

Summary report

# 12<sup>th</sup> SDMX Experts Workshop

7 – 11 October 2024



International  
Labour  
Organisation



**BIS**



**United  
Nations**



**OECD**



**WORLD BANK GROUP**

# CONTENTS

---

1	Introduction .....	2
2	Artificial Intelligence .....	3
3	Enterprise-level implementation of SDMX / implementation scenarios .....	5
4	SDMX.io .....	9
5	Semantic standardisation .....	11
6	Data Modelling and Semantic Versioning .....	13
7	New releases of the SDMX standard .....	15
8	Using VTL in statistical processes: successful implementations, issues and improvement requests .....	17
9	SDMX and interoperability of standards .....	19
10	Global data exchange initiatives .....	21
11	Business case for SDMX .....	23
12	Micro data in SDMX .....	25

# 1 INTRODUCTION

---

The SDMX Experts Workshop is a biennial event which brings together SDMX experts from national statistical offices, central banks, international agencies and the private sector. The objective of the SDMX Experts Workshop is to discuss and promote good practices, share lessons learned and advance the standard to adapt to new technologies and evolving requirements.

The 12<sup>th</sup> edition of the SDMX Experts Workshop was held in Amsterdam from 7 to 11 October 2024. The workshop was organised by a committee appointed by the SDMX Sponsor Organisations. It was hosted and financed by the European Commission (Eurostat).

The event was physically attended by 87 SDMX experts, with others following the event remotely via live stream. 37 physical participants came from international organisations, 37 came from national administrations and 13 from private sector organisations.

A total of eleven plenary sessions were held over the first four days of the workshop, whereas an advanced capacity building session was organised on the fifth day. For each of the plenary sessions, this report presents a summary of the discussions and identifies a short set of recommended follow-up actions addressed to the SDMX governance bodies and to the SDMX community at large.

## 2 ARTIFICIAL INTELLIGENCE

---

### Session description

AI can revolutionize every stage of the statistics lifecycle, from initial data collection to final publication, discovery, and analytics. During the session led by Rafael Schmidt (BIS), speakers shared updates on AI initiatives, research, and progress since the 2023 SDMX Global Conference with workshop participants providing valuable feedback on where and how AI can be most usefully employed.

### Key takeaways

The session underscored the significant benefits that AI can bring to data processing, analysis, and reporting, beyond its already recognized potential in data discovery. Specific areas of interest included AI-powered harmonization, metadata management, and data modelling. Within the broader SDMX framework, reducing barriers to entry and enhancing interoperability with other standards were also highlighted as promising AI applications, while sharing use cases, code, and engaging with AI industry experts emerged as preferred methods for collaboration.

Despite advancements in usability and content of institutional web data portals, practical natural language data discovery has remained elusive until recently. Now, the IMF, in collaboration with EPAM, a US software engineering firm, has developed Stat-GPT, a data discovery platform with a conversational natural language interface. Jeff Danforth (IMF) presented Stat-GPT's journey from prototype in June 2023 to evaluation release in October 2024. Stat-GPT further aims to address the issue of data fragmentation by providing a single access point to statistics from multiple sources, consolidating datasets from the IMF, World Bank, ECB, and other institutions. The innovative technical architecture outlined by Maksim Samusenka (EPAM) centres on a large language model (LLM) employing prompt engineering, embeddings, and multiple agents for executing SDMX data queries and similar tasks.

In another promising data discovery development, Alessandro Benedetti and colleagues at Sease, a UK-based AI and search technology company, have been working with SDMX sponsor organizations to enhance the accuracy of responses to data consumers' questions. By combining traditional Apache Solr search engine technology with LLMs, they aim to disambiguate users' natural language queries, extract relevant information, and formulate structured queries. Despite some inherent LLM limitations, this approach demonstrates a relatively simple solution for improving the accuracy of results from natural language questions on statistical data.

The session also highlighted the importance of enhancing interoperability between data standards. Tom Roberts from SmartXData, a US company exploring data as tradable commodities, presented research on aligning ISO 20022 and SDMX metadata using AI. This approach aims to facilitate seamless data sharing and enable ISO 20022 data analytics using the SDMX metamodel which is well-suited for this purpose.

Making the SDMX standard more accessible has been a longstanding goal of its sponsors. Julian Kurz, Samuel Melm, and Brian Buffett from the BIS presented an initiative that significantly advances this effort: an SDMX conversational chatbot. This LLM/RAG powered chatbot allows users to ask questions about the SDMX technical standards and statistical guidance. Plans are in place to expand its information sources with capacity-building materials, SDMX software tool manuals and other relevant content. Development and testing are ongoing, focusing on ensuring the accuracy and relevance of the chatbot's answers.

In conclusion, the session showcased the transformative potential of AI in the field of official statistics. From enhancing data discovery and interoperability to making statistical standards more accessible, AI is poised to play a crucial role in advancing the capabilities and reach of statistical organizations. The collaborative efforts and innovative solutions presented during the session highlight the community's commitment to leveraging AI for the benefit of all stakeholders involved in the statistics lifecycle.

