

CONFERENCE OF EUROPEAN STATISTICIANS

Workshop on Statistical Data Collection

10-12 October 2017, Ottawa, Canada

WP.2-7

26 September 2017

Moving to Automated Collection using SDMX

Gareth McGuinness (International Monetary Fund)

gmcguinness@imf.org

Abstract

The IMF Statistics Department is moving from a collection model where country authorities submit Excel-based reporting forms, to one where countries disseminate data in SDMX format, allowing automated machine-to-machine retrieval by the IMF. Around 40 countries of varying statistical capacity have set up such functionality via the IMF's Data Standards Initiatives, which require (Special Data Dissemination Standard Plus) or recommend (enhanced General Data Dissemination System) SDMX dissemination. Challenging aspects of this change have included assisting country authorities to adopt an SDMX infrastructure, moving from presentation templates to code-based collection instruments, reassuring data providers that the new "pull" model is just as robust and reliable as the previous "push" model, and managing security of confidential data. Successfully overcoming these challenges has led to many of these countries reporting data using this new model.

Moving to Automated Collection using SDMX

Gareth McGuinness^{*1}

International Monetary Fund, gmcguinness@imf.org

The shape of the change

Since the 1940s, the International Monetary Fund (IMF) has been collecting macroeconomic and financial data from countries, for use internally and to redistribute these data to IMF member countries and the global public.

The collection paradigm for most of this history has been the report form, a tabular presentation of data that has, in recent decades, most often been provided to the IMF Statistics Department in Excel files. To help increase the efficiency of the global official statistics environment, reduce reporting burden for countries, and improve quality and timeliness of IMF data, the department is focusing on the use of automated machine-to-machine collection. This means moving from a collection model where country authorities submit Excel-based reporting forms, to one where countries disseminate data in Statistical Data and Metadata Exchange (SDMX) format, allowing automated machine-to-machine retrieval by the IMF.

Around 40 countries of varying statistical capacity have set up such functionality as part of their participation in the IMF's Data Standards Initiative. Recent developments in this initiative have led to the Special Data Dissemination Standard Plus (SDDS Plus) requiring countries to disseminate using SDMX and the enhanced General Data Dissemination System (e-GDDS) recommending SDMX dissemination.

Challenging aspects of this change have included moving from presentation templates to code-based collection instruments, assisting country authorities to adopt an SDMX infrastructure for dissemination, reassuring data providers that the new "pull" model is just as robust and reliable as the previous "push" model, and managing security of confidential data.

From forms to codes

Moving to SDMX to data exchange between country authorities and the IMF requires a shift from form-based reporting to code-based reporting. For many years, the principal means of countries providing data to the IMF Statistics Department has been by completing and reporting Excel templates provided by the IMF. Many countries submit the Excel files using the IMF's Integrated Collection System (ICS), which provides a secure online mechanism for submitting the files. Some countries provide the files by email, which are then uploaded to the ICS by IMF staff.

While the report forms contain the IMF's codification of the datasets for submission, they are principally intended as a visual aid for compilers showing the relationships between the various data items. Being in Excel, the forms also offer data providers the flexibility to edit the data directly on the report form. However,

¹ The views expressed herein are those of the author and should not be attributed to the IMF, its Executive Board, or its management.

the forms also have several draw-backs. Providers can make changes to the template rendering them unreadable (even though this requires circumventing protections placed on the file by the IMF). The report forms are highly susceptible to transcription and other errors – much time is spent discussing such errors with country authorities. The Excel report forms are also not easy to populate automatically, as such procedures often require specific cell references making them complex to create and cumbersome to maintain. Finally, it is not straightforward for the IMF to automate consumption of data, often requiring manual intervention by STA staff before a report form can be successfully consumed.

By contrast, the SDMX data exchange paradigm is based on the use of mutually agreed coding structures and content. For most of its collections, the IMF accepts data coded using multilaterally-agreed data structures (where they exist) or using a time series style data structure formed around the IMF's internal data catalogue.

The use of this structured and standardized approach supports the unambiguous identification of data being provided by country authorities, while also enabling consumption of data by multiple international organizations and other users, reducing the burden on the data providers. Generating code-based output directly from data providers' production systems (usually as an xml file) is also much easier than for a highly-formatted Excel file, and such output is much easier to consume by the collecting agency, reducing the chance of errors in both provision and collection. What is lost from this approach is the ability for humans to easily read the data in the SDMX machine-readable format, which can make troubleshooting errant files more difficult. In addition, there is no easy way to manipulate data once an SDMX file is created, so any changes must be made in the originating database, although as this is best practice in any case, this could be viewed as a positive factor for SDMX data exchange.

The IMF Statistics Department's experience working with countries has shown that the benefits from moving to code-based SDMX exchange outweigh the disadvantages of such an approach. Countries appreciate the opportunity to reduce their burden, by automating the production of the data to be exchanged, while providing the data in a standardized format that can be retrieved by many users.

