

The Current State of Play

11. The Fund has taken steps to address some of the challenges regarding data, with significant recent efforts in this regard. But will these be sufficiently holistic and well-funded to be sustainable? Will they support the organization in being proactive, not just reactive, in the realm of data?

12. To address such questions, the following sections present the evidence gathered for this evaluation on: (i) data issues pertaining to the Fund’s core strategic functions of surveillance and lending; (ii) efforts to address information gaps; (iii) the quality of data; (iv) internal data management practices; and (v) data dissemination and international cooperation. The key theme throughout is how data support the Fund’s strategic operations, and the sections are prioritized accordingly; for example, good data management is a means, not an end in itself, to better enable data to play its role in surveillance and lending.

A. Meeting the Fund’s Core Operational Needs

The global financial crisis changed the Fund’s approach to surveillance and created a surge in demand for more data in new areas.

13. The crisis underscored the importance of mainstreaming macrofinancial analysis into bilateral surveillance and better integrating bilateral with multilateral surveillance. It thus provided considerable motivation for revamping the IMF’s toolkit for detecting macrofinancial risks and risks associated with global interconnectedness.

14. The resulting, more integrated conceptual frameworks all depend heavily on data, making it increasingly difficult to disentangle the data issues related to the three main branches of surveillance—bilateral, multilateral, and financial¹²—and to lending. Each of

them, to varying degrees, face the fundamental data dilemmas of trade-offs between: accuracy versus timeliness, granularity versus aggregation, international comparability versus country specificity, and confidentiality versus transparency. And according to the IEO’s survey of IMF staff, each of these core operations is adversely affected by data deficiencies (Figure 1).¹³ Despite considerable overlap, the following discussion considers separately the three types of surveillance and also lending, as each poses some unique data issues.

(i) Bilateral surveillance

The fundamental question is whether data are adequate for surveillance . . .

15. Bilateral surveillance¹⁴ is the cornerstone of the Fund’s operational work. The workhorse accounting structure underlying this surveillance is the financial programming framework, an integrated macroeconomic framework that demonstrates how the data for a country’s various economic sectors—real, monetary, fiscal, external—are interlinked, allowing the Fund to construct a picture of the overall economy.¹⁵ But any analysis based on this framework can only be as good as the data supporting it, which will also reflect the approaches used by IMF staff to address data gaps and inconsistencies.

is technically not an independent, third “branch” of surveillance, but rather, as articulated under the Integrated Surveillance Decision, an integral part of both bilateral and multilateral surveillance. Nevertheless, in practice, the IMF has often treated financial surveillance as a separate entity. See, for example, IMF (2012c).

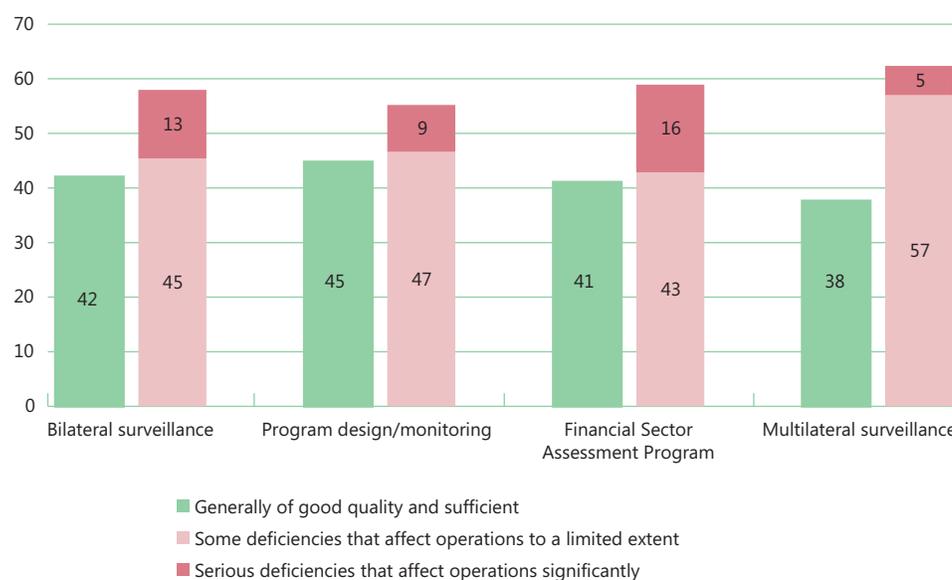
¹³Although this report focuses on surveillance and lending, data deficiencies also can have a bearing on other important areas of Fund work, such as calculating quota shares to guide decisions regarding relative size and distribution of members’ actual quotas.

¹⁴Most notably, the Article IV consultations that the IMF conducts (typically) on an annual basis with each of its member countries.

¹⁵A shortcoming of the financial programming framework is that the financial sector is still not fully integrated into the framework.

¹²The Articles of Agreement only recognize two forms of surveillance—bilateral and multilateral. Thus, financial surveillance

Figure 1. Staff Perceptions of Data Quality and Availability for Operations
(In percent)



Source: IEO Survey.

16. The Greek crisis provides a compelling illustration of the importance of high-quality data for IMF surveillance—and for global economic stability. Greece’s debt crisis erupted in late 2009, when a new government revealed that the projected fiscal deficit and government debt had been grossly understated by the previous government. This disclosure alarmed financial markets, ultimately precipitating a “sudden stop” of financial inflows and the need for a bailout. But this was not the first time that Greece’s data had been found wanting: a 2004 Eurostat report showed that Greek government deficit and debt figures had been misreported as far back as 1997, and that the deficit had not been below the Maastricht limit of 3 percent in any of these years (Eurostat, 2004). How did the Fund miss the warning signals of problematic data?

