Data Structure Definitions for global use" that helped broaden adoption of SDMX for automated data exchanges, and limited MSD implementation experience outside Europe. Suggestions offered to use the same artefact type (DSDs) for data and referential metadata were discussed. While such an approach would offer some immediate solutions in the area of available tools, it would present new challenges for

the maintenance of existing metadata flows and code lists, particularly, no easy solution was identified that would allow linking data and referential metadata with this approach. It was therefore concluded, that such an approach is not within the reach of the current version of SDMX, and its pros and cons should be evaluated in further depth before considering it for version 3.0.

2.5.2 Actions
- The SDMX working groups to complete the work on the guideline on standardising reference metadata reporting in SDMX;
- Analyse how to improve the technical standard to ease the implementation of referential metadata exchange.

2.6 HOW TO IMPLEMENT VTL IN SDMX WORKFLOWS
Chair: Vincenzo DelVecchio (Banca d'Italia)

2.6.1 Summary
The objective of VTL is to provide a standard language for defining validation and transformation rules that would make the "Transformations and Expressions" package of SDMX 2.1 operational. In July 2018, VTL 2.0 was published, which contains many improvements compared to previous versions of the language. With the work on VTL 2.0 completed, the time has now come to implement it and use it. The session discussed some of the current challenges (e.g. the need to define appropriate SDMX formats for storing and exchanging the rules and linking them to data objects) and outlined avenues for future work.

The first presentation, held by Maurizio Capaccioli (Eurostat), described the main characteristics of the VTL with particular attention to the new features introduced by the version 2.0, recently published.

In the second presentation, given jointly by Stefano Pambianco (European Central Bank), Chris Nelson (Metadata Technology) and Stratos Nikoloutsos (Agilis), the ongoing work for implementing the VTL in SDMX and the implementation plan were described. In order to make VTL operational, specific SDMX messages need to be defined for exchanging VTL rules and appropriate methods have to be specified both for referencing SDMX artefacts and for manipulating them according to their data types.

The following presentation, "The Banks' Integrated Reporting Dictionary (BIRD) and the VTL experience" held by Silvia Giacinti (European Central Bank), focused on the current use of the VTL for the banking industry. In fact, the BIRD initiative, promoted by the European Central Bank (ECB), is already using the VTL for specifying the transformation rules that the European banks need to perform in order to produce the reports required by a number of European authorities, starting from the input data extracted from their operational systems. This initiative involves the European commercial banks and the software vendors that operate on this market.

The fourth presentation, "VTL sandbox" given by Maurizio Capaccioli (Eurostat), showed a VTL software tool developed by Eurostat. The VTL Sandbox is composed of a parser, which checks the correctness of the VTL statements and converts them into an executable sequence of elementary operations, and an interpreter, which translates these elementary operations into SQL and executes them on a relational database. The VTL Sandbox supports VTL 2.0 and is freely available on GitHub.

In the following discussions, it was clarified that the VTL is being implemented for SDMX 2.1 and that, according to the current plans, the implementation will be available by the end of June 2019. At a later

stage, when SDMX will evolve towards new and more advanced versions, the SDMX implementation of VTL will evolve accordingly. This two steps approach was recommended to make the VTL available even before the release of future SDMX versions.

Many participants underlined the importance of SDMX and VTL dealing with microdata (more than registers and surveys). In particular, it was noted that some big European microdata[1] collections deal with multiple measures through SDMX attributes, whereas VTL supports multi-measures natively. Naturally, the ongoing SDMX implementation of VTL, based on VTL 2.0, will inherit the SDMX 2.1 limitations relevant to microdata and will not exploit directly some existing VTL features.

Many institutions or initiatives are going to implement systems based on VTL. Examples of both open source and commercial tools/prototypes were shown. It was clarified nonetheless that SDMX tools for VTL can only truly take off ground when the specifications relevant to the SDMX implementation will be available, that is, starting from the second half of next year.

From early implementers experience and from the perspective of the VTL, unifying the representation of data and metadata in the future SDMX versions appears desirable, and would achieve a significant simplification of the standard.

2.6.2 Actions
- Complete the implementation of VTL in SDMX;
- Consider evolving SDMX toward a better support of the microdata and a unification of data and metadata structuring artefacts.