Country adoption of SDMX infrastructure

The sponsors of SDMX define Statistical Data and Metadata eXchange as “an international initiative that aims at standardising and modernising (“industrialising”) the mechanisms and processes for the exchange of statistical data and metadata among international organizations and their member countries.”²

The seven international organizations that sponsor SDMX are the Bank for International Settlements (BIS), the European Central Bank (ECB), Eurostat (Statistical Office of the European Union), the International Monetary Fund (IMF), the Organisation for Economic Cooperation and Development (OECD), the United Nations Statistical Division (UNSD), and the World Bank. Each of these organizations is a major force “at world and regional levels in the collection of official statistics”.³

A key part of the statistical process for these organizations is collecting data from member countries' government agencies and other national official sources. The development of SDMX was intended to provide the technical underpinning to support this process, by assisting country authorities to disseminate data in a

² https://sdmx.org/?page_id=3425.

³ https://sdmx.org/?page_id=3425.

format that could be automatically harvested by international organizations and other data users. Disseminating data using SDMX is not a trivial undertaking, particularly for those agencies without existing SDMX skills and infrastructure.

Many agencies in advanced economies have invested in building SDMX infrastructure and disseminate data using the format. In Europe, for example, central banks and national statistical offices (NSOs) are required to provide data to pan-European agencies using SDMX. In North America, many official statistical offices have independently developed SDMX capabilities. Many of these European and North American countries are disseminating data in SDMX as part of their adherence to the IMF's SDDS Plus standard.

Once you go outside the realm of advanced economies most countries are yet to develop significant SDMX infrastructure, although knowledge, understanding, and interest is growing. To tap into this burgeoning appetite, the IMF has sought opportunities to provide such countries access to cloud-based infrastructure containing straightforward tools to disseminate data using SDMX.

In Africa, the IMF has partnered with the African Development Bank (AfDB) to use their Open Data for Africa platform (ODP) for countries to disseminate SDMX.⁴ Country authorities can upload data to the ODP using Excel, and the ODP provides a full suite of dissemination functionality, including dashboards, visualization, data queries, and the ability to publish in SDMX. 14 countries in Africa are disseminating data in SDMX on the ODP, allowing them to cease traditional form-based reporting, as the IMF can retrieve the data automatically using the SDMX application programming interface (API) provided by the ODP.

For countries outside of Africa, the IMF has made available the cloud-based IMF SDMX Central tool.⁵ This website provides a streamlined process to convert data from Excel to SDMX. Once converted, country authorities simply need to post the SDMX files on their website, from where the IMF can automatically harvest the data. Eleven countries are using IMF SDMX Central to convert data into SDMX for dissemination on their websites.

From “push” to “pull”

Country authorities are used to proactively submitting data report forms to the IMF, either via e-mail or using the ICS System, and receiving a notification that their submission had been received. Under the SDMX approach, country authorities simply disseminate data and wait for the IMF to collect it. Currently, the IMF does not send any acknowledgment when updates are made, as dissemination is not tied to a specific individual.

This places the data provider in a position where they have to trust that the IMF has undertaken its responsibilities to collect the data that has been disseminated. When discussing with providers in advance of adopting this approach, we stress to be a responsibility to provide data to the IMF is met by disseminating in SDMX format. However, to ensure no ambiguity exists, the IMF is considering options for notifying providers once we harvest the SDMX data they have disseminated.

After several cycles of providing data in this manner, data providers have expressed reassurance that the pull model is just as robust and reliable as the previous push model.

⁴ See <http://dataportal.opendataforafrica.org/> for more information.

⁵ <https://sdmxcentral.imf.org/>

Private and confidential

Much of the collection work undertaken by the IMF is to retrieve data for use internally within the Fund. Some of these data are highly confidential, in which case the use of a public, open data dissemination paradigm is obviously not suitable. The AfDB and IMF worked with the ODP software provider to implement a secure mechanism for obtaining SDMX data from private datasets within the ODP, allowing the automated exchange of machine-to-machine readable data.

Currently, other country authorities wishing to submit confidential SDMX data must disseminate data via a secure web service within their own IT infrastructure, so that the IMF can pull the data confidentially. This requires the agency to have sufficient skills, technology, and resources to establish such a service, which creates a potential barrier to adoption of this approach.

Continued expansion

The IMF's experience of introducing SDMX-based data provision has demonstrated the feasibility of widespread adoption of this approach to exchanging data between country authorities and the IMF. Many countries are reporting data to the IMF using this new model, ranging from the most advanced economies with significant statistical capacity to low income countries facing skills, technology, and resource constraints. As a result of this promising progress, the IMF is continuing to expand its activities in this area.

This is not to say the future holds no challenges. As noted above, more work is needed on options for confidential transmission and receipt of data using SDMX. And work continues to improve confidence in the new paradigm, by successfully demonstrating the effectiveness and efficiency for both data providers and collecting organizations.