Distr.
GENERAL

WP.22
9 May 2012

ENGLISH ONLY

**UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE (UNECE)**
**CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION**
**STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD)**
**STATISTICS DIRECTORATE**

**Meeting on the Management of Statistical Information Systems (MSIS 2012)**
(Washington, DC, 21-23 May 2012)

Topic (iv): Collaboration

# FUTURE DEVELOPMENTS IN PGI WITH SDMX-BASED DATA COLLECTION

**Supporting Paper**

Prepared by René Piché, International Monetary Fund (IMF)[1]

## I. Introduction

1. The Principal Global Indicators (PGI) website provides internationally comparable data for the Group of 20 economies (G-20) and the five members of the Financial Stability Board that are not part of the G-20 to facilitate the monitoring of economic and financial developments for these jurisdictions. The PGI website was launched in April 2009 in response to the global financial crisis and is an undertaking of the Inter-Agency Group on Economic and Financial Statistics (IAG).[2] It provides information on major economic indicators that are available at participating international agencies covering financial, governmental, external, and real sector data, with links to data available at websites of international and national agencies. The site is available at http://www.principalglobalindicators.org/.

2. The various aspects of recent PGI development and enhancements were presented at the May 2011 MSIS Meeting.[3] This paper reviews the initiatives underway for the PGI to better leverage the SDMX standards for delivering faster and better data while reducing the reporting burden on national data providers, and

[1] The views expressed herein are those of the author and should not be attributed to the IMF, its Executive Board, or its management.
[2] The IAG comprises the Bank for International Settlements, the European Central Bank, Eurostat, the International Monetary Fund (Chair), the Organisation for Economic Co-operation and Development, the United Nations, and the World Bank. It was established in 2008 to coordinate statistical issues and data gaps highlighted by the global crisis and to strengthen data collection.
[3] See "Principal Global Indicators – Innovation in communication of official statistics" available at http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.50/2011/wp.21.e.pdf

duplication of efforts in international institutions. It is organized as follows: section II describes the governance structure for the PGI website, section III describes the current set up for the PGI database and website and its limitations, section IV reviews the main objectives behind the redesign of the PGI data processes and first steps in implementation, while section V describes the main challenges for the long-term implementation of an SDMX-based website and how they are being addressed.

## II. The IAG and the G20 Data Gaps initiative

3. The PGI website is one of the key initiatives described in *The Financial Crisis and Information Gaps* report jointly prepared by the Financial Stability Board (FSB) Secretariat and International Monetary Fund (IMF) staff. The report was endorsed by the G-20 Finance Ministers and Central Bank Governors in November 2009.[4] It contains 20 recommendations for closing information gaps (the so-called G-20 Data Gaps Initiative) and the implementation of these recommendations is coordinated by the IAG.

4. The development of the PGI website falls under recommendation 20 of the Data Gaps Initiative, which is related to improving communication of official statistics. Recommendation 20 states: *The G-20 economies to support enhancement of the PGI website, and close the gaps in the availability of their national data. The IAG should consider making longer runs of historical data available*. The IMF is the lead agency for implementing Recommendation 20 and hosts the PGI, working in close collaboration with other IAG agencies for data collection and ongoing enhancements it is making to the PGI website.

5. The IAG meets regularly (through videoconference) to oversee the implementation of the recommendations under the G-20 Data Gaps Initiative. These meetings provide the opportunity to review the work program for the PGI website and guide the short- and medium-term priorities for its further enhancement. In the short-term, the high priorities for enhancing the PGI website are to close gaps in the availability of national data, and improve data dissemination timeliness and quality. Ways of improving the efficiencies in data supply are being investigated, including the expanded use of the SDMX standards for the exchange and dissemination of official statistics.

## III. The Current PGI Process and its Limitations

6. The PGI website is hosted and managed by the IMF. When the website was initially developed, its main objective was to quickly provide access to G-20 data in comparable presentations and units of measure. The short development timelines suggested that it was best to make use of the existing IMF Statistics Department (STA) databases, which contain data reported to STA by IMF member countries, for populating the PGI database. IMF-sourced data were supplemented by selected data from other IAG agencies, such as data from the *Main Economic Indicators* (*MEI*) from the Organisation for Economic Co-operation and Development (OECD), and Quarterly External Debt Statistics (QEDS) from the World Bank. Both *MEI* and QEDS data are regularly used by IMF economists and were available from internal IMF databases supported by existing data exchange protocols with OECD and World Bank. The European Central Bank (ECB) and Eurostat also contribute data to the PGI, as they are the IMF data source for selected European Union country data on monetary statistics (ECB) and national accounts and government finance statistics (Eurostat).