... and the IMF staff’s answer is a qualified “maybe.”

17. Greece has not been alone with respect to flawed data. Data deficiencies have adversely affected the bilateral surveillance of all categories of countries—advanced, emerging, and low-income countries (LICs), albeit to different degrees (Figure 2)—with almost 60 percent of IMF staff survey respondents noting such deficiencies regarding their primary country assignment. Lack of data or inadequate quality were each cited by about 90 percent of these survey respondents.

These findings on data deficiencies and the adverse impact on surveillance reinforces those of the Fund’s 2014 and 2011 Triennial Surveillance Reviews and various past IEO evaluations (Annex 4).

18. A number of cases have been documented in which problematic data reporting hampered the Fund’s conduct of surveillance and led to faulty analysis. Reichmann and Monasterski (2016) discuss about a dozen such country cases that have arisen since the 1990s.¹⁶ It is also highly likely that many data-induced shortcomings have left no traces in Fund documentation, and that in most such cases, the Fund could not have detected data problems that might affect its analysis, absent the explicit admission of the member country. Australia’s Bureau of Statistics, generally considered among the best, is one notable example. It admitted to a benchmarking issue in its 2010 and 2011 official employment figures, overstating the strength and the weakness, respectively, in the labor market. This led to perceptions that the Reserve Bank’s decision to push up rates in 2010 and to reverse course the following year could have been influenced.¹⁷

¹⁶Most cases where the Fund has documented data that have undermined analysis have occurred in the context of Fund-supported programs, reflecting the much greater attention the Fund gives to data when its own financial resources are at risk.

¹⁷*Sydney Morning Herald*, July 3, 2012. See also on this issue: <http://www.abs.gov.au/AUSSTATS/abs@nsf/Previousproducts/6202.0>Main%20Features2Apr%202012?opendocument&tabname=Summary&prodno=6202.0&issue=Apr%202012&num=&view> and <http://www.rba.gov.au/publications/smp/2012/aug/box-e.html>.

Figure 2. Staff Perceptions of Data Quality and Availability for Bilateral Surveillance
(In percent)



Source: IEO Survey.

19. The most common reason for data deficiencies, according to the survey of staff, is a country's limited capacity (including cost constraints), but a more troubling reason, cited by close to 20 percent of staff survey respondents, is the authorities' unwillingness to provide the data. While in some cases non-provision was due to cost considerations, more than half of such instances were due to confidentiality concerns about how the IMF would handle the data. The survey of data providers also indicated a strong regional component, with about 40 percent of respondents from Asia and from Middle Eastern oil-exporting countries expressing concerns about confidentiality. Worse still, 10 percent of IMF staff (with higher numbers for those working on emerging markets) claimed that intentional manipulation of data was responsible for data inadequacies.¹⁸

The IMF has a broad-ranging toolkit to address data deficiencies . . .

20. What instruments does the IMF have at its disposal to question official statistics and to address data deficiencies during the conduct of bilateral surveillance? In addition to staff judgment and experience,

¹⁸Intentional manipulation is often a case of Goodhart's Law, the popular formulation of which is "When a measure becomes a target, it ceases to be a good measure." Goodhart's Law (named after an economist who was a member of the Bank of England's Monetary Policy Committee) refers to the vulnerability of a statistical indicator to manipulation once it is used to define a policy target.

data inconsistencies are often discovered through the use of the IMF's financial programming framework.¹⁹ Problems can also be detected by checking flow data against stock data. In about half of country cases with data deficiencies, staff survey respondents said that they had to come up with their own estimates for the problematic data.

21. In the context of Article IV consultations, IMF staff are expected to candidly assess the adequacy of member countries' statistics for surveillance (IMF, 1995a), with major deficiencies discussed in the main text of the Article IV staff report, along with a more detailed review in a Statistical Issues Appendix (SIA). This guidance is aimed at raising the profile of data issues in surveillance and to prompt corrective action if warranted, with staff proposing remedial measures or technical assistance, if needed.

. . . but doesn't always deploy it effectively

22. The Fund's regular reviews of data provision had identified a number of problems with the SIA arrangement (Box 1):²⁰ (i) lack of candor in staff's discussion of data deficiencies, with an "upward bias in the

¹⁹Until recently, financial programming was typically not applied to advanced economies, a factor which may have contributed to the undetected buildup of the large imbalances prior to the financial crisis.

²⁰The system currently in place was approved and reviewed, respectively, during the 2008 and 2012 reviews of data provision to the Fund for surveillance (IMF, 2008 and 2012b).

Box I. A, B, or C? Grading a Country's Data Adequacy for Surveillance

The IMF's current framework for data provision for surveillance was first defined in 1995 (IMF, 1995b), with some amendments since then. One of the elements that was added to Article IV consultations was a Statistical Issues Appendix (SIA), which includes an overall judgment on the adequacy of data provision for surveillance and, where relevant, a discussion of the implications of data deficiencies and recommendations for improvement. How candid have these judgments been?

Initially, assessments in SIAs included only two categories: adequate or inadequate. In 1995, 59 percent of a sample of 50 member countries were deemed to have "adequate" data provision.¹ This number climbed to 70 percent by 2003. In 2005, an intermediate category was added, allowing the following options for assessing data provision:

- Adequate for surveillance (A)
- Broadly adequate for surveillance, but with some shortcomings (B)
- Inadequate for surveillance (C).