#### Recommendations for follow-up actions

- Coordinate SDMX/AI initiatives to jointly maximise what can be achieved and avoid duplication of effort.
- Focus on applying AI to those stages of the statistics lifecycle where it can deliver the most value, specifically: discovery / dissemination, process, analysis and reporting.
- Explore how AI can be applied to harmonisation, metadata management and data modelling.
- Explore how AI can further help to reduce the SDMX barrier to entry and enable interoperability with other standards.
- Strengthen collaboration on AI initiatives by exchanging use cases and code, engaging with AI/ML experts from other institutions and sharing the funding burden through co-investment.

### 3 ENTERPRISE-LEVEL IMPLEMENTATION OF SDMX / IMPLEMENTATION SCENARIOS

---

#### Session description

This session was moderated by David Barraclough (OECD) and Allen Boddie (IMF). It described a wide range of use cases where SDMX is being applied to discuss implementation challenges and solutions. The focus was at the strategical level such as business processes and non-technical innovations for successful SDMX integration rather than specific platforms.

The session ended with a breakout exercise.

#### Key takeaways

The **OECD's presentation on Task force for Implementing SDMX at the enterprise level** explained that there is hardly any approved guidance on strategically implementing SDMX at scale, even though the knowledge and experience exists. The presentation described the knowledge gap in the content-oriented guidelines. The joint-working group task force's job is to help fill this gap by compiling an approved guideline to help implementors adopt SDMX more rapidly and at lower cost.

There was a recruitment call for additional implementation experts to the task force, including a volunteer for chair. This was highly successful; the task force doubled its participants, and we have a broader geographical and statistical representation. The next steps are for the task force to meet in November to discuss the findings from this session.

In the **ADB's presentation on Enhancing Data Management through the SDMX standard**, it was seen that SDMX is an enabler for innovation and operation efficiency. There is strong regional and global demand to implement SDMX, with challenges such as the need for upskilling and demand for open-source tools are acknowledged. To address this need, online e-learning courses are an effective upskilling strategy, and initiatives by ADB and others in the global SDMX community are closing the knowledge gap. ADB's Data Division envisions setting-up a new technical assistance project in 2025 to support selected developing member economies in Asia and the Pacific on SDMX adoption. ADB is open to collaborate with development partners and NSOs in Asia and the Pacific interested in adopting SDMX into their national statistical system.

In the **AfDB's presentation on the Use of SDMX in a multi-standards environment as a base to build transversal platforms**, it was seen that African countries wish to make their data FAIR, reduce the gap between production and dissemination, satisfy data users requests and reduce data reporting burden, and enhance tools ownership and foster skills autonomy. At sub-regional level, the needs are to Streamline Data Collection and Transmission, have better Coordination with Member Countries, and improve country data quality.

There are several challenges such as siloed projects, data fragmentation, limited expertise and the need to retain expertise, resource and IT infrastructure, coordinating across stakeholders, and data quality. Some solutions foreseen are to have a formal process to initiate request for technical assistance, start with a data analysis step to understand data inventory, a plan to sustain expertise in the country, and move to cloud-based IT solutions.

In **INEGI's presentation on the use of SDMX in a multi-standards environment as a base to build transversal platforms**, it was explained that transversal platforms are a structured set of technologies supporting the aspects defined by an architecture to satisfy all the needs related to the statistical lifecycle in an institution, and they provide an environment for the whole statistical process and enables statistical interoperability. They should be used in conjunction with a governance framework such as the Data Governance Framework for Statistical Interoperability (DAFI). The reasons to adopt them were enumerated, such as to improve IT governance, cost-cutting, accelerate development, better interoperability, strengthen cybersecurity, etc. A transversal platform in statistics should also map to the ModernStats standard (GSBPM, GAMS0) to enhance interoperability and coherence with a stats office statistical lifecycle. It was mentioned that one drawback today trying to use SDMX for these platforms is that SDMX 3.0 is not widely implemented, so management of microdata is very limited.

In **iStat's presentation on SDMX strategy in IStat: state of art and perspectives**, the EU-funded modernisation program undertaken by iStat was described. The main aims were to increase quality in production, dissemination and exchange of statistical data, according to international regulations, guidelines, standards and best practices, reduce costs in the medium term (reuse as much as possible of methodology, meta-information, software tools and libraries), and provide sustainable solutions in the long term. The enablers were to reengineer the statistical business processes and increasing the usage of metadata within the data life cycle, use European regulations and agreements with international and national organizations, and migrate from legacy dissemination systems towards plug & play IT architectures and tools. There followed a description of the iStat Toolkit system and how it meets these objectives. Some lessons learnt were to identify priorities carefully, ensure that staff have the knowledge, and have a clear implementation plan.

In **Statistics Netherlands (CBS) presentation on the Role of Data and Metadata Standards in Data Strategy**, three objectives were described: Societal challenges are central, increasing access to data, and maintaining high quality. But the current situation is a fragmented data landscape. The overall vision is to implement a 'data-centric' data management. CBS will base their data architecture with 5 steady states that map to GSBPM. They will use .Stat Suite to replace their some of their silos and Statline system as .Stat Suite supports the implementation of the steady-state paradigm. Also, SDMX provides an established information model and many open-source tools. Challenges were that this requires a cultural change in the office, there are limited data modelling skills, there multiple modernisation project vying for resources, REST API may be limited for large amounts of data, and (again) support for microdata in SDMX is not yet mature. In summary, to have FAIR data you first need metadata standards, and don't under-estimate the innovative power of international communities.