7. The PGI data mostly correspond to the data disseminated monthly in the IMF *International Financial Statistics* (*IFS*) publication. However, in order to meet the more timely requirements of the PGI, its database is updated directly from STA's data production environment, where data are updated on an ongoing basis, while the *IFS* public database is updated once a month.

---

[4] The various FSB-IMF reports on the G-20 Data Gaps Initiative are available from the PGI website at http://www.principalglobalindicators.org/about_iag.aspx .

8. Notwithstanding this improvement, the original data sourcing approach rapidly showed its limitations. The timeliness of IMF-sourced data does not match the timeliness of data disseminated by countries on their national website. This difference became obvious when STA started to regularly compare the timeliness of PGI indicators with the timeliness of data disseminated by PGI economies subscribing to the IMF Special Data Dissemination Standard (SDDS). Twenty-two of the twenty-five PGI economies subscribe to the SDDS. [5] SDDS subscribing economies undertake to disseminate data for 18 data categories on a SDDS-prescribed National Summary Data Page (NSDP) at set periodicity and timeliness. The NSDP is then used by the IMF to monitor countries' observance of the SDDS. An additional timeliness issue was related to the data exchange arrangement for the OECD *MEI*; data were transmitted once a month to the IMF, while the *MEI* data, similarly to *IFS*, are updated on a daily basis in the OECD production environment.

9. The SDMX initiative provides technical standards for data exchange and their implementation for the PGI website could significantly increase the timeliness of and the access to the PGI data. The SDMX initiative is sponsored by the same seven international agencies that established the IAG and the PGI website and IAG members suggested making greater use of SDMX standards to improve data sharing amongst IAG agencies and, as a result, improve the content and timeliness of PGI data. The success of the SDMX-fed iPhone/iPad PGI application highlighted the benefits of further leveraging the SDMX standards.

10. The OECD *MEI* provided the first opportunity to leverage SDMX standards, as the data are available from the OECD.Stat data warehouse, which includes an SDMX web service. However, some changes were required to the underlying data model used for the *MEI*, as there was no explicit dimension identifying the unit of measure, making it difficult to develop an error-proof way of reading the data and mapping it to the internal IMF environment. *MEI* data are now updated daily on PGI, using a pre-defined SDMX query that pulls data from OECD.Stat and automatically updates the PGI database.

11. The next step involved examining how to better utilize data available from the existing ECB and the forthcoming Eurostat SDMX web service. This review highlighted the limitations of the current design of the PGI database and website. PGI data are largely fed from the IMF *IFS*, but presented in comparable units of measure (e.g., growth rate and percent of GDP). IMF database managers created additional time series in comparable units of measure to respond to the needs of PGI. These data are pushed daily to the IMF internal data warehouse, and extracted for dissemination on the PGI website. For OECD *MEI*, colleagues at the OECD created these additional time series in comparable units of measure and included them in the daily *MEI* data exchange protocol. However, this approach (i.e., provide data transformation in percent of GDP) could not work in a data sharing environment where the most timely GDP data is pulled from Eurostat's web service and the balance of payments from ECB's web service. Therefore, the IMF PGI Team is reviewing the PGI data work flow to fully leverage the potential offered by the SDMX standards.

## IV. A New Data Architecture for the PGI Website

12. The current data architecture for the PGI database has reached its useful limit. A new PGI data processing environment is needed to support more flexibility in the selection of the most timely data source as well as providing ways to disseminate data in comparable units of measure while using the timeliest data for both the numerator and the denominator.

13. For example, in the current environment, it is difficult to change the data source for the balance of payments data disseminated in the monthly balance of payments (BOP) electronic publication of the IMF. These BOP data are very detailed (hundreds of time series) and reported by the member country directly to the

[5] The Euro Area is not a subscriber of the SDDS, as it is not an IMF member country. However, it complies with the requirements of the SDDS, including the dissemination of a National Summary Data Page.

IMF's Statistics Department. The level of detail of the data often implies some delay in the provision of the data to STA. On the other hand, data disseminated on the SDDS National Summary Data Page are usually more timely, but too sparse to meet IMF reporting requirements. However, these NSDP data are sufficient to meet the dissemination requirements of the PGI website.

14. The new PGI database environment will provide flexibility in the selection of the most timely official data source by removing reliance of IMF internal data processing environments, which are geared towards the production of much more detailed data than what is needed for PGI. The objectives are to identify the most timely available data source, regularly pull these data in the PGI database, monitor on an ongoing basis that they meet the requirements of the relevant PGI indicators by comparing them with data from other known data sources, and apply the relevant PGI transformations prior to the dissemination on the PGI website. In that context, the qualifier "available data" implies computer-friendly official data, or more precisely, data available in SDMX formats.