By 2007, the sample percentage judged as either category A or B jumped to 90 percent, with only 4 percent

¹Note that the Article IV consultations for those deemed to have inadequate data provision were still completed.

given a category C rating (another 6 percent were unclear as to rating). Yet in a survey of mission chiefs that was conducted at the time, more than half of the respondents noted problems with data provision that hampered surveillance, with 40 percent reporting that their teams routinely had to prepare estimates for key data. Why this discrepancy between SIA assessments and mission chief responses?

The IMF's 2008 Review of Data Provision (IMF, 2008) concluded that the relatively rare use of Category C may have stemmed, in part, from concerns that it would undermine the relationship with country authorities or would raise questions about how surveillance can be conducted at all if data are "inadequate." This led to more changes in the framework, including elimination of the term "inadequate" from category C, recasting it as "Data provision has serious shortcomings that significantly hamper surveillance."

Did this change improve candor? By the time of the 2012 Review of Data Provision (IMF, 2012b), the percentage of countries classified as C had increased threefold to 12 percent. But a supporting survey of mission chiefs still indicated a huge discrepancy, with 59 percent of the respondents stating that important data deficiencies had hampered surveillance, thus suggesting continued reluctance to use the lowest rating. In response, new guidelines were issued in 2013 (IMF, 2013a) to encourage staff to use more candor and to provide clearer instructions to staff on the classifications.

characterization of data adequacy;" (ii) excessive workload on staff, deriving from the requirement to document and propose remedial measures to address data issues; (iii) poorly focused SIAs, with limited coverage; and (iv) lack of attention by the Executive Board.²¹ These problems were confirmed by the evaluation's survey and interviews; for example, according to the staff survey, formal data adequacy assessments are softened, as only 46 percent of cases in which data are perceived as inadequate are reported as such in SIAs. In response to the 2012 Review of Data Provision to the Fund (IMF, 2012b), the Fund issued a guidance note (IMF, 2013a), updating and clarifying how staff are to address any data shortcomings in the Article IV report. The updated guidance note aimed, in part, at improving compliance with the intent of the SIA, but—as discussed in [Annex 5](#)—little appears to have changed since it was issued.

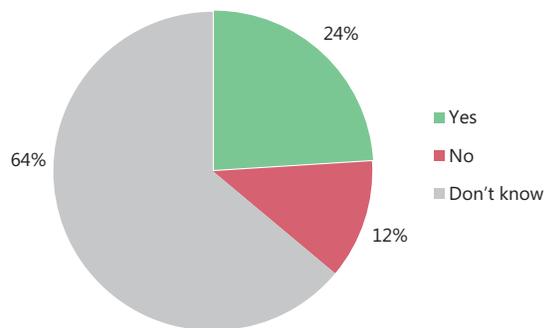
²¹As senior IMF staff members pointed out to the evaluation team, the Board's "lack of attention" to data quality issues at times reflected peer protection and political considerations.

23. Perhaps the most serious indictment of the SIA is its relative obscurity. Neither the Board nor IMF staff pays much heed to the SIA,²² with more than half of staff survey respondents noting that country teams lacked the resources and time to make thorough assessments. More worrisome, though, are the survey results of country authorities (i.e., data providers), fewer than one-quarter of whom were familiar with the SIA for their own country ([Figure 3](#)). This implies that the SIA does not provide the intended incentive for countries to improve their data.²³

²²As an example, for the 2007 United States Article IV consultation, the SIA noted that "Coverage of international capital flows in external sector statistics has been improved, with the June 2007 releases of BOP and IIP data on financial derivatives." This identical statement, highlighting 2007 data, appeared in the SIAs from 2008 until 2014, when an attentive staff member finally changed the date to June 2014. Of course, the U.S. SIA was not alone in conveying incorrect information. This evaluation found errors in a number of SIAs, as confirmed by country authorities during interviews.

²³Interviews with country authorities showed that a major reason for their lack of familiarity with the SIA was its issuance in a separate supplemental document for the Board meeting. Most of the authorities only read the main section of the Article IV report.

Figure 3. Did Your Country's Last Article IV Report Include a SIA?
(In percent)



Source: IEO Survey of Country Authorities and Data Providers.

24. The IMF has some legal scope to question countries on issues of data accuracy and availability, as embodied in the policies on misreporting and breaches of obligations under Article VIII. Potential breaches of obligations in the context of surveillance have occurred with some frequency; the 2012 Review (IMF, 2012b) noted that, in the preceding four years, “sustained concerns were raised with eight members about their willingness to share data required for Fund surveillance to the best of their ability.” Seven of those cases referred to nonprovision of data and were resolved within a year, while one related to provision of inaccurate data and resulted in the Board issuing a decision of censure, calling on the member to adopt remedial measures. This case—Argentina—remains unresolved (Box 2).²⁴

²⁴The most recent Board meeting on this breach of obligations was held in May 2015, with no change in the stance adopted by the Fund.

Box 2. Argentina and the Breach of Obligations

In January 2007, the Argentine government changed the personnel in charge of producing the consumer price index (CPI) at the National Statistics and Census Institute (INDEC). Concerns about the integrity of the CPI started to be voiced soon thereafter.