2.7 HOW TO REPRESENT GEOSPATIAL DATA IN SDMX
Chair: Juan Munoz Lopez (INEGI, TWG)

2.7.1 Summary
The geo-referencing presentation showed different ways to represent geographical and geodesic information, for different uses. The attendance agreed that there is a need to include this kind of information in the standard. The presenter made an invitation to provide opinions during January of 2019 about the use cases to consider in a first stage and to express their intention to participate in the task force of the TWG, which will be working on the technical specifications for geo-referencing in SDMX.

As an example that reinforces the need to represent geographical information in SDMX, Eurostat presented the geographical grid for the integration of the Census information from the European countries.

2.7.2 Actions
- A task force of the SDMX-TWG will be working on a proposal for the technical specifications to represent geospatial data in SDMX.

---

[1] For example, The European Group Register (from the Eurostat/European Statistical System), the Register of Institutions and Affiliates Database, the Security Holding Statistics database and the AnaCredit system (from the ECB/European System of Central Banks). The latter is should collect about 100 billion records per year.

2.8 How SDMX should coexist and link to other standards such as XBRL, DDI, JSON-stat

Chair: Alessandro Bonara (ECB)

2.8.1 Summary

The first presentation compared the use of JSON-stat and SDMX-json and how they co-exist today with other standards. JSON-stat is a format for data dissemination, which has existed for several years and has a large implementation base in international organisations and statistics Offices. The presentation underlined that a "sea" of formats exists that could be categorized into exchange, download or web API's. JSON-stat and SDMX-json both belong to the latter category.

The support/encapsulation of JSON-stat in SDMX-json could be related the SDMX Roadmap 2020 item "Making data usage easier", as JSON-stat has a broad set of users and compatible tools who are not currently within reach of SDMX-json. JSON-stat is semantically simple where SDMX-json is semantically rich and contains more information. Both standards have their own advantages but could benefit from being more compatible. The SDMX Secretariat will consider if this could be part of the work programme.

In the second presentation, SDMX was compared to the RDF standard. Many NSIs use RDF for metadata management, and data and metadata could be integrated with statistical and geographical information. RDF could also be used as a link between SDMX and other standards like DDI. Some SDMX-based RDF vocabularies already exist on www.w3.org including the Data Cube model and DCAT (Data Catalogue Vocabulary).

To integrate SDMX and RDF, a number of standard content-oriented resources could be considered such as code lists, cross-domain concepts and SDMX glossary, and a list of statistical subject-matter domains. These resources could be published as RDF and would improve their formal consistency and accessibility, and allow them to be linked to external resources.

Not so easy, but still valuable, would be to translate the Metadata model from the SDMX Information model to a RDF version. This could be useful for NSIs that use RDF as Metadata management systems. RDF offers interesting opportunities for SDMX for publishing and linking metadata.

The final presentation touched upon the interoperability between standards, which is part of the SDMX 2020 Roadmap. Ole Sorensen (ECB) presented the work that has been done between ECB and Banco de España on mapping the XBRL data to SDMX. Many organisations collect and disseminate data in either of these formats and could benefit from having a mapping between SDMX and XBRL(DPM). Pilot mappings were successful, although mechanical transformations did not lead to optimal data models, and revealed that translations, between standards may be model dependent and may need to be fine-tuned, rather than mechanical.

2.8.2 Actions
- Finalise the mapping between SDMX and XBRL information models and submit for review;
- As part of the SDMX 3.0 work, consider a guideline/protocol/process on how users could create new user-defined formats for SDMX.

2.9 Enabling the extension and composition of SDMX codelists

Chair: Abdulla Gozalov (UNSD)

2.9.1 Summary

A standard way to reuse and extend code lists without having to recreate them has been identified as the highest priority feature needed in SDMX. As the implementation of SDMX gains coverage in terms of both geography, number of installations and subject matter, it has become clear that participants in data exchanges need to be able to introduce extensions to global DSDs. The ability to inherit from and extend global code lists would enable the counterparties to quickly and easily customize global artefacts to implement data exchange and dissemination for local or regional use while maintaining consistency and compatibility. This session identified and reviewed benefits, challenges and pitfalls in connection with the extension of code lists, and helped develop its implementation in the upcoming version of the standard.