In **HMS analytical software's presentation on an integrator's perspective on the evolution of tools in the context of changing business and IT demands**, it was explained that data practitioners need tools that are effective and simple to use, automated, customisable, and well-maintained. SDMX enables this. HMS showed some automated processes such as real and near-time sync of structural metadata between systems, the use cases where they apply, and the systems involved. The conclusion was that tools, IT, and business need to be considered to achieve an SDMX implementation project's outcome.

The session ended with a **breakout exercise** where the plenary room was split into quarters. Each group was moderated by David Barraclough, Allen Boddie, Edgardo Greising, and Luca Gramaglia.

Each group was asked the same questions which they discussed for 20 minutes:

- What were the main challenges for their organisation in adopting and using SDMX across the organisation?
- What could help them overcome the challenges that they had?

The discussions were summarised and presented in plenary. The summary is available in Box 1 on the next page.

#### Recommendations for follow-up actions

- These takeaways will be discussed in the joint working group task force for Implementing SDMX at the enterprise level;
- Continue to refine SDMX capacity building, it is one of the key enablers mentioned;
- Consider if people can actually find the existing training – if it exists it must also be discoverable, communicated, and compelling;
- Implementation of SDMX 3.0 is in-demand for agencies with microdata use cases;
- SDMX should be better positioned as a strategy-oriented, business-oriented solution rather than just a technical one. There could be an “SDMX for managers/business analysts” course, and a marketing strategy could be drawn up to produce senior-management level communications.



### Box 1: Summary of breakout session on enterprise-level implementations of SDMX

In addition to general fears of changes to established processes, participants highlighted several themes around the challenges they faced and what could be done to overcome these challenges.

**A lack of buy in from senior management with no clear data strategy or enterprise level mandate to implement SDMX make it difficult to prioritize an SDMX project.** This generally contributed to a lack of understanding by statisticians that SDMX was more than a data format or IT project. It was felt that better documenting the SDMX business case as well as case studies on successful adoption throughout the data lifecycle would help senior staff see the value in implementing SDMX from an efficiency standpoint. This could then lead to more resources being put into the project. It was also felt that outreach at senior level meetings would be beneficial to the promotion of the standard. Implementing SDMX as a bottom-up project from IT, especially without executive support, should be avoided as it would be regarded as a technical project with little benefit to the business.

**A general lack of education or technical knowledge as well as learning resources around SDMX make it difficult to get staff engaged with an SDMX project.** Generally, participants felt that capacity building materials were lacking as was promotion of the standard. Expanding the SDMX material to better cover introductory information needed for people new to the standard was recommended. It was proposed that a “unified”, e.g. ISO, certification on SDMX skills would help managers recognize this skillset within their staff as well as in external consultants, and help motivate staff to upskill in SDMX.

**Choices around tooling and integration with legacy systems make it difficult to implement an SDMX project on the IT side.** Most tools still lack coverage of SDMX 3.0, whose features are often needed in microdata projects. It was suggested that a tools and best practices guide would be helpful. Any improvement in the linkage between common data science tools was also seen as an enabling factor for staff involved in SDMX projects.

**Improvements in Data management which often accompany SDMX projects lead to the perception that SDMX is complex.** Organizations often take the move to SDMX as an opportunity to improve Harmonization between datasets and to improve data quality. These activities are needed to get substantial benefit out of SDMX, but it was often seen as a challenge to get statisticians on board. It was suggested that a standardized SDMX governance framework would help organizations to plan how to best improve data management as part of an implementation project.

## 4 SDMX.IO

---

### Session description

This session provided an sdmx.io update and roadmap along with a deep dive into the 2023 IFC Survey on the adoption and use of SDMX in central banks and a presentation of selected sdmx.io ecosystem tools: FMR Workbench; pysdmx; Global Discovery Service (GDS); the Test Compatibility Kit (TcK).

### Key takeaways

**sdmx.io** – the presentation reviewed the purpose of sdmx.io and provided an update on achievements to date and the workplan. It was clear from the subsequent presentations that ecosystem and elearning efforts are responding to community needs. The roadmap emphasised AI initiatives such as an SDMX Chatbot and elearning courses to address the knowledge gap; an SDMX for Data Scientists initiative to increase awareness and use of SDMX data sources; and an expanded ecosystem which includes multiple VTL tools.

**Manage Artefacts with FMR Workbench** – the power of the (sdmxio) ecosystem model was presented along with a demo of the FMR Workbench. The NBB presented their integrated ecosystem of FMR Workbench, the Excel Matrix Generator, .Stat Suite, and XMLSpy working seamlessly together in simulated NBB metadata management activities.

**2023 IFC Survey Results** – the final results of the 2023 SDMX Tools survey of Central Banks were presented. SDMX adoption continues to expand with 79% of respondent banks using or implementing SDMX, compared with 64% in 2016. The main blocking factors identified were resources and support for SDMX in tools. Central Banks who identified SDMX as extremely or highly relevant also indicated a knowledge gap in their ability to implement SDMX. This knowledge gap is consistent across all data lifecycle processes. A wide range of open source tools are in use with most used by only 1 or 2 central banks. Central Banks reported being satisfied with the tools and with the support resources.

**The SDMX Global Discovery Service** – the purpose of the SGDS was presented, namely, as more SDMX Registries are deployed and used in production, the need for SDMX Artefacts **discovery** and **sharing** is becoming more important. The current status of the SGDS and a demo of the UI were presented along with the roadmap. The integration of the TcK as a key element for validating API endpoints was shared. Slido polls conducted for SGDS indicated that participants find it difficult to discover SDMX data/metadata sources and that they are interested in using both the UI and API features of the SGDS.

