

Documentation on

Direction of Trade Statistics (DOTS)

Database and Browser on CD-ROM

Table of Contents:

- A. Introduction
- B. Contents of the Direction of Trade Statistics Database
- C. Estimation Methodology
- D. Consistency Between Partner Country Data
- E. Metadata in Time Series
- F. Two Formats of the DOTS Database--Scientific Notation and MS Access: Differences between their Time Series Record Structures

A. INTRODUCTION

The Direction of Trade Statistics (DOTS) Database and Browser on CD-ROM, a product of the International Monetary Fund (the Fund or IMF), contains exports and imports data by partner countries and areas of Fund member countries and nonmembers such as Cuba and North Korea. The DOTS CD-ROM contains all trade series for all countries that appear in the *Direction of Trade Statistics Yearbook (DOTSY)*.

The product is designed to meet the needs of two sets of users – those who prefer to access the database through an interactive interface and those who prefer to access or process the full contents of the database directly .

For those who prefer to access the database interactively, the browser software provides a familiar and easy-to-use Windows interface for browsing the database, selecting series of interest, displaying the selected series in a spreadsheet format, and saving the selected series for transfer to other software systems, such as Microsoft Excel. It includes four complementary views for browsing the database: (1) a table view corresponding to the tables contained within the *DOTSY* publication; (2) an economic concept view that provides access to similar concepts across countries; (3) a cross country matrix view that allows retrieval of a single concept and period for selected countries and partners; and (4) a view/search facility based on the structure of the time series codes.

For users who access or process the full contents of the database directly, the DOTS database text or “flat” file can be extracted from the CD-ROM in scientific notation format. It is approximately 300 MB in size. Users should employ the DUMP utility that installs with the browser to generate the flat file. After the browser is installed, click on Start/Programs/IMF CD-ROM/Dump Utility. The Dump Utility screen accepts entries for the database file (E:\data\dot.mdb- where E: is your CD-ROM drive) and output file

(your choice). The program begins after the database and output files are selected via the user interface buttons. The processing time to create the expanded data file varies according to the type of computer being used. For example, a Pentium III with 128 MB of RAM will execute the DUMP program in slightly less than an hour. The data items are stored internally in a compressed form, which maintains six significant digits of accuracy for every observation and therefore allows for a greater degree of precision than in the printed issue.

For assistance in using the browser software, an extensive on-line help facility is incorporated with the software. In addition, a list of frequently-asked questions (FAQ) is available in Print_Me.

For assistance in using the flat file, the remaining sections of this document provide information on the scientific notation database format. Section B of this document is a note on the coverage of DOTS data. Sections C and D of this document provide background information on the hierarchy of estimation types and on the inconsistency in reported data between countries. Sections E and F provide technical information aimed primarily for users of the database files (i.e. the scientific notation file format).

B. Contents of the DOTS Database

The Direction of Trade Statistics (DOTS) database presents figures on the value of merchandise exports and imports by trade partners as well as area and world aggregates showing trade flows between major areas of the world. The DOTS database on CD-ROM provides users with more than approximately 100,000 time series of annual, quarterly, and monthly trade data for approximately 186 countries. These data are also published in the *Direction of Trade Statistics Yearbook (DOTSY)*. Data on the CD-ROM are available from 1980. Beginning in 1981, nonreported monthly data are estimated following a defined estimation procedure. A detailed explanation of this procedure is found in *A Guide to the Direction of Trade Statistics (1993)*, which is incorporated in the CD-ROM.

“Country” does not always refer to a territorial entity that is a state as understood by international law and practice; the term also covers some nonsovereign territorial entities, for which statistical data are provided internationally on a separate basis.