The session had four presentations from UNSD, OECD, and ILO (two presentations). It was shown that the lack of code list extensibility in the current version of the standard has significant adverse implications and makes both the standard and global artefacts less usable.

Use cases showed that extensibility would address, specifically: extending the global reference area code list with subnational area codes; extending the code list for observation status for a specific subject-matter domain; and adding tax code lists, which are reporter-specific and change frequently.

ILO provided further use cases, namely extending code lists to support classification variants, such as the various versions of ISIC, or the different reference area encoding systems including M49 and ISO; using a concept with several concatenated code lists (composite dimension), where currently several breakdowns sharing a code list must duplicate that code list; and breakdowns that are currently undefined but known to be populated in the future. Addressing these issues in the current version of the standard either requires several DSDs, or leads to a major structural updates.

The ILO representative then presented possible solutions identified by the TWG: adding codes dynamically in a DSD; enabling a code list to reference and extend another code list; enabling a concept to use a Hierarchical Code List; enable a concept to reference multiple code lists; modifying the Structure Maps mechanism to enable a composition of multiple code lists; and using VTL to implement the composition of a new code list. After an extensive analysis, TWG identified Inheritance (referencing and extending a parent code list) as the most appropriate option. If any duplicates occur, codes down the hierarchy override identical codes from any parent code list(s). This proposal will be submitted to the SDMX sponsors for approval. Extension at the level of dataflows and provision agreements is under review.

In a discussion that followed, the participants generally approved of the selected approach. A request was raised, however, to make it possible to not just add but also subtract or hide codes from parent code lists, which may be necessary in some contexts.

2.9.2 Actions
- SDMX TWG to continue working on including code list extensions in the SDMX technical standard as part of SDMX 3.0;
- SDMX SWG to consider impact of code list extensions on the modelling guidelines, and other guidelines;
- Both working groups to consider possibility of removing/hiding codes when extending a code list.

2.10 EVOLVING THE SDMX WEBSERVICE

Chair: Siddesh Kaushik (World Bank)

2.10.1 Summary

This session had three presentations two on the SDMX Web Service and one on the Technology Compatibility Kit (TCK)[2]. The TCK is a tool that certifies the SDMX compliance and coverage of RESTful web services. The TCK is designed to work as a black box test tool which queries for all the specified resources and features under the SDMX Rest API. A test report summarises the results from the testing, indicating the level of compliance, coverage and supported features.

The presentation was followed by a panel discussion of experts. Overall, the response to the changes in the Web Service and TCK was well received. Some of the suggestions that evolved from the session were,

Requests for Web Service features include:

- Control of the response format via a query string;
- Ability to obtain number of records to be returned for a web service query;
- Need for a query builder to help end users create a query;
- While incorporating VTL in Web Service one has to be mindful of special characters that are currently not allowed in the standard.
- Allow querying attributes in addition to dimensions

2.10.2 Actions
- SDMX TWG to consider including the web service features requested above in the standard;
- SDMX TWG to continue developing certain features of the TCK and deliver production version.

2.11 PULL MODE DATA EXCHANGE BASED ON SUBSCRIPTIONS AND NOTIFICATIONS

Chair: Álvaro Diez Soto (Eurostat)

2.11.1 Summary

The aim of this session was to present the use case of data sharing among international organisations in the context of macro-economic statistics.

The session included four presentations where presenters:

- introduced the use case;
- identified the business requirements;
- explained the technical architecture based on SDMX for a pull mode exchange;
- presented the remaining issues for having the use case fully implemented.

The presentations were well received by the experts and the main conclusions after the lively discussion were:

---

[2] GitHub repository for TCK: https://github.com/sdmx-twg/sdmx-tck

- Building a distributed data cube with multiple data providers and multiple consumers is one of the most complex scenarios in data exchange. If the implementation of this complex scenario is successful, most other scenarios will trivially follow as special cases;
- When pulling data, it is highly recommended that the organisation exposing the data implements the updateAfter parameter in the SDMX webservice, so that it outputs only the affected data since a given time. This parameter makes it easier for consumers to avoid undesirable timeouts, and optimizes the amount of data retrieved;
- The SDMX global registry should be updated to the latest version of fusion registry as some of the functionalities needed for the pull exchange are not included in the current version;
- The RSS feed for registry notifications should be a part of the next version of SDMX.