**Introduction to pysdmx** – an overview of pysdmx was presented along with key features and the roadmap. It was highlighted that version 1.0 will be release soon with the following features: SDMX 3.0 compliance, provides an information model, retrieves metadata from SDMX registries, reads SDMX data and returns a Pandas Data Frame. The usage of pysdmx

at the BIS and at Meaningful data covers process automation, data visualisation, and provision of data and metadata to other applications such as VTL.

**SDMX TcK State of Play and Roadmap** – the purpose of the TcK, a tool for measuring compliance and coverage of an SDMX RESTful endpoint against available SDMX REST API versions was reinforced. The architecture and an overview of how it works was shared. The presentation clearly communicated the coverage of the current production release of the TcK as well as the major tests covered by various test releases. Participants were split between Reference Metadata Queries and Schema Queries on the priorities for the next features to be covered.

#### Recommendations for follow-up actions

- Address SDMX adoption barriers by promoting pilot exercises or sharing successful implementation business cases. Invest in initiatives to help close the knowledge gap. To the extent possible, ensure tools are updated to support the latest versions of the standard. Consolidate tools to reduce duplication and inefficiencies.
- Support the ecosystem approach of fit-for-purpose tools which come together to solve practical use cases. Reduce feature overlap in tools. Establish interoperability good practices.
- Continue to support efforts to increase visibility, accessibility, and usability of SDMX REST API endpoints through services such as TcK and SGDS in addition to tools such as pysdmx.
- Extend the IFC SDMX Tools Survey of Central Banks to cover the full range of official statistics producers.

## 5 SEMANTIC STANDARDISATION

---

### Session description

This session had two groups of presentations:

- The first group discussed efforts in various groups to semantically standardise concepts, code lists and classifications – and offer guidelines and shared SDMX representations globally.
- The second group presented cases with the limits of standardisation attempts for code-lists and concepts and offered a first deliberation of potential additions to the standard that can introduce a deeper, more articulate semantic layer.

### Key takeaways

**SDMX Classifications:** Given that international classifications reflect harmonisation efforts in the international statistics community, they are best candidates for SDMX implemented concepts and code-lists. While some initial work has been done and global artefacts are available in the GR, further work, including some refactoring of initial, divergent, and not necessarily best practice designs is warranted.

The SDMX Classifications Task Team has been mandated by the sponsors to improve availability of standard classifications in the GR: with clear, composable, easy to maintain code-lists (with explicit mappings if resources permit between concurrent variants and consecutive versions). An ambition is to link to the classification custodian, find explicit/invested maintainers, who can ensure timely and regular updates of the classification artefacts.

**SDMX guidelines for units of measure:** The Units of measure Task Group launched by the SDMX SWG has produced the draft guidelines for Units of measure modelling (including neighbouring concepts and code-lists: e.g. unit multipliers, price base etc.).

The guidelines promote alignment with units of measure standards in sciences, the use of dimensional analysis for deriving new units of measures and provides recommendations to special cases typical in economic and social statistics (monetary value, counting various statistical populations).

The draft is submitted to the SWG, and later for public review and sponsor approval.

**The need to establish semantic relationships between concepts – uses cases from the**

**OECD data explorer:** In the migration process from a legacy data-warehouse to a new SDMX based one, OECD has embarked on a large, organization scale harmonization exercise. While the exercise was successful in reducing variation of code-lists and increased interconnectedness of datasets, it also revealed the limits of standardization.

Standardisation, when pushed to extreme, could impact readability, the ergonomics, and discoverability of the data. Mechanisms are needed to maintain semantic richness and contextual flexibility. The objects, within and beyond the standard that could achieve these were enumerated (concept roles, annotations, category-schemes or skos/xkos), but further work is needed to find the best approach, most in line with the standard.

**A Unified Approach to SDMX Interoperability:** Lack of sufficient semantic richness in the SDMX standard and other obstacles of full standardisation (legal, stickiness of partially aligned bubbles) create metadata silos. A proposal was put forward to introduce a sub-layer for SDMX, based on simple entities + relationships model (where both the entities and relationships can be identified and introduced flexibly) – e.g. in which entities can underpin various codes in related code-lists. In general, such connectivity, granular reference ability, and semantic richness (which could emerge from the open relationship types) is desirable. The dilemma is really whether this should be achieved based on available LOD standards. Additionally, maintainability and governance at such granular level was mentioned as a challenge.

#### Recommendations for follow-up actions

- The SDMX Classifications Task Team will continue its work prioritising the Reference area code-lists (clean separate code lists, with Geographical hierarchy and political hierarchy separated), work towards establishing connections with classifications experts, and extend the scope of SDMX represented classifications.
- The units of measure Task Group has submitted the guidelines to SDMX SWG for review. In response to suggestions from the audience will revisit the text of the guidelines to explicitly address the scenario of multi-measure data-models.
- Continue the research into achieving better interoperability and preserving semantic richness with the standard (recommended mechanisms within the standard, improvement of the standard – with the addition of an interoperability layer, or interfacing with existing LOD standards.)