The country composition of area aggregates distinguishes three main categories: industrial countries, developing countries, and a group of countries summarized as other countries n.i.e. (not included elsewhere). Developing countries are divided into five areas: Africa, Asia, Europe, the Middle East, and the Western Hemisphere. Other countries n.i.e. comprise Cuba, North Korea, and prior to 1990, Eastern Germany. Also included are memorandum items for the European Union (EU), Oil Exporting Countries, and Non-Oil Developing Countries. Data in these memorandum lines show trade by all members of the group with the respective country. Similarly these memorandum lines appear on the area and world tables. The memorandum line represents trade by all individual members of the group. For example, the EU line in the

world table represents trade by all EU members with the world, including trade among members. It does not present the EU data as if the EU was a single country. Users interested in such statistics would need to subtract intra-EU trade.

Trade included under “special categories” may cover military trade, transactions in free zones, or transactions in aircraft or ships. These data have not been designated to a partner country by the trade statistics compilers, nor have they been included by item in the reported data as an area not specified.

At the bottom of each country table two sets of analytical lines appear, based on the data in the table. The first set shows the distribution (in percent) of the country’s total exports and imports by area of the world. This distribution may not add to 100 percent, because trade not specified by partner country or area and entries shown as “special categories” are not included. The other set of lines shows annual percentage changes for total imports and exports by area of the world.

All figures are presented in U.S. dollars. Most countries report data in their national currency. U.S. dollar equivalents are obtained by converting the data at exchange rates published in series **rf** or **rh** in the country pages in *IFS*. The area and world totals are obtained by summing the individual country data. Export data for individual countries are shown f.o.b., whereas import data usually are shown c.i.f. In some cases import data are reported and shown f.o.b. However, for the calculation of area totals and for the matrix view, import data reported f.o.b. are adjusted to a c.i.f. basis by applying a c.i.f./f.o.b. factor of 1.1.

Data that countries report to the Fund on exports and imports by trading partners vary in frequency and currentness. The 23 industrial countries and about 40 developing countries report their data by month on a regular basis. In recent years, these countries have represented more than 85 percent of the value of recorded world exports and imports. Other countries report monthly data that are less current, or they compile and provide only quarterly or annual data. In general the data for individual countries are reported by official national agencies to the Fund or through their official publications. For some countries, the data reported to the United Nations Statistical Division and maintained in the *COMTRADE* database also have been used. For most member countries of the European Union, monthly data are sourced from the *COMEXT* database maintained by EUROSTAT.

C. Estimation Methodology

Partner country information makes it possible to estimate trade not only for countries less current in their reporting but also for countries for which data are not available from other sources. The estimation procedures were modified and enhanced beginning with data that appeared in the 1989 yearbook. The estimation procedures are described in *A Guide to Direction of Trade Statistics* (1993), available from the Fund’s Publications Section and also incorporated onto this CD-ROM. Reported data, including total imports and exports reported for publication in *IFS*, are the basis of all estimates in *DOTSY*. The entire

direction of trade statistics (DOTS) database is continuously supplemented with estimates. Area and world totals are compiled from reported data and these estimates. For a given year the percentage of world trade that is estimated declines over time, as data reported by countries replace the estimates.

When partner data are used directly or indirectly to derive estimates, the data first are adjusted by a c.i.f./f.o.b. factor of 1.10 to allow for the cost of freight and insurance. Reported imports c.i.f. are divided by the c.i.f./f.o.b. factor to give partner country estimates of exports f.o.b. Similarly, reported exports f.o.b. are multiplied by the c.i.f./f.o.b. factor to give partner country imports c.i.f. For example, if country B has not reported data from its own records but country A has done so, then A's data for imports from B (reported c.i.f.) are divided by 1.10 to give the f.o.b. value of B's exports to A. Conversely, A's data for exports to B (reported f.o.b.) are multiplied by 1.10 to give the c.i.f. value of B's imports from A. Given the absence of timely data on cost, insurance, and freight, the 10 percent c.i.f./f.o.b. factor represents a simplified estimate of these costs, which vary widely across countries and transactions.