These concerns led several private sector entities and provincial governments to compute their own indices that showed consumer prices growing at significantly higher rates than those reported by INDEC. For example, Cavallo (2013), using data collected by the Billion Prices Project at MIT—such as prices in major supermarkets available on line between October 2007 and March 2011—replicated the components and weights of the CPIs in five Latin American countries (Argentina, Brazil, Chile, Colombia, and Venezuela). He found that, while the online price indices for the other four countries tracked well both the level and dynamic behavior of inflation, in Argentina they exceeded the official index by a factor of about three.

The apparent underreporting of CPI has implications for other key variables of significant importance for economic analysis. Inasmuch as the official CPI enters their calculation, measures of poverty or of the real effective exchange rate would be underestimated while the real growth of the economy would be overestimated. Underreporting would also have notable financial implications given Argentina’s issuance of inflation-linked peso bonds.

Failure to provide information to the Fund (i.e., a breach of obligation under Article VIII, Section 5) is defined to cover both non-provision of information and provision of inaccurate information (IMF, 2004b). The latter failure is considerably more difficult to substantiate than non-provision

of required information. This caused the Fund to take a measured approach to Argentina’s case. Eventually, in July 2011, Argentina was found in breach of its obligation due to inaccurate reporting of official data for the CPI and GDP. Technical assistance was offered, which resulted in several recommendations to correct the known deficiencies. However, delays in the implementation of key remedial measures led the Executive Board in September 2012 to issue a Statement of Concern, followed by a Declaration of Censure in February 2013. The latter established a timeframe for the adoption of the remedial measures, noting that failure to follow this timeframe could result in a declaration of ineligibility to use the general resources of the Fund.

Over the following two years, Argentina took some measures to address the difficulties, for example, introducing a new *national* CPI (the previous one was limited just to Buenos Aires) and revised GDP data (now with base year in 2004) in early 2014. These actions, however, did not fully assuage the Fund’s concerns, and further actions—related to the transparency of the process—were required before the Executive Board could withdraw the Declaration of Censure.

Argentina was an early subscriber to the SDDS and never lost that status. While the SDDS is a dissemination standard, not a quality standard, the Fund does issue an annual observance report as a form of monitoring, covering, among other metadata dimensions, the integrity and quality of the data. Argentina’s reports for 2012 through 2014 still indicated that Argentina met the integrity dimension, while the discussion of quality relegated the information on the Declaration of Censure to a footnote.

25. Greece is a timely case study as to how well (or how poorly) the IMF used its toolkit to deal with data deficiencies. In its assessment of Greek statistics in the run-up to the disclosures of false data, the Fund had been rather sanguine, with only occasional expressions of mild concern (“... data are adequate for surveillance but should be strengthened” was a common refrain). Not only was surveillance inadequate in this regard, but Greece had engaged in several important statistical milestones with the Fund: Fiscal Transparency ROSCs in 1999 and 2006, an SDDS subscription in 2002, and a full data ROSC in 2003—none of which brought to light the seriousness of the data problems. In 2010, in conjunction with Board approval of the Fund’s initial IMF-supported financial program with Greece, the Managing Director issued a report to the Board (IMF, 2010) on a breach of obligations under Article VIII, Section 5. The Board determined that Greece had taken sufficient remedial actions, including enacting a new law granting independence to the national statistical agency (ELSTAT). Yet—as discussed in [Box 3](#)—the independence of ELSTAT remains a concern five years later.

(ii) Multilateral surveillance

The perennial dilemma for multilateral surveillance data is international comparability versus country specificity . . .

26. Multilateral surveillance, always an important component of the Fund’s operations, took on an even larger role with the Fund’s adoption of the Integrated Surveillance Decision in 2012. This decision made the Article IV consultations a vehicle for both bilateral and multilateral surveillance, and helped to push forward work on policy spillovers and interconnectedness. In the context of this wider scope, data needs have grown markedly.

27. Multilateral surveillance (and cross-country analysis even more so) poses a special challenge for data, as it is predicated on comparability across countries²⁵—that is, on the same concept being defined and measured in the same way everywhere. But global standards do not necessarily suit local conditions. Particular country circumstances unavoidably result

²⁵In contrast to cross-country analysis, multilateral surveillance, which often focuses on spillovers and interconnections, does not always necessitate perfectly standardized cross-country datasets.

in different definitions, measurements, or coverage of economic variables, implying that concepts can be homogeneous across countries only to a certain degree. How can the IMF ensure that it is not “comparing apples and oranges” in its multilateral and cross-country work? And what does the analysis mean if the data are not fully comparable?

28. The IMF’s work on methodology and capacity development in the area of statistics has gone a long way to strengthen comparability. This is particularly true for the databases maintained by STA, which emphasize data that meet methodological standards. But the main sources of data for much of multilateral surveillance are area departments, where data are more likely to conform to country specificities or be based on staff estimates.

29. IMF staff recognize the challenge posed by lack of comparability. According to the IEO’s survey of staff, almost two-thirds of those engaged in multilateral surveillance claim that data deficiencies hamper surveillance to some degree, with lack of comparability across countries overwhelmingly cited as the main reason.²⁶ In sharp contrast to IMF staff views, the IEO survey of external data users indicated that almost 90 percent believed IMF data are comparable across countries,²⁷ a misperception that could pose a reputational risk to the Fund.