## 6 DATA MODELLING AND SEMANTIC VERSIONING

---

### Session description

In a pivotal session chaired by Yamil Vargas (IMF) and Glenn Tice (BIS), participants explored the transformative role of data modelling within the framework of Statistical Data and Metadata Exchange (SDMX). The discussions underscored the importance of effective data modelling in enhancing the usability and accessibility of statistical data.

### Key takeaways

The session commenced with a presentation by Jeff Danforth (IMF) and Maksim Samusenka (EPAM) on the **AI Powered SDMX-Data-Modelling Tool**. They introduced an innovative approach that integrates **Generative AI (GenAI)** into the data modelling process. This co-pilot-like system aims to facilitate the conversion of legacy structures to SDMX components, thereby streamlining workflows. By utilizing GenAI for tasks such as adding new concepts and modifying existing artefacts, the presenters highlighted the potential for significantly improving productivity while ensuring that human oversight remains central to the process.

Glenn Tice (BIS) followed with insights into the **FMR Data Modeller UI**, a tool designed to enhance statistical domain modelling within the SDMX framework. He explained how this tool serves as a solution to common challenges faced by statisticians, including the need for cohesive domain models and bridging expertise gaps. The FMR Data Modeller UI provides a structured approach to modelling statistical domains, integrates seamlessly with the SDMX Matrix Generator, and FMR's main features as a comprehensive metadata repository. This allows users to create, browse, and retrieve metadata structures all within a single, user-friendly platform.

Next, Yavuz Coban and David Barraclough (OECD) presented on the **Decomposition of Indicators to Identify Statistical Concepts**. Their focus was on methodologies for breaking down complex indicators into fundamental statistical concepts. This decomposition is crucial for increasing accessibility, simplifying data queries, and enhancing comparability. The presenters underscored the risks of misinterpretation associated with non-decomposed variable concepts, emphasizing the need for clarity in statistical dissemination.

Finally, Allen Boddie (IMF) addressed the principles of **Semantic Versioning in SDMX**. He provided an update on the taskforce's efforts to revamp the guidelines for versioning SDMX artefacts, initially established in 2015. Allen detailed the categorization of various SDMX artefacts and introduced wildcarding dependencies to maintain consistency in versioning practices. He outlined the next steps for releasing the updated guidelines as **Content-Oriented Guidelines** for SDMX 3.0 and 3.1, with plans for future integration into the **SDMX Technical Specifications**.

In conclusion, this session fostered rich discussions around the future of data modelling in the context of SDMX. The collective insights from the presentations on the integration of AI, advancements in the FMR Data Modeller UI, the importance of indicator decomposition,

and the principles of semantic versioning offer a roadmap for improving the clarity, usability, and accessibility of statistical data across various domains. This collaborative effort aims to pave the way for innovative practices and tools that will benefit statisticians and data users alike.

#### Recommendations for follow-up actions

- SWG to formalise guidelines for semantic versioning of SDMX artefacts.
- Data modellers to reflect on the opportunities to improve productivity using GenAI and feedback comments to the IMF on the proposal for a co-pilot-like system to facilitate the conversion of legacy structures to SDMX components.
- Where possible, modellers should decompose complex indicators into fundamental statistical concepts to improve accessibility, simplify data queries and enhance comparability of datasets.
- Institutions with statistical domain modelling use cases to comment on the BIS proposal to converge the OECD's Matrix Generator with FMR and consider co-investment in the project.

## 7 NEW RELEASES OF THE SDMX STANDARD

---

### Session description

This session focused on the latest and upcoming developments in the SDMX standard, including:

- An overview on SDMX 3.0 implementation.
- SDMX 3.1: horizontally complex DSDs.
- SDMX 3.2: presentational metadata (and broader capabilities of the feature).
- A discussion on actionable steps or recommendations for future developments.

### Key takeaways

**SDMX 3.0 implementation** – The IFC survey on SDMX adoption revealed blocking factors to adopting (or migrating to) SDMX 3.0: lack of resources, absence of a business case, and the need to adapt tools. Most open-source tools are expected to support the major features of SDMX 3.0 by Q3 2026. Poll results indicated a preference for core libraries to support new versions within less than a year of release, and tools within 1 to 2 years. Tools should support new SDMX features as a core principle. Data Reporting is a key driver for initial SDMX adoption which is subsequently extended throughout the data lifecycle. Reporting flows, and associated tools, should keep up with new versions of the standard to avoid technical debt and outdated implementations.

**SDMX 3.1: horizontally complex DSDs** - The presentation highlighted the problem of working with highly multi-dimensional datasets, which can result in vertically or horizontally complex DSDs which are costly to update. The proposed solution addresses this complexity with a new content constraint on the dataflow level: dimension constraint. With this change, dimensions can be added to existing DSDs without disrupting existing dataflows. This facilitates the development of DSDs with pure concepts and clean code lists, while reporting DSDs are reusable for dissemination without loss of interoperability or usability.

**SDMX 3.2: Custom Structure Definitions** - CSDs were introduced as a new structural metadata artefact planned for end 2025. CSDs aim to satisfy diverse metadata requirements by defining objects, specific properties (eg pivot tables) and references to other SDMX artefacts. CSDs will be referenceable and discoverable via the REST API. Poll results indicated that annotations, vendor-specific tools, and other metadata models are currently used to address non-SDMX metadata modelling. The main use case indicated in the poll for CSDs is data dissemination, followed by internal process-related metadata.