Briefly, the estimation procedures described in *A Guide to Direction of Trade Statistics* (1993) are as follows. In the absence of some or all of the monthly DOTS, quarterly or annual reported DOTS are used. If quarterly data are available, then they are distributed over the relevant months using (a) the available monthly DOTS reported for other partner countries, (b) *IFS* monthly totals, or (c) an even distribution. If only annual data are available, then estimation procedures include, in addition to the techniques described above, DOTS reported from the most recent annual report or extrapolations of the most recent data or estimates. Where possible, estimates are computed first for months, and then annual totals are obtained by summation.

If data on total trade are reported for publication in *IFS* but DOTS are not reported, then the sum of the direction of trade estimates are constrained to coincide with the *IFS* totals. In such cases, the data for total exports and total imports shown in the lines *IFS* World Total and *DOTS* World Total will be identical. If *IFS* data are not available, then extrapolation will determine the estimated value of trade, and *IFS* data will not appear on the country page.

When information is not reported and is inadequate to support the estimation techniques, the data are extrapolated using a matrix of trade among broad country groups. Projected trade growth by these country groups, consistent with trade growth estimates used in the Fund's *World Economic Outlook*, is combined with the available DOTS to derive extrapolation factors.

Estimates are not provided for trade flows between countries where data are unavailable for both trading partners during the latest ten years or more.

In summation, only a small portion of world trade is omitted from the DOTS; this portion comprises (1) a small amount of unreported trade among the developing countries; (2) a small amount of unreported trade between developing countries and the countries

comprising the group “other countries n.i.e.”; and (3) the trade among “other countries n.i.e.”

Comment boxes are shown for some data observations to identify data that have been derived from partner countries or have been estimated.

D. Consistency between Partner Country Data

It is sometimes assumed that corresponding export and import data between partner countries should be consistent. That is, the exports from Country A to B should be equal to the imports of Country B from A, after taking into account the insurance and freight costs under the generally observed case that Country B imports are valued on a c.i.f. basis. The DOTS estimation system uses this assumption in cases where one partner has not reported data.

However, notwithstanding the inclusion of insurance and freight in imports c.i.f., it should be noted that there are several complications that can cause inconsistency between exports to a partner and the partner’s recorded imports f.o.b., or between imports f.o.b. from a partner and the partner’s recorded exports. The principal reasons for inconsistent statistics on destination and origin for a given shipment are differences in classification concepts and detail, time of recording, valuation, and coverage, as well as processing errors:

(a) Classification concepts and detail:

Destination/origin classification concepts in the presence of transshipment: Much trade is consigned via other countries, e.g., a substantial volume of German trade passes through the Netherlands. An inconsistency in reported trade data can arise when the trade of the originating country is classified as going to either a first or proximate destination that is a transshipment point, whereas the ultimate destination country classifies the same trade according to its ultimate source. Alternatively, the originating country may classify the shipment as going to the ultimate destination country, whereas the ultimate destination country classifies the same trade according to a proximate source that is a transshipment point. The lack of uniformity in the classification approach actually results from the application of different national customs instructions used in determining origin, transshipment, and destination countries.

Although there would be no classification inconsistency if all countries were to use a proximate destination/origin approach, this does not meet the analytical and operational needs of many data users whose interests align with the ultimate destination and/or origin of trade. That is, analytical information following the ultimate destination/origin principle requires countries to further distinguish, by trading partner, between the exports and imports involved in transshipment through their territory (re-imports and re-exports) from exports whose ultimate origin, and imports whose ultimate destination, is in their territory. This, however, is often a difficult distinction to make in actual situations, and publication of such data is not uniform international practice. International

recommendations note that no single way of attributing partner countries is ideal in view of the prevailing conditions under which trade is recorded.

Lack of destination/source country detail in published statistics: These conditions cause immediate problems in identifying trade counterparty information from published data. Destination/origin country for military and other confidential items and government goods may not be published or may be published only by regional aggregates. A destination can also be indeterminate and thus not be published, because the transshipment dispatcher may lack knowledge about the ultimate destination at the time of consignment. Regional aggregates without country detail may also be published when the trade of individual partners in a region is below a threshold level.