30. Problems with non-comparability have been highlighted in some IMF work. A notable example is Dippelsman, Dziobek, and Gutiérrez Mangas (2012), which underscores how failure to follow international guidelines for reporting of public sector debt (arguably one of the most important macroeconomic indicators) or inadequate documentation of data definitions “can lead to major misunderstandings in the fiscal policy debate.”²⁸

31. The present evaluation also considered comparability of data by examining the definitions of government deficit that were used for performance criteria in the 48 IMF-supported programs approved from January 2011 to April 2015 ([Annex 6](#)). The combination of different components resulted in nine different definitions in terms

²⁶The importance of comparability was confirmed by the 2014 Triennial Surveillance Review (TSR) survey of IMF mission chiefs; when asked to check those factors most important for increasing the use of cross-country studies in surveillance, 85 percent chose greater availability of *comparable* cross-country data.

²⁷By a slight margin, *World Economic Outlook (WEO)* data are (wrongly) believed to be more comparable than those of *International Financial Statistics (IFS)*.

²⁸The authors use Canada as an example to illustrate how different definitions of the public sector give rise to very different debt levels, with debt-to-GDP ranging from 38 percent on a narrow budgetary definition to 104 percent, using the consolidated general government.

Box 3. Greece: Policy-Based Evidence-Making and the Perils of Statistics

In October 2009, the Greek authorities disclosed to Eurostat that government deficit and debt data for 2005–09 needed to be revised. The revisions, completed in November 2010, were of an exceptional scale and resulted in the forecast deficit for 2009 moving from 3.7 percent of GDP to 15.4 percent of GDP, while the government debt moved from 99.6 percent of GDP to 126.8 percent of GDP.

The revisions reflected methodological weaknesses and unsatisfactory technical procedures in the Greek statistical system, but also inappropriate governance as exemplified by lack of clear responsibilities between institutions, diffuse personal responsibilities, and opaque empowerment of officials “which left the quality of fiscal statistics subject to political pressures and electoral cycles” (European Commission, January 2010). The contemporaneous Fund report on Breach of Obligations under Article VIII, Section 5 (IMF, 2010) stated that “the institutional setting at the time failed to ensure the independence and accountability of the National Statistical Service of Greece and other services involved in the production of fiscal data and public debt data.”

The problems reported in October 2009 were not new; in fact, both Eurostat and (to a far lesser degree) IMF staff had repeatedly indicated that Greek statistics were notoriously weak and plagued with problems. A 2004 report by Eurostat triggered “the first Greek data crisis” by showing that Greek government deficit and debt figures had been misreported since as far back as 1997, and that in none of these years had the deficit been below the Maastricht limit of 3 percent per year. Subsequently, as noted in Eurostat’s 2010 report, Greek government deficit and debt statistics were the subject of “continuous and unique attention for several years.”

On its part, Fund staff took a generally approving stance with only occasional expressions of mild concern. Congratulations were offered on the occasion of Greece completing the Fiscal Transparency ROSC in 1999, subscribing to the SDDS in 2002, and completing a data ROSC in 2003. On the latter, staff observed (IMF, 2003a) that: “. . . Statistics-producing agencies in the main have a legal and institutional environment that supports statistical quality. . . . All agencies demonstrate professionalism and are transparent in their practices and policies. In particular, the strong laws protecting confidentiality, rules for civil servants, and internal regulations of the central bank provide a clear set of ethical standards for staff. . . .” By 2006, in the Fiscal Transparency ROSC that was specifically prepared after the 2004 data crisis, staff was still maintaining a positive line: “Greek budget processes give assurances of

integrity about fiscal data through independent audit and recently strengthened statistical reporting.” In most consultation reports, staff took the general line of “. . . data are adequate for surveillance but should be strengthened” (e.g., the 2006 and 2007 Article IV consultations). Notwithstanding staff’s generally accommodating attitude, muted concerns about data weaknesses and calls for “further improvements” were an almost constant feature of consultation reports. Only by 2009, on the eve of the government’s acknowledgment of data deficiencies, did staff take a more forceful line, with the Article IV consultation for that year including a quite specific and detailed list of failings in Greek statistics.

Admittedly, even in the best statistical systems, it can be difficult to uncover truth when those in charge are bent on hiding it. Moreover, analyses may be unduly obstructed by insufficient financial sector data—as bemoaned in the 2005 consultation report—by differences across sectors in the coverage or definition of variables, by the complexity of intergovernmental fiscal relations in Greece, or by opaque financing activities—such as the off-market swaps in which Greece frequently engaged—but a more thorough application of the financial programming framework should have allowed staff to get an inkling of the sizable ongoing irregularities.

IMF staff had on several occasions (viz., the 2005 and 2006 consultations) called for granting independence to Greece’s national statistical service. This finally came about when the creation of an independent new office, ELSTAT, was made a condition of the 2010 program and part of the remedial action proposed by the authorities subsequent to the May 2010 report on breach of obligations. Under a new chief statistician, Greek government finance statistics were accepted by Eurostat without reservation in 2011–15, in contrast to the repeated reservations of the previous years—indicating a marked improvement in the quality of Greek statistics. However, ELSTAT’s independence continued to be challenged by vested interests (e.g., a criminal investigation was launched in 2013 against the chief statistician regarding revisions to historic data on public finances and debt), raising doubts about the underlying commitment of the country to truly independent statistics and pointing to risks of re-politicization in the future. Five years after being set up, ELSTAT’s independence was still not assured, as suggested by the Euro Summit of July 12, 2015 when “. . . Given the need to rebuild trust with Greece . . . safeguarding of the legal independence of ELSTAT . . .” needed to be included among the required measures.

of coverage, a heterogeneity that was further magnified by measuring the criterion on a cash or accrual basis, and in above- or below-the-line terms. This wide variety of concepts often carried over to the data reported in the *WEO*, thus putting paid to the notion that the numbers included in *WEO* are strictly comparable.^{29,30}

... highlighting the importance of having countries adopt standard data templates and—in the meantime—providing clear metadata.