2.11.2 Actions
- Request that the SDMX global registry is updated to the latest fusion registry version;
- SDMX TWG to consider improvements to registry notifications as part of SDMX 3.0.

2.12 STRATEGIES TO CREATE GENERAL-USE SDMX PLATFORMS FOR THE STATISTICAL WORKFLOW

Chair: Abdulla Gozalov (UNSD), Rajiv Rajan (Paris 21)

2.12.1 Summary

The growing consensus among SDMX practitioners is that, in order to make SDMX more appealing to extended user base, on one hand, there is a need to hide technical complexities from its end users and on the other, extend the processes coverage through the SDMX tools.

This session looked for experiences and new ideas in: 1. specificities and functionalities, and 2. abstraction and coverage of SDMX platforms, through various examples. The objective was to deliberate and synthesise a comprehensive framework that can guide the strategic development and implementation of general-use SDMX platforms.

The panel members provided their inputs and the strategies took shape through discussion with participants. The aspects mentioned were:

1. Understanding the scope of the general-use SDMX platforms: How much of the statistical business processes the platforms ought to cover? Does it have any dependence on the use cases?
2. Mapping the ecosystem: Positioning of the SDMX platform in the midst of the known and potential actors. What new standards, tools and technologies out there could have an impact of the envisaged platforms;
3. Envisioning the use cases: Beyond international organisations and central banks, national statistical offices will continue to play a crucial role
4. Making it happen: What would it take to create general use SDMX platforms - technology, collaboration, and resources?

Rajiv Ranjan (PARIS21) introduced the topic and introduced the panel members. Touching upon the first aspect mentioned above, he provided a brief overview of the Data Flow Assessment Framework that PARIS21 is developing to support National Statistics Offices (NSOs) in conducting data flow assessments with the aim to implement data dissemination solutions.

Abdulla Gozalov (UNSD) provided the historical perspective of SDMX, especially its evolution from just being a data sharing mechanism to covering the entire data value chain. He also highlighted the lack of technology stack for various functions in the value chain. His presentation was titled, "The Expanding Scope of SDMX: Benefits and Challenges". He ended his presentation with quick questions: With the benefits too attractive, can the challenges be overcome?

Kamel Abdellaoui (National Institute of Statistics-Tunisia) presented the characteristics of the tools used in NIS and the metadata driven approach and the integration with the SDMX-RI (Eurostat). He also talked about the success story on various tools implementations and implementation challenges in Tunisia. He presented "SDMX - INS experience using Istat toolkit".

Finally Jens Dosse (OECD) using his presentation titled ".Stat suite for sdmx-based data lifecycle management" presented the modular solution of .Stat Suite and provided an example of how it is being piloted in Cambodia to encompass the data management aspects leading up to dissemination.

2.12.2 Actions
- Formulate the open questions into strategies that could be used as a handbook and/or training material in order to develop and evolve SDMX platforms;

- Ensure that issues discovered during the implementation of SDMX platforms, are addressed in further development of the standard.

2.13 WHICH SKILLS AND ROLES ARE NEEDED WHEN IMPLEMENTING SDMX?
Chair: Rafael Schmidt (BIS)

2.13.1 Summary
From a technical perspective, the SDMX standard is broad, complex, and has a considerable learning curve. In addition, many organizations adopting the SDMX standard are operating with limited resources. This session elaborated on the necessary skills and roles to achieve success while implementing the standard, given the aforementioned aspects. A key message is: Keep it as simple as possible!

The participants agreed that the SDMX standard is large, complex, and difficult to learn. This is not surprising as it is a metadata-driven solution to a difficult problem – how to support multiple types of statistical data through the processes required for data reporting, data collection, data dissemination and, recently, mathematical calculations. Therefore, even if the entirety of the SDMX standard is complex, the participants discussed how to keep usage of SDMX simple for the end-user.

The first presentation focused on how such simplicity could be achieved, e.g. by concentrating on the SDMX Information Model, and hiding the complexity of the technical implementation. This demands well thought out technical solutions to demanding use cases, good software architecture that leads to easy to understand training material, hiding the complexity of the standard from the day-to-day work of programmers, and the willingness of both technical people and statisticians to work together to solve the issues facing statistical organisations.