(b) Time of recording:

There is a lag between shipment of an export and receipt of an import, especially for goods that are shipped over long distances. To avoid timing inconsistencies, international national accounts and balance of payments accounting standards recommend a change of ownership principle to assign a date to transactions. Under this principle, an exported good changes ownership at shipment from the exporting country, at which time the exporting country records an export and the destination country records an import (f.o.b.). However, for practical reasons, the customs authorities of the importing country record the receipt of these goods later, when they arrive at the receiving country's frontier. This can cause a timing inconsistency. The result is that the same shipment can appear in an importing country's statistics for a month, quarter, and/or year that is later than the period that appears in the exporting country's statistics.

(c) Valuation:

Possible reasons for inconsistent valuation of a given shipment between source and destination include inconsistent currency version, evasion, anti-evasion procedures, values not known at the time of consignment, and differences in treatment of particular costs or procedures for assessment.

(d) Coverage:

Shipments to and from free-trade zones and bonded warehouses, exclusion of military and other confidential items and government goods, value thresholds for customs registration of shipments, returned goods, and other goods missed by customs (or surveys) are examples of coverage differences that can result in inconsistencies.

(e) Processing errors:

As a result of reporting and processing lags, trade data for a given period are often released before all customs documents for the period have been processed. These data are sometimes not revised, or, if data are revised, errors are nevertheless made in assigning the date on which goods are shipped or received and the late data are assigned to the

wrong month, quarter, and/or year. Errors can also be made in assigning a destination to exports and an origin to imports during customs clearances, or in cases when the ultimate destination is changed after the initial consignment during transshipment, the change is not incorporated into published statistics via the release of revised data.

These issues should be considered in interpreting the data. International cooperation in harmonizing and reconciling customs procedures and definitions may assist in reducing differences, and, in a few cases, countries have reconciled data or have used partner data to make their own estimates.

A particular problem arises in DOTS for countries that have large inconsistencies with partners and sometimes fail to report data. In such cases, there will be anomalies in the time series between those periods based on reported data and those based on partner data.

E. Metadata in Time Series

Each time series on the CD-ROM is composed of a descriptor record (containing metadata--control codes, operation codes, and descriptive information--about the time series) and numerous data records (also containing a number of metadata items that describe some aspect of the series).

Metadata items described in this section comprise:

Descriptor record:

- 1) time series code, consisting of country/area code, topic/subject code (section, subsection, classification, and qualification), version code, publication code, and partner country code
- 2) descriptor-record indicator
- 3) data type
- 4) magnitude (dimension/unit)code
- 5) decimal code
- 6) English descriptor
- 7) country/area name
- 8) magnitude name
- 9) base year
- 10) data-provider code
- 11) scale factor
- 12) data-type code
- 13) year-end period
- 14) consolidation code
- 15) last-update date
- 16) subject matter
- 17) calculation sequence
- 18) national descriptor
- 19) responsibility code

Data record:

- 1) starting-year identifier
- 2) data frequency code
- 3) break/footnote indicator

E.1 Descriptor Record Metadata**E.1.1. The Basic 13-Character Time Series Code**

Time series in the DOTS database on CD-ROM are uniquely identified by a thirteen-character time series code. A description of the time series code is given below. Briefly, the time series code, CCCTTTTTVSPPP, is constructed as follows:

- CCC - Three-character country/area code (always numeric)
- TTTTT - Five-character topic code (usually 2 numerics followed by 3 alphabets)
- V - Single-character version code (always the alphabet Z since there is only one version in DOTS)
- S - Single-character publication code (always alphabet D)
- PPP - Three-character partner country code (always numeric)

(As a matter of convention, a period [i.e., the punctuation mark "."] is always used on the CD-ROM as a position holder to indicate a blank in the time series code [e.g., "60571..dzd718"].)