32. These findings underscore the importance of providing clear metadata for all IMF-disseminated data. However, notwithstanding the IMF's guidelines to staff to this effect, an examination of a large sample of Article IV consultation reports showed that none of them complied with this requirement (*Annex 5*). An inattentive or impatient economist could simply download the desired data, compare apples and oranges, and draw the wrong policy conclusions.³¹ Nevertheless, even with excellent metadata, the diversity of definitions can greatly impair the ability to do cross-country work.

(iii) Financial surveillance

Data issues for financial surveillance are among the most challenging ...

33. Financial sector surveillance, in the aftermath of the global financial crisis, has become even more central to the Fund's core operations. However, data issues are particularly challenging here, given the sensitive (and often confidential) nature of the data, the need for granularity and comprehensiveness (e.g., “off-balance-sheet” exposures), and lack of consistency. Data are often

²⁹Nominal GDP provides another example of comparability issues in *WEO* data. While most countries still measure GDP using the 1993 System of National Accounts (SNA), some, including most of the advanced economies, have now moved to the 2008 SNA. Typically, GDP, as measured under the 2008 SNA, is larger than that under the older system (e.g., U.S. nominal GDP was almost 4 percent larger, while it is estimated that, were China to move to the newer system, its economic size could be as much as 16 percent larger).

³⁰The *WEO* makes adjustments to some data to improve comparability. For example, the *WEO* has migrated balance of payments data to the methodology used in the sixth edition of the *Balance of Payments and International Investment Position Manual (BPM6)*, even though many countries still submit data under the previous *BPM5* methodology. The *WEO* also converts data on a fiscal-year basis to a calendar-year basis.

³¹In interviews with external data users, many admitted that they use multiple (noncomparable) IMF data sources (*IFS*, *WEO*, country reports, Working Papers) to fill in missing data for cross-country studies.

nonexistent or opaque in some critical areas, particularly on cross-border linkages and the shadow banking sector.

... due, in large part, to the often market-sensitive nature and need for granularity of data.

34. Financial surveillance is constantly struggling with the tension between granularity and aggregation. Aggregate data can mask critical vulnerabilities—that granular data might reveal—and may not be usable with some of the Fund's new analytical tools. For example, network analysis (used to examine issues of interconnectedness) needs quite granular data. But the Fund's hands are essentially tied by its Articles of Agreement, as it cannot require countries to provide institution-specific data.

35. The Financial Stability Assessment (FSA), a component of the FSAP and a key instrument of the IMF's surveillance, illustrates the data challenges facing the IMF. According to this evaluation's survey of staff, the data collected for FSAPs are perceived as the most problematic.³² Just under a third of the survey respondents from the Monetary and Capital Markets Department (MCM) believed data were sufficient for conducting an FSAP exercise. Notably, almost 90 percent said data problems had hindered the conduct of stress tests, while about three-quarters said the analysis of potential cross-border spillovers was hampered by data problems. Poor quality data was cited by 40 percent of respondents.

36. A 2014 review of the FSAP (IMF, 2014d) made clear the role of data—in particular, availability and quality for stress testing—in underpinning (or undermining) the program's effectiveness. While many countries voluntarily provide these data to the FSAP team—subject to strict confidentiality protocols—the FSAP review noted that this practice is not universal (as confirmed by this evaluation's survey results), with advanced countries the least likely to share supervisory data. And even when the needed data are available, FSAP teams are typically not equipped to assess their accuracy or the quality of underlying assets.

... underscoring the importance of building trust, yet being candid about data limitations.

37. Considering the reasons why country authorities are loath to share data, about a third of MCM survey

³²Also, compared with staff working on the *WEO*, staff involved with the *Global Financial Stability Report* were much more likely to note problems with lack of data, comparability, and uncertain quality.

respondents cited legal constraints,³³ but another third cited issues of trust. Data providers who were interviewed regarding access to market-sensitive financial data noted that banking supervisors tend to trust, in order, other supervisors, central banks, the Bank for International Settlements (BIS), and only then the IMF.³⁴ Further, as noted by the 2014 Triennial Surveillance Review, “some of the Fund’s counterparts have become less willing to share data as the crisis has subsided.” This is clearly problematic for the Fund’s FSAP, which in such cases must rely on publicly available data and/or on stress tests conducted by supervisors and the banks themselves. According to some interviewees, the results from the FSAP team’s stress tests differed at times from those of the stress tests conducted by the authorities or the banks, largely because of differing access to data.

38. These findings highlight two issues: (i) there is a tension between the mandatory character of FSAs and the voluntary provision of the data they require,³⁵ and (ii) the limitations of the associated risk assessment need to be clearly communicated by the FSAP team. The evaluation survey of MCM staff is revealing in this regard, with 40 percent of respondents advocating mandatory data provision to help address data deficiencies, and only half agreeing that the Financial System Stability Assessment (FSSA) report had clearly noted the problems with data quality or access. To help address the first issue, the Fund could clarify its confidentiality protocols to the membership to encourage the voluntary provision of the needed information.³⁶ On the second issue, the 2014 FSAP review noted that the standard disclaimer on all FSSA reports should be expanded to highlight any data limitations. But this evaluation found no change in the standard disclaimer in the most recent FSSA reports, including some with serious data access and quality issues.