The second presentation elaborated on the foundation for a comprehensive, coordinated and efficient global training portfolio. The presentation aimed at answering the following questions: Who are the categories of users who need SDMX training? What kind of skills does each user category need to

develop? What kind of training programmes / materials are best suited to reach these users? How can responsibilities be split across the Sponsors to create modular and reusable training material?

A set of implementation actions were discussed and agreed on for further follow up with the relevant SDMX bodies.

2.13.2 Actions
- Map existing SDMX training material and initiative to the presented framework (skills/roles);
- Add the above mapping to the SDMX.org learning page;
- Conduct a gap analysis to identify new, out-of-date and improvements to training material;
- SDMX sponsor organisations to assume responsibility for training material for a certain user profile;
- Create standard training packages according to the above actions, and make them available on SDMX.org

2.14 DEVELOPING AN SDMX CAPACITY BUILDING STRATEGY
Chair: Márta Nagy-Rothengass (Eurostat)

2.14.1 Summary
This session provided an overview on what works, what is missing and what can be improved when designing and carrying out capacity building activities. The session comprised the views of SDMX trainers, trainees, as well as a former trainee who eventually became an SDMX trainer.

A list of suggestions and considerations emerged during the discussion:

- Trainee's view:

    - Understanding of national specificities in terms of tools and technologies helps acquire SDMX knowledge.

- Trainer's view:

    - Data modelling and IT tools should be taught as separate training courses;
    - Step-by-step is the best approach, including the presentation of information model and tools;
    - Training courses are most effective in the context of specific projects.

2.14.2 Actions
- Provide input to the actions related to developing training material from the previous session "Which skills and roles needed when implementing SDMX?", particularly:
- Understand national specificities
- Separate Data modelling and IT tools into different courses or subjects (except in the context of specific projects when the interaction of statisticians and IT experts is highly desirable)
- It is best to do trainings in the context of specific projects of the trainees

2.15 TRAINING FOR SDMX TRAINERS

Chair: Luca Gramaglia (Eurostat)

2.15.1 Summary

This session presented a list of existing training material available for specific skills and roles. A set of templates for SDMX training courses as well as UNSD's experience in designing and carrying out training courses were also presented.

The discussions identified some known issues as well as provided a set of recommendations, in particular:

- It was pointed out that translating SDMX training material into languages other than English would be easier if a set of standard training material existed;
- The proposed framework of skills and roles would benefit from the addition of certain roles (e.g. Manager, Data Steward, System architect);
- Several recommendations were made related to designing/delivering SDMX training courses:
  - For courses on project-specific implementations, the combined presence of a statistician and IT specialist may be useful;
  - Offering virtual machines with pre-packaged configurations of relevant SDMX software to be used during training courses would be beneficial.

2.15.2 Actions

- Provide input to the actions related to developing training material from the previous session "Which skills and roles needed when implementing SDMX?", particularly:
- Consider updating the presented training framework to include the suggested roles;
- Ensure that SDMX course trainers can gear the material to the audience and have access to the relevant skills and knowledge. For example, when presenting to non-IT trainees, avoid explanations that depend on IT knowledge; and for project-specific training, trainers should gain knowledge of the relevant project;
- Create a freely-available virtual machine with pre-packaged configurations of relevant SDMX software to be used during training courses, such as the SDMX Convertor and SDMX Reference Infrastructure.
- The 'Learning' section on sdmx.org should be revamped to match the proposed training framework;

-

3 LESSONS LEARNED

The 2018 Expert Workshop was organised according to a new format, which favours the participation of SDMX experts and the discussion of technical issues. This new approach effectively means a break with past editions and the SDMX Global Conference, which focuses on the promotion of SDMX and showcases best practices.

The members of the SDMX Secretariat have agreed that the new format has proved to serve its purpose, especially in supporting the development of the standard. Yet a series of proposals and/or lessons learned have also been identified:

- The physical meeting of the SDMX Secretariat as side meeting to global events has been demonstrated to work well and should be kept;
- A more suitable venue that sits participants comfortably, does not require swapping places among members of the same delegation, and ensures that screens are visible should be booked in next editions;
- In order to facilitate the discussion and the development of tangible deliverables, some expected outcomes should be defined in advanced and added to the session's abstract;
- The establishment of an advisory group and the development of the term of reference may be beneficial in turning the Expert Workshops into a more result-oriented event.