E.1.1.1. Country/Area Code in Time Series Code

The country code identifies the country or area to which the time series data relate. "Country" in this publication does not always refer to a territorial entity that is a state as understood by international law and practice; the term also covers the euro area and some territorial (nonsovereign) entities, for which statistical data are provided internationally on a separate basis.

The country code always consists of three numeric digits. The first digit typically provides regional grouping as follows:

- 0 World or all country aggregates
- 1 Industrial countries
- 2,3 Developing Countries, Western Hemisphere, and Non-oil Developing countries
- 4 Middle East
- 5,8 Asia
- 6,7 Africa
- 9 Centrally managed or formerly centrally managed and other economies, Oil-exporting countries, and the European Union

A complete list of the country codes/currency names is provided in a separate file entitled DOT.cty in Print_Me, in both numerical order and alphabetical order. Data pertaining to an area provide aggregated data for countries that comprise that area.

E.1.1.2. Topic/Subject Code in Time Series Code

The topic/subject code is a five-position alphanumeric code indicating the topic or subject for each time series, consisting of section, subsection, classification, and qualification codes. Of the five positions, the first two characters are always numeric; the next three characters are always alphabetic. On the CD-ROM, the "." is always shown when a position would otherwise be blank. No topic/subject codes are printed in the DOTS hard-copy publication; however, the internal (CD-ROM) codes are 70..D for exports, 71..D for imports c.i.f., 71.VD for imports f.o.b., 70P and 71P for percent distribution, and 70C and 71C for annual percent change.

Section Code in Topic Code

The first position of the topic code, referred to as the section code, identifies the institutional, topical, or subject area of this series. The section code used in DOTS is:

7 international transactions

Subsection Code in Topic Code

The second position of the topic code is known as the subsection code. The subsection codes used in DOTS are:

0 exports
1 imports

Classification Code in Topic Code

The third and fourth positions of the topic code, when present, provide classification distinctions. The classification codes used in DOTS are:

P. percent distribution
C. annual percent change
.V imports f.o.b.

Qualification Code in Topic Code

The fifth position of the topic code, referred to as the qualification code, when specified, provides information about special characteristics of the data. The qualification code used in DOTS is:

D U.S. dollars

E.1.1.3. Version Code in Time Series Code

A one-position alphabetic code in position 9 of the time series code, used to identify the various revisions of concept and content of a time series. For the DOTS CD-ROM, there are no multiple versions; the version code is always Z.

E.1.1.4. Publication Code in Time Series Code

An alphabetic code in position 10 of the time series code, used to identify the IMF publication containing the series. All series in DOTS contain publication code D.

E.1.1.5. Partner Country Code in Time Series Code

A three-position numeric code in positions 11-13, used to designate the partner country in those time series that specify transactions between countries.

Additional Metadata Items in the Descriptor Record:

E.1.2. Descriptor Record Indicator

A one-position alphabetic code "H" in position 14, distinguishing the descriptor record from the data record.

E.1.3. Data Type

A one-position alphanumeric code, which represents three different fields in a one-position value. The three fields have been added to the CD-ROM descriptor record as data type (stock, average, or flow), year-end code (fiscal year-end), and consolidation code, described later in this section.

This code also defines the way the months relate to the quarters and the way the months and quarters relate to the years. That is, this code indicates whether lower-frequency data can be or are derived from higher-frequency data. It also identifies the end of the fiscal year of the time series data. The following table provides information on this code that refers to the DOTS database ("conditionally" means derivation is performed only if the desired observation does not already exist). DOTS data are flow data and refer to December end-of-year.

Type Code	of Data	End of Fiscal Year	Quarters Derived from Months	Annual Derived from Quarters
F	Flow	December	Always	Always
N	Flow	December	Conditionally	Never
P	Flow	December	Always	Never
R	Flow	December	Never	Always

5 Flow December Conditionally Conditionally

E.1.4. Magnitude (Dimension/Unit) Code

A one-position alphanumeric code that indicates the unit of measurement (weights, volumes, currencies, rates, etc.) and the scale (units, thousands, millions, etc.) of a series. This code also specifies a multiplier to be applied to observations when processing revisions to the series or displaying observations from the series.