**SDMX-REST API** - The past developments and the new features of the SDMX REST API (time travel and pagination) were presented. Participants expressed interest in possible new features that go beyond data and metadata retrieval, such as business functions for format



conversion, data validation, and data mapping. The poll showed moderate interest in exploring REST alternatives like gRPC (44%) and GraphQL (22%).

**SDMX Global Discovery Service** - The implementation and information model of the SGDS were presented. SGDS responds with three types of messages for agency, endpoint, and pubLocation, and it offers a mechanism for resolving a URN (involving parsing the URN, identifying the AgencyID, and returning the resource if response is HTTP 200). Participants showed intention to use the URN parser and high interest in the endpoint discovery feature.

#### Recommendations for follow-up actions

- Address adoption barriers for SDMX 3.0 (and for future releases) by promoting pilot exercises or sharing successful implementation business cases. Data aggregators, which are mostly IOs, to migrate outdated SDMX 2.0 data reporting implementations to a newer version of the standard.
- Encourage timely release of core libraries and tools to support new versions. Tools should support new SDMX features as a core principle.
- Extend the relevant parts of the survey to a wider audience to gather more insights.
- Consider integrating business functions into the SDMX REST API and explore REST alternatives.
- Include a review process that allows widely used CSDs to be proposed as a new SDMX artefact.

## 8 USING VTL IN STATISTICAL PROCESSES: SUCCESSFUL IMPLEMENTATIONS, ISSUES AND IMPROVEMENT REQUESTS

---

### Session description

This session provided an overview of real-world use cases for VTL, focusing on what is missing/improvable in VTL language and tools.

### Key takeaways

After some years of activities to define and consolidate the language, VTL has definitively started the process for a wide adoption by national and international organisations, since different open source engines and several tools are available to use it in production statistical processes.

The new form of documentation, which will be officially published at the end of the year with the VTL 2.1 version, will facilitate users to get familiar with the language and it will enable possible new implementers of calculation engines and supporting tools, given the improved readability and quality of the manuals and the availability of input and output data and structures for each of the examples given for every operators; the new format will allow to increase their number, adding all the test cases used by the tools' implementers.

The number of PoCs and pilot projects in which the VTL Task Force has been involved has greatly increased in the last year and many organisations are considering to introduce VTL in their collection and production processes; a useful guideline using a GSBPM-like approach has been presented to help defining all the activities needed to successfully integrate VTL in the statistical data processing of the organisations.

Some VTL tools will join the SDMX.io initiative in the next months; in particular, integration of VTL engines with FMR will enable SDMX users not only to exchange VTL scripts together with data and metadata (already possible for SDMX 3.0), but also to execute them in order to complement SDMX validation and transformation capabilities with VTL language.

A very important ESCB business initiative (Integrated REporting Framework) and the related Common Data Management IT project are going to adopt VTL for statistical collection and production processes; the National Central Banks who are in charge of developing the calculation engine in a cloud native environment are evaluating the currently available open source engines and will contribute to enforce VTL language's quality and reliability.

### Recommendations for follow-up actions

- VTL pages in the SDMX website should be refreshed and more user-friendly; VTL TF will contact SDMX Web editors for this task.

- Training sessions and webinars should be prepared and advertised in order to make VTL easier and more accessible for final users; VTL TF will try to put together all the existing training material about VTL in order to organise events and to publish self-service courses.
- Further efforts should be aimed at implementing APIs to invoke VTL services for parsing and executing VTL scripts.
- SDMX Secretariat and Sponsor organisations will consider the possibility to start the process for VTL to become ISO Standard.

## 9 SDMX AND INTEROPERABILITY OF STANDARDS

---

### Session description

SDMX's integrated approach and comprehensive guidelines make it a powerful tool for achieving interoperability in statistical data exchange. This session aimed to portray these features of the standard and show how SDMX can be integrated with other standards.

### Key takeaways

**SDMX and Interoperability of Standards (Maris Rusev – d-fine)** - Interoperability is crucial for the financial sector to become more data-driven. The relationship between SDMX (Statistical Data and Metadata Exchange) and XBRL (eXtensible Business Reporting Language)/DPM (Data Point Metadata) models was explored in a presentation based on an implementation used in Bundesbank. Emphasis on validation and transformation techniques to build strong data architectures is important. A case study on transforming DPM XBRL data points into time series for SDMX use, especially regarding European Sustainability Reporting Standards (ESRS) had been done. Interoperability enhances data sharing, fostering better decision-making and policy formation.

**DPM 2.0 Overview (Fernando Wagener – ECB)** – DPM 2.0 (also known as DPM Refit) aims to improve interoperability with other reporting standards, particularly SDMX. The presentation focused on how the DPM 2.0 model applies to non-supervisory purposes, using IreF (Integrated Reporting Framework) as an example. It also provided examples and prove on how the DPM model 2.0 integrated with SDMX in general, but mainly 3.0.

**Statistical Organizations and the Official Statistics Landscape (Olav ten Bosch - CBS)** - Statistical bodies provide a wide range of outputs (articles, graphs, data tables) that support research, policymaking, education, etc. SDMX plays a growing role but must interact with other standards to ensure a better user experience aligned with FAIR (Findable, Accessible, Interpretable, Reusable) principles. The presentation discussed the need for interoperability in official statistics and the role of open-source software used by data scientists and policymakers. The dream would be to have one R package to access all official statistics content from all data providers. The questions was given, Could Linked Data be more easily picked up with AI?