39. The Fund has made notable strides, nonetheless, with respect to data needed for financial surveillance.

³³Some countries with legal constraints find ways to allow the FSAP team to “access” the data without actually violating the law (e.g., letting the FSAP team into the room to watch the conduct of supervisory stress tests).

³⁴The 2013 IEO evaluation, *The Role of the IMF as Trusted Advisor*, also found that country authorities placed more trust in the BIS than the Fund in the handling of confidential data (IEO, 2013).

³⁵In September 2010, the Executive Board decided to make the Financial Stability Assessment (FSA) mandatory for systemically important financial sectors in response to the shortcomings revealed by the financial crisis. Previously, all FSAs, as part of an FSAP exercise, were conducted on a strictly voluntary basis.

³⁶See “Confidentiality Protocol—Protection of Sensitive Information in the Financial Sector Assessment Program,” IMF, *Selected Decisions*, Thirty-Second Issue, p. 108.

Substantive progress has been made since the global crisis on collecting data on Globally Systemically Important Financial Institutions (G-SIFIs), the nonbank financial institutions (NBFIs), shadow banking, and Financial Soundness Indicators (FSIs).^{37, 38} The Fund’s efforts to collect data on NBFIs and shadow banking is particularly important to allow the Fund to expand its coverage of stress tests to the nonbank sector (an increasingly important player in many countries’ financial sectors), and to help member countries limit regulatory arbitrage, a potential precursor for a future crisis. The Fund has also developed new analytical tools that benefit from the expanded set of financial data.

(iv) Use of Fund resources

Data deficiencies can affect program design and monitoring . . .

40. Data quality and availability are also extremely relevant for IMF lending. Staff must be able to count on information adequate to allow the design of a program fit for the intended purpose. This has usually been the case, but in some instances, staff has indicated that policy programs would have been formulated differently if more accurate information had been available (Reichmann and Monasterski, 2016) (Box 4). From 2000 through March 2015, there were 62 cases of misreporting vis-à-vis data in the context of Fund-supported programs,³⁹ up sharply from the nine cases in the previous 15-year period from 1985 to 2000. Occasionally, inaccurate or incomplete information about a member country’s observance of a program performance criterion may give rise to a “noncomplying purchase” and the issuance of a misreporting notification to the Executive Board.

41. Even when data allow for adequate diagnosis and formulation of policies, the specific design of performance criteria is influenced by considerations of data accuracy, availability, and timeliness. Trade-offs are unavoidable among these factors, and the resulting criteria will seldom be totally homogeneous across time

³⁷These datasets are part of the Data Gaps Initiative.

³⁸The improvement in the collection of FSIs is especially noteworthy, with 101 countries currently providing at least the core indicators as of mid-2015, compared with 57 in 2007. Nonetheless, FSIs notably suffer from lack of comparability across countries, as they are based on very heterogeneous definitions of capital, nonperforming loans, etc.

³⁹Of these 62 cases of provision of incorrect data, 11 were considered “de minimis,” 38 received waivers, and only 13 required corrective actions, usually involving early repurchase or repayment.

Box 4. Faulty Data and Faulty Analysis: Past Examples

Instances of data that subsequently prove to be wrong or incomplete are probably frequent, but are usually of little consequence and therefore go unreported. However, a number of cases of data-induced faulty analysis were documented in reports on breaches of obligations under Article VIII, Section 5 or misreporting in programs (Reichmann and Monasterski, 2016). The following are examples of the type of cases that can occur:

Hungary (1982–89)

In November 1989, the government revealed that both domestic and external debt had been underreported since the mid-1970s. The misreporting involved a misspecification of the net credit to the government and the consequent misreporting of monetary and balance of payments statistics as well as the public debt. In the February 2000 review of misreporting cases (IMF, 2000), staff stated that “. . . Hungary’s widespread, systematic and substantial misreporting of data clearly resulted in a fundamentally distorted view of the program by the staff. . . . Had the staff been aware of actual [developments] the program would not have been submitted for Board approval with the same quantified criteria. Had correct data been known, it would have at least affected the staff’s assessment of the size of corrective actions needed. . . .”

Jordan (1996–97)

Under an extended Fund arrangement, the authorities provided staff with erroneous information on national accounts and fiscal data. Revisions provided in mid-1998 indicated that GDP growth had been substantially lower

than first reported—around 1 percent per year instead of 5 percent—and, consequently, fiscal revenues had been substantially lower than reported. As a result, Jordan’s budget deficit had been higher and had to be financed by recourse to nonbank sources. In the 1999 Article IV consultation report, staff indicated that “. . . the data set that [had been] available had portrayed a fundamentally distorted picture of the state of the Jordanian economy and performance under the extended arrangement . . .” and “. . . staff [had been] working on the basis of a wrong view of economic developments in Jordan, which had a major impact on the assessment of performance. . . .”