The valid codes in DOTS are as follows:

- 1 - Millions of U.S. dollars
- 3 - Billions of U.S. dollars
- 5 - Percent

E.1.5. Decimal Code

The decimal code in position 18 of the descriptor record is not used to scale the data. This field normally indicates the number of decimal places for presentation purposes.

E.1.6. English Descriptor

A 36-position field that describes the subject of the time series.

E.1.7. Country/Area Name

A 24-position field that identifies the reporting country to which the data pertain.

E.1.8. Magnitude (dimension/unit) name

A 16-position field that states the magnitude of the time series.

E.1.9. Base Year

Not applicable to DOTS.

E.1.10. Data-Provider Code

This is not used in DOTS

E.1.11. Scale Factor

The scale factor is not being applied to control the data presentation. This two-position field normally indicates the scale factor that could be used for the presentation of data in decimal format.

E.1.12. Data-Type Code

A one-position alphanumeric code that identifies the data in the time series as stock, flow, average, or unknown. Valid values are S, F, A, or U. DOTS data are flow data.

E.1.13. Year-End Code

A three-position alphabetic year-end code identifies the end of the fiscal year of the time series. DOTS time series refer to DEC year-end.

E.1.14. Consolidation Code

A one-position numeric code that indicates whether and under what conditions lower-frequency data are to be derived from higher-frequency data. It defines the way monthly observations relate to quarterly ones, and monthly and quarterly observations relate to yearly ones.

Consolidation Codes:

<u>Code</u>	<u>Descriptor</u>
0	Never consolidate
1	Always calculate longer-period data
2	Calculate longer periods only if they are void
3	Always calculate quarterly, but never calculate annual
4	Calculate quarterly only if void, never annual
5	Exclusively used for DOTS data and follows the rules outlined below:

<u>Condition</u>	<u>Rule</u>
<u>Quarterly Consolidation</u>	
If observation exists and has no footnote(i.e., data are reported).	Do not consolidate.
If observation exists and has a footnote(i.e., is consolidated from estimated monthly data) or if no observation exists, consolidate and add footnotes as follows:	Three months have no footnotes(i.e., reported) = no footnote. Two months are reported and one month has a footnote (i.e., is estimated), footnote = G. One month is reported and two months estimated, footnote = L. If all three months are estimated, quarterly footnote = Y.
<u>Annual Consolidation</u>	
If observation exists and has no footnote (i.e., data are reported).	Do not consolidate.

<p>If observation exists and has a footnote (i.e., is consolidated from estimated monthly data) or if no observation exists, consolidate and add footnotes as follows:</p>	<p>12 months are reported = no footnote.</p> <p>6-11 months are reported and the rest are estimated, footnote = S.</p> <p>1-5 months are reported and the rest are estimated, annual footnote = T.</p> <p>If all months are estimated, footnote = Y.</p>
--	--

E.1.15. Last-Update Date

An eight-position field indicating the date on which the time series was last updated in the form YYYYMMDD, where YYYY is the year, MM is the month, and DD is the day of this date.

E.1.16. Subject Matter

A two-position alphabetic code that classifies the time series as to its general topic.

E.1.17. Calculation Sequence

A one-position alphabetic code that indicates in which hierarchical sequence a time series has been calculated.

E.1.18. National Descriptor

A 36-position alphanumeric field that describes the national description of a series.

E.1.19. Responsibility Code

The responsibility code is a data protection feature designed to prevent unauthorized access to or updating of a time series. The six-position alphabetic code identifies the custodian responsible for the maintenance of the time series.

E.2. Data Record Metadata

E.2.1. Starting-Year Identifier

A four-position numeric field that specifies the starting date for data in that record.

E.2.2. Data Frequency

A one-position alphabetic field that identifies data frequency (A for annual, Q for quarterly, M for monthly).