**Interoperability of .Stat Suite, FMR, and NSIWS (Pedro Corranza – OECD)** – The presentation questioned whether a single sign-on and unified user rights management can enhance the interoperability of the technical architecture for these systems. As an example, the basic authentication that is recommended in the SDMX API was mentioned as being an issue in relation to security and we should remove it from the standard. The last question to the audience was : What can we do with authentication and authorization, could SDMX adopt or recommend secure common standards in relation to authentication and do we make authorisation part of the standard?

**SDMX and Semantic Web Standards – (Luca Gramaglia - Eurostat)** - The presentation explored how higher interoperability between SDMX and semantic web standards could be mutually beneficial. Proposals for future initiatives based on previous attempts at integrating these standards are discussed. Could a new working package related to semantic web ontologies and SDMX and create real world uses cases? Linked Open Data question could maybe be related to the presentation from Olav (CBS).

#### Recommendations for follow-up actions

- Cross check the list of tools provided by Olav Ten Bosch in his presentation against the [sdmx.org](http://sdmx.org) tools page.
- Follow up whether a working package for TWG/SWG should be created investigating the interoperability between semantic web standards and SDMX and investigate the Linked Open Data, how this can be achieved.
- For the TWG group to investigate the authentication and authorization methods and how they can be integrated and recommended by the SDMX standard.

## 10 GLOBAL DATA EXCHANGE INITIATIVES

---

### Session description

The first global DSDs were introduced in 2013. The session took stock of the challenges faced and lessons learned in over one decade of managing and maintaining global DSDs. Workshop participants discussed in particular potential solutions for the four issues below:

- How global DSDs could be made available for a larger number of statistical domains.
- What modelling issues have been encountered in the creation of global DSDs, and how they could be tackled.
- How to improve the maintenance of global DSDs.
- How to remove potential obstacles to interoperability and other implementation issues when using global DSDs.

### Key takeaways

Several statistical domains were mentioned as potentially benefiting from the creation of global DSDs (e.g. population statistics, environmental statistics and climate change, tax revenue statistics). However, it was recognised that the development may be easier in domains where there is already methodological consolidation and where shared data collections across different international organisations already exist (e.g. health statistics, energy statistics and statistics on research and development). Other potential domains could be identified via a public call for interest. In order to push for the creation and implementation of such global DSDs, workshop participants highlighted the need to reach out to the responsible governance bodies in each domain, but also the need to provide additional incentives for national organisations to adopt global DSDs. One way to provide such incentives would be to create global DSDs that are also suitable for dissemination of include dissemination DSDs in global DSD packages.

Several modelling issues in the current global DSDs were also highlighted. Participants mentioned in particular that global DSDs are often complex due to the large number of dimensions, that some dimensions are used in an inconsistent manner and that the semantics associated to certain codes differ across organisations. Global DSDs also rely on domain-specific concept schemes which often duplicate cross-domain concepts already available in other concept schemes. The introduction of the new SNA 2025 could be an occasion to fix these issues in the global macro-economic statistics DSDs. This would however require the involvement of modelling experts at an early stage in the development of the new framework.

Regarding the maintenance of global DSDs, workshop participants highlighted that certain functionalities available in SDMX 3.0 (e.g. semantic versioning or code list extensions) could reduce the workload associated to the maintenance of global DSDs. Participants also

suggested that, while releases of global DSDs could keep on occurring annually, the review of the pending issues could occur on a quarterly basis, in order to more evenly spread out the work associated to the review over the course of the year. The lack of clear channels to raise and follow up on issues related to global DSDs and cross-domain concepts / code lists was also highlighted as a weakness in the current maintenance process.

Workshop participants agreed that the rationalisation of formats foreseen in SDMX 3.0 could help in avoiding interoperability issues due to different expected formats. The Technical Working Group could also promote the principle that SDMX-compliant tools to support all formats for a given SDMX version and implement this principle as part of the SDMX Test Compatibility Kit (TCK). Furthermore, the currently under-used pull mechanism could be further promoted by having better support for update notifications in available SDMX tools and by updating the subscription notifications to rely on more modern technologies (currently they are based on email).

#### Recommendations for follow-up actions

- The SDMX Sponsors should identify a shortlist of priority statistical domains for which global DSDs should be developed.
- The SDMX Macro-Economic Statistics Ownership Group should ensure that SDMX experts are associated to the development of the SNA 2025 framework, so that potential modelling issues could be identified at an early stage.
- The Statistical Working Group should define a clear channel through which issues related to cross-domain artefacts can be reported and tracked.

## 11 BUSINESS CASE FOR SDMX

---

### Session description

The session started with a presentation about the evolving applications of SDMX and the initiative to update the Business Case. Thereafter, the Experts workshop members were asked to classify / rank / prioritise different use cases for SDMX implementation and to what extent they should be emphasised in the business case for SDMX. An intensive discussion followed that provided valuable insights into the current use of SDMX, including unorthodox but promising use cases.

### Key takeaways

There is strong support for the review of the SDMX Business Case. SDMX as a facilitator of data governance emerged as a most important use case, with many reports including those where SDMX is used primarily for that purpose.