Ukraine (1996–98)

Ukraine misreported the level of its international reserves continuously during 1996–98 and in the negotiations on a follow-up extended arrangement. The misreporting involved multiple transactions that impaired the liquidity of the foreign assets involved and, more egregiously, two “round-tripping operations” which artificially inflated the reserves. Days before the Board meeting on the requested arrangement, it was revealed that almost \$700 million of reserves was illiquid, leaving usable reserves of less than \$300 million—forcing an impromptu redesign of the program. The corresponding staff paper stated “. . . With the new information on Ukraine’s external reserve position, and the pressure in the market, the authorities have had little choice but to move the exchange rate band . . . the staff has reluctantly accepted the reimposition of the export surrender requirements . . . [and] further modifications of the program might be unavoidable. . . .”

or countries. Usually, the wider the coverage of a performance criterion, the better it reflects policy aspects that have a bearing on the program’s objectives. But wider coverage may run afoul of the availability and timeliness of the required data, forcing an inevitable narrowing of the criterion’s scope.⁴⁰ Over 60 percent of staff acknowledged the influence of data conditions in the formulation of performance criteria.⁴¹

⁴⁰This narrowing of the scope can have a critical impact on policy implications. For example, based on interviews with the relevant country authorities, the Fund missed about 25 percent of GDP in public debt, in a recent financial program, by failing to include data on public-private partnerships and state-owned enterprises.

⁴¹The same percentage of staff noted that the program included undertakings to improve data provision or quality.

... and performance criteria must often be tailored to fit the availability of data ...

42. Even the variables that data are intended to measure may differ across countries, reflecting the particular historical and political developments that determine a country’s institutional organization and hence the definition and scope of a given economic sector or instrument. This is particularly the case in regard to the concept of government or the public sector (Annex 6). The resulting differences in definition mean that a balance must be struck between the Fund’s need to treat members evenhandedly in the application of conditionality and its need to tailor performance criteria to fit the circumstances of each case.

... but an IMF-supported program can also help improve data quality and availability.

43. Often the existence of a program can have a reciprocal effect on the quality, timeliness, and availability of data. The due diligence that staff is required to do before including data in a performance criterion can result in the correction of data that are found wanting or in efforts to develop and provide the data needed. Such positive effects, plus the intersectoral consistency checks provided by the financial programming framework, have been felt more by developing and emerging economies than advanced economies,⁴² as the former have been more frequent users of Fund resources.

B. Addressing Information Gaps

The IMF's efforts to address data gaps have resulted in a significant expansion in data ...

44. By and large, the collaborative arrangements in place for data provision have served the Fund well, with most member countries providing data that far exceed those required under Article VIII. Even more so, since the global crisis, there has been a notable rise in the amount and breadth of data (much of which is in the financial realm) that member countries provide to the Fund. For example, 138 economies currently report monetary and financial statistics according to the IMF's standardized report form (SRF),⁴³ up from 83 as of end-2007.⁴⁴

45. Much of this strengthening of data provision is due to concerted efforts—on the part of the IMF (especially STA), other members of the IAG, and member countries—to address data gaps identified by the global crisis.⁴⁵ In particular, significant progress has been made in implementing the recommendations of the G20 Data Gaps Initiative (DGI); all G20 members and many

⁴²In fact, until the global economic and financial crisis with its origin in advanced countries, many desks on such countries did not use the financial programming or other macroeconomic framework to check for intersectoral data consistency. This became particularly evident when some member countries of the European Union (EU) came to the Fund for financial programs in the aftermath of the crisis.

⁴³Among STA's many databases, the SRF data are the most used by area department staff.

⁴⁴Notwithstanding this impressive progress, several G20 countries and other economies with systemically important financial centers still do not report with the SRF.

⁴⁵The Managing Director's Global Policy Agenda (IMF, 2015d) noted that closing data gaps should be a key area targeted by the Fund's capacity development activities.

non-G20 economies have enhanced their data provision to the IMF (IMF, 2014e), and efforts to collect a broader array of financial data (including FSIs) are also proceeding apace. Most of the associated conceptual work for the DGI has been completed, and more generally, the number and types of data-based analytical tools have expanded significantly.

... but do the benefits outweigh the costs?

46. After a crisis, data suddenly become a forethought, rather than an afterthought. This raises the question: were data gaps a core reason or a scapegoat for missing the recent global economic and financial crisis? The answer to this question is an important one, as it can help determine the direction for future surveillance. In fact, the failure to foresee the impending crisis cannot be attributed to lack of data (Box 5). With hindsight, it became clear that a substantial amount of existing data had pointed to growing vulnerabilities in several key areas.⁴⁶ Failure to foresee the crisis stemmed more from ignoring or misinterpreting these warning signals than from the absence of signals, a view shared by many of this evaluation's interviewees.

47. Nevertheless, filling in key data gaps could substantially strengthen surveillance. But this also comes with costs, particularly for those responsible for collecting or providing the new data. Almost three-quarters of the respondents to our survey of data providers believe that the benefits of the Fund's new data initiatives outweigh their costs, yet 40 percent felt the IMF was asking for too many data and almost half said the initiatives would pose a very heavy burden on reporters. This was particularly the case with the respondents from advanced countries, who are most affected by the new demands under the DGI. Among the respondents from low-income countries, only one in five indicated concern in this regard.

48. The proliferation of data and analytical tools also risks the possibility of the Fund failing to strike the right balance between collecting information and being able to process it efficiently and analyze the results. Indeed, while two-thirds of staff respondents to the survey indicated that the additional data from the new

⁴⁶IEO (2011a) notes, for example, that had the IMF conducted the Vulnerability Exercise for Advanced Countries prior to the crisis, using data that were available in 2006 would have pointed to the United States, United Kingdom, and Iceland as being at high risk of financial crisis.