E.2.3. Break/Footnote Indicator

Browser:

The CD-ROM browser contains comment boxes to indicate that the observed data has been estimated. The following comments may appear in any of the estimated data cells:

Represents distribution of quarterly or annual DOT data using monthly pattern displayed by monthly reported partner country data covering the entire quarter/year.

Represents distribution of quarterly or annual DOT data using monthly pattern of IFS data covering the entire quarter/year.

Represents evenly distributed data using quarterly or annual DOT figures.

Represents distribution of IFS data by applying partner country pattern of the most recently reported annual DOT data.

Identifies data derived from reported partner country records.

Identifies data derived from estimated partner country records.

Represents extrapolated data.

Represents 2 months of reported data and 1 month of estimates.

Represents 1 month of reported data and 2 months of estimates.

Estimated by other methods, sometimes including the use of partner country records.

Represents 6-11 months of reported data and 1-6 months of estimates.

Represents 1-5 months of reported data and 7-11 months of estimates.

Identifies data consolidated from estimated monthly/quarterly partner country records.

Flat file:

Users of the flat file are provided with a one-position alphabetic character used to identify a break/footnote. The absence of an alphabetic break/footnote indicator in a series indicates that the figures on trade with individual trade partners are obtained from a country's own records. Figures representing summations of individual data or estimates to area and world totals and aggregations of countries to areas and the world also carry symbols. The composition of the total or aggregate in terms of reported data versus estimated figures can be established by reviewing symbols attached to the figures at the

detailed level. These indicators are automatically generated. A list of DOTS symbols follows:

Monthly data

- C --- Identifies distributed cumulative data
- A --- Represents distribution of quarterly or annual reported data using monthly pattern displayed by monthly reported partner country data covering the entire quarter/year.
- B --- Represents distribution of quarterly or annual reported data using monthly pattern of IFS data covering the entire quarter/year.
- D --- Represents evenly distributed data using quarterly or annual reported figures.
- I --- Represents distribution of IFS data by applying partner country pattern of the most recently reported annual DOTS data.
- P --- Identifies data derived from reported partner country records.
- Q --- Identifies data derived from estimated partner country records.
- R --- Identifies data derived from estimated partner country records.
- X --- Represents extrapolated data.

Quarterly data

- G --- Identifies consolidated quarter with 2 months of reported data and 1 month of estimates
- L --- Identifies consolidated quarter with 1 month of reported data and 2 months of estimates
- V --- Identifies consolidated quarter with figures estimated by applying partner data for the entire quarter
- Y --- Identifies consolidated quarter with figures estimated by other methods than by applying partner data for the entire quarter; also used in world and area totals

Annual data

- S --- Represents 6 - 11 months of reported data and 1 - 6 months of estimates
- T --- Represents 1 - 5 months of reported data and 7 - 11 months of estimates
- V --- Represents consolidated derived solely from partner records
- Y --- Represents consolidated data estimated by other methods than by applying partner data for the entire year; also used in world and area totals

**F. Two Formats of the DOTS Database--Scientific Notation and MS Access:
Differences between their Time Series Record Structures**

F.1. DOTS Database in Scientific Notation

The flat file containing observations in scientific notation consists of time series, which contain descriptor records and data records:

F.1.1. Descriptor Record Structure

As described above in section E.1, the first record of a time series is called the *descriptor record*, is up to 162 characters in length, and is identified by the letter "H" in position 14 of the record. This record contains control codes and descriptive information—metadata—about the time series.