15. The new PGI data processing architecture will provide an environment that will maximize the potential of using data pulled using SDMX web services. It cannot simply be a data dissemination hub, as logic and controls must be implemented in a processing environment to select the most timely data source, ensure it corresponds to the concepts that should be disseminated, and, where needed, restate the data in comparable units of measure. As such, it may be conceived as an enhanced version of the hub approach developed for the SDMX proof of concept implementation, the Joint External Debt Hub.[6]

16. In a first step, the objective is to maximize data sharing across IAG agencies to populate the PGI website. The IMF has already developed a generic SDMX data retrieval/update module. This module provides three functions. It allows running SDMX data queries at set intervals against data sources (web services) to pull data on an incremental basis.[7] It provides a generic mapping tool, so that the SDMX data file could be mapped to the IMF internal coding system. This is a must in the early stages of SDMX adoption, as there are multiple Data Structure Definitions (DSDs)[8] currently being used for encoding data for the same statistical domain. Finally, the module provides a data validation and review tool that compares the SDMX-sourced data with data in the internal databases prior to loading the data internally.

17. In the short-term, it is not possible to replace the data that the IMF collects for *IFS* and other datasets with similar data pulled from other IAG agencies. While similar, other IAG members' datasets are not identical and, for some statistical domains, some key series that the IMF collects are not available from other IAG members. However, these IAG data meet the requirements of the high-level indicators available on the PGI website.

18. As a result, the IMF is now developing a data processing system whereby data from multiple sources would be stored in an input environment, their respective timeliness and ability to meet PGI requirements analyzed, and the most timely official data source would be selected as the PGI indicator for a specific country/indicator. The approach will support providing the data transformations that are required for the PGI website, as well as the types of data validation that are expected for such a flagship inter-agency website.

---

[6] Please see the OECD paper presented at the 2004 MSIS conference and available at: http://www.google.com/url?q=http://www.unece.org/stats/documents/ces/ac.71/2004/4.e.pdf&sa=U&ei=PC2gT8-6Keay0AGR5-WOAg&ved=0CCYQFjAJ&usg=AFQjCNHClBQ4-dGoBu2yiXk6miFuyERHcw

[7] The SDMX technology allows setting SDMX data queries that return data updates since the web service was last queried. This increases the speed of data transmission and facilitates the data validation process, as the files being processed are usually much smaller.

[8] A DSD is the SDMX encoding structure, where data are defined using dimensions (and their associated code lists) and attributes (coded or free format) according to the SDMX information model.

## V. Challenges for Longer-term Objectives

19. In the longer-term, some of the high-level indicators disseminated on the PGI website could be fed from an SDMX-formatted NSDP. These data correspond to the timeliest official data available, as the SDDS prescribes that SDDS data disseminated by countries must be made available the same day on the NSDP. However, the current prescribed format for the NSDP is html, which makes it very difficult to decrypt by computers and requires a high-level of customization in doing so. The IMF will be investigating modifying the format of the NSDP to utilize the SDMX format. This would make the NSDP fully computer-readable and help improve the timeliness of the PGI. This change will require active consultation with the SDDS subscribers and is expected to take one to two years before being completed.

20. As a step to further promote the adoption of SDMX, the IAG is actively coordinating the development of "DSDs for Global Use". According to the "SDMX Action Plan 2011 to 2015" released by the SDMX Sponsors[9], the SDMX initiative will deliver DSDs for global use for two key statistical domains: balance of payments and national accounts statistics. This will respond to the call from many official data compiling agencies for a set of internationally agreed DSDs for reporting data to international organizations (IOs). In the case of balance of payments, the DSD will cover the provision of data for over 18 data collection exercises currently conducted by IOs using a set of common dimensions and code lists. This is expected to promote the adoption of SDMX and facilitate the data reporting to IOs. The balance of payments DSD will be available for pilot-testing by countries during June 2012, while the DSD for the complete set of national accounts will be available by the end of 2012.

21. The main objective of the current redesign of the PGI data processing environment is to position the PGI website to fully leverage the benefits of the larger adoption of the SDMX standards by IOs and member countries. As a result, the PGI website will be able to deliver the timeliest data, while at the same time minimizing the response burden on countries through more efficient data sharing amongst IOs.

[9] Please see http://sdmx.org/wp-content/uploads/2011/10/SDMX-Action-Plan-2011_2015.pdf .