SDMX has been referred to as a “best-kept secret”, and it was strongly suggested that the initiative should be more actively promoted. Engagement with senior management of statistical offices was identified as a key aspect of the promotion, and a session on SDMX Business Case aimed at senior management was proposed for the next year’s SDMX Global Conference.

The importance of SDMX facilitating data interoperability, both within the enterprise and on a global scale, was highlighted by many participants. The implementation of metadata-driven processes was identified as a key benefit. The benefits of SDMX in moving data from production to dissemination, particularly in the context of recently introduced support for micro-data, was also emphasized.

The benefits of SDMX in the context of data discovery, quality measurement, driving innovation, and improving transparency, were all highlighted by the participants. Finally, it was suggested that the use of SDMX beyond statistics should be considered, and several examples to that effect were provided including the use of SDMX for administrative records.

The following tentative conclusions can be made from the slido poll run at the end of the session:

- The uses of SDMX for data collection, reporting, dissemination, and data/metadata management, have received a fairly even distribution, with 58-70% of the participants reporting their usage in each of those areas.
- Dissemination and data/metadata management emerged as the most important use cases in the participants’ view, followed closely by data reporting and collection.
- Data/metadata management as well as emerging use cases including the AI, were identified as areas where the use of SDMX should urgently be increased.



- A plurality (48%) of the participants reported that facilitation of data harmonization was a medium priority at their offices. 18% reported interoperability was a high priority, and 33% a low priority.
- 75% of participants reported that harmonization should receive further effort going forward.

#### Recommendations for follow-up actions

- The SDMX Business Case Review Task Force should use the takeaways from the session in its work, in particular with regard to the uncommon and unorthodox use cases.
- Further communication efforts should be directed at senior management of statistical offices, including high-profile events such as the UN Statistical Commission.
- The importance and value of data harmonization should be communicated and strongly emphasised both to the technical experts and senior management.
- It was proposed that the initiative should be renamed due to its narrow reference to Exchange (of statistical data and metadata). The proposal received very strong support. The SDMX Secretariat should discuss the issue and potentially propose a new name for the initiative while keeping the logo and acronym due to their wide recognition.

## 12 MICRO DATA IN SDMX

---

### Session description

This session explored SDMX for microdata from two key perspectives. From a business-oriented standpoint, it examined microdata modelling and its related aspects. It also addressed the implementation challenges along with alternative technical solutions.

### Key takeaways

The session commenced with a short presentation by the chair, Bilyana Bogdanova (BIS), who introduced the cross-working group (CWG) on micro data. He outlined its objective to deliver a concept-oriented guideline, its organization, structure, and achievements to date.

Olgerd Unger (ECB) introduced the G-20 Data Gap Initiative rec 14, focusing on data access and sharing. He shared findings from a comparison of various data sharing standards (SDMX, XBRL, ISO 20022, DDI), highlighting SDMX for its consistency and comparability across systems, efficient data integration, global recognition and applicability to the entire production chain. He stressed the importance of interoperability and robust governance, while also addressing challenges such as integrating standards with existing IT infrastructure and balancing implementation costs with organizational benefits.

Next, Daniel Suranyi (ECB) emphasized that with SDMX 3.0 any data can be modelled. though different terminology (macro/ micro) may be an issue. He provided a progress report on the CWG's efforts, noting that the group has collected and analysed 13 use cases from 8 institutions. Members identified several challenges, such as handling multiple measures, managing relations between concepts and Data Structure Definitions (DSDs), dealing with arrays, and code list governance. The current work agenda of the group includes modelling the use cases in SDMX 3 (native) and developing SDMX 3 workarounds. The next step is to summarise the work in the concept-oriented guideline on micro data and best practices respectively.

Additionally, the analysis of the above modelling use cases has revealed some potential implementation gaps. In an interactive presentation, Edgardo Greising (ILO) outlined these gaps, proposed potential solutions, and gathered feedback from the expert audience on their criticality. Key issues identified include the need for improved querying for concepts, linking datasets, and managing aggregation rules. The presentation also emphasized the importance of integrating with other standards and efficiently handling large volume microdata set queries. Proposals were made to enhance the SDMX API and data modelling capabilities.

Maayan Kellerman (Bank of Israel) presented BoI's "SMARTer" project. The presentation emphasized the creation of a standardized and integrated microdata platform using SDMX, addressing challenges such as data silos, governance, and complex data access. Expected

challenges include locating/ developing appropriate tools, modelling issues, data sensitivity, linking micro datasets and performance issues.

Finally, David Barraclough (OECD) presented a real-world example on the use of annotations to enable macro-micro drilldown in visualisations. The proposed mechanism includes defining uncoded dimension - unique identifiers for microdata and coded dimensions to indicate aggregates as well as standard data flow annotations.

The session fostered vivid discussions and valuable contributions by the audience. Important topics emerged on linking, interoperability with other standards (DDI), integration, and more importantly, reconciliation of micro and macro data. The session highlighted the need to follow a user-centric approach, connecting the data to the user and the necessity for collaborative efforts to advance in this area.

#### Recommendations for follow-up actions

- Continue the work of Cross Working Group on micro data to have a first version of the concept oriented guideline for the SDMX global conference.
- Continue the workstreams work on modelling and delve into selected use cases (financial datasets, surveys and registers), expanding on the work initiated in the side event meeting.
- From a technical perspective, reflect on the feedback collected in the Slido poll during the session and engage in relevant discussions within CWG and with the TWG.