The descriptor record structure in a time series is as follows:

<u>Position</u>	<u>Field Length</u>	<u>Class</u>	<u>Data Field Name</u>
1-3	3	Numeric	Country code or area code
4-8	5	Alphanumeric	Subject code
9	1	Alphanumeric	Version code
10	1	Alphanumeric	Publication code
11-13	3	Alphanumeric	Partner country code
14	1	Alphanumeric	Descriptor record indicator
15	1	Blank	
16	1	Alphanumeric	Data type code-packed
17	1	Alphanumeric	Magnitude(dimension/unit) code
18	1	Alphanumeric	Decimal code
19-54	36	Alphanumeric	English descriptor
55-78	24	Alphabetic	Country name
79-94	16	Alphanumeric	Magnitude(dimension/unit) name
95-98	4	Alphanumeric	Base year
99	1	Alphanumeric	Data provider code
100-101	2	Alphanumeric	Scale factor
102	1	Alphanumeric	Data type
103-105	3	Alphanumeric	Year-end code
106	1	Alphanumeric	Consolidation code
107-114	8	Alphanumeric	Last update date
115-116	2	Alphanumeric	Subject matter
117	1	Alphanumeric	Calculation sequence
119-154	36	Alphanumeric	National descriptor
155-160	6	Alphanumeric	Responsibility code
161-162	2	Blank	

F.1.2. Data Record Structure

Following the descriptor record of a time series are the associated *data records*. The annual records are sequenced first, followed by quarterly records, followed by monthly records. A time series can contain a variable number of data records. The exact number

of records in any time series is dependent upon the availability of the annual, quarterly, and monthly data.

The first four positions of the annual, quarterly, and monthly data records specify the starting year for data in that record. The fifth position identifies an annual data record by "A," a quarterly data record by "Q," and a monthly data record by "M." The data consist of 12 thirteen-position fields (twelve numeric positions and a preceding break/footnote indicator position). Monthly data records always begin with January in the first data field, and quarterly records with Quarter 1 in the first data field. For example, a data record with 1990M in positions 1 through 5 would contain the 12 monthly data fields of 1990. If January and February are unavailable, the first two data fields will be blank.

The data fields appearing within a time series are represented in scientific notation in units. For presentation purposes, these fields can be scaled by the dimension/unit or decimal code contained in positions 17 and 18 of the descriptor record. For example, a data field of 1.23208E+09 with a dimension/unit "1" would take on a value of 1232.08 million U.S. dollars while the same data field with dimension/unit of "4" would take on a value of 1.2 billion national currency (dimension/unit 1 = millions of U.S. dollars; 4 = billions of national currency). See subsections E.1.4 and E.1.5 for the description of magnitude(dimension/unit) and decimal codes.

The data record structure in a time series is as follows:

Position	Length	Class	Data Field Name
1-4	4	Numeric	Year identification
5	1	Alphabetic	Data frequency--A(Annual), Q(Quarterly), M(Monthly)
6-18	13	Alphanumeric	Break/footnote indicator and data1
19-31	13	"	" data2
32-44	13	"	" data3
45-57	13	"	" data4
58-70	13	"	" data5
71-83	13	"	" data6
84-96	13	"	" data7
97-109	13	"	" data8
110-122	13	"	" data9
123-135	13	"	" data10
136-148	13	"	" data11
149-161	13	"	" data12
162	1	Blank	

F.2. DOTS Database in MS Access Format

The DOT.mdb file is an MS Access database that is accessible with MS Access 97. It has been constructed by using DAO (Data Access Objects) 3.5. Although the MS Access database is available for user access, users should note that the International Monetary Fund reserves the right to alter the structure of the database without notice. For this reason, users who develop their own programs to access the data are advised to use the scientific notation format, rather than the DOT.mdb format.

The DOT.mdb file contains two tables: TimeSeries and Observations. The TimeSeries table contains the main attributes of time series as described in section C. The Observations table contains all time series observations and has the following structure:

OID: TimeSeries code

OstartYY: the year

Ovalue1 to Ovalue12: the monthly value (1=January,....., 12=December)

Ovalue13 to Ovalue16: the quarterly values (13=Q1,....., 16= Q4)

Ovalue17: the Yearly value

Flag1 to Flag12: the monthly footnote (1=January,....., 12=December)

Flag13 to Flag16: the quarterly footnote (13=Q1,....., 16=Q4)

Oflag17: the yearly footnote