

STATISTICS DIRECTORATE

National Accounts and Economic Statistics - International Trade Statistics

THE OECD ITCS DATABASE IN THE UN/OECD JOINT TRADE DATA COLLECTION AND PROCESSING SYSTEM

**6th OECD INTERNATIONAL TRADE STATISTICS EXPERT MEETING (ITS) & OECD-EUROSTAT MEETING OF EXPERTS IN TRADE-IN-SERVICES STATISTICS (TIS)
Tour Europe, Salle des Nations, 12-15 September 2005**

This document has been prepared by Mr. Gregory Legoff, OECD, for information and discussion under point b) of item 3 of the draft agenda : Recent developments at OECD

For further information, please contact :
gregory.legoff@oecd.org
andreas.lindner@oecd.org

JT00188440

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format

THE OECD ITCS DATABASE IN THE UN/OECD JOINT TRADE DATA COLLECTION AND PROCESSING SYSTEM: TRADE DATA ISSUES

Introduction

Some five years ago, the OECD and the United Nations Statistics Division joined forces in the area of merchandise trade data collection and data processing.

The overriding objective has been to reduce the respondents burden for OECD countries through OECD being the central collector of detailed merchandise trade data and, in doing so, synchronize and harmonize both UNSD and OECD data processing routines leading to identical data values published by both organizations. It goes without saying that each organization retains its “proprietary” way and media of data dissemination, including also possible different aggregations.

By its very nature and also novelty character in international co-operation, this has been a multi-year process in which OECD-UNSD co-operation has been gradually deepened and extended. Both organizations are now finalizing the complex project of a joint system to collect and process the data. In line with the action plan, it will be implemented and fully operative at the beginning of 2006.

Over the past three years, many joint meetings were held in both headquarters. A common data processing practice was established of which this paper presents the main elements.

The basic national data will be standardized at the most detailed commodity level available; these data can be shared with other international organizations. As detailed in the following points, records will be standardized with respect to country codes, commodity codes and quantity unit codes. Values in national currency will be converted into US dollars using the same exchange rate conversion factors. All statistical practices will be harmonized particularly concerning confidential trade treatment and conversions between classifications. As explained below, exchange rates will be updated automatically and data for both organizations will be revised and synchronized accordingly.

1. Country codes

UNSD and OECD currently use different country codes and labels to process and disseminate trade statistics.

On the one hand, UN Concepts and Definitions recommend that the statistical territory as described by the country itself should be used for the definition of the statistical territory of that country as a reporter and as a partner. On the other hand, the OECD needs to keep the coherence with the Eurostat nomenclature and the OECD Trade in Services database.

OECD and UNSD have agreed to maintain the most detailed level of country codes currently existing in each database and to maintain (or create when necessary) aggregated countries/territories to match the different needs of both organizations. All aggregated countries/territories having a numerical code will be processed like other partner countries in order to convert quantities properly. Each organization will have

the choice to publish the aggregated or the detailed countries/territories according to its own definitions. UNSD numerical codes will be used for data processing. The OECD will use ISO alpha 3-digit for dissemination.

2. Non-standard HS codes

The OECD has created special codes to process as accurately as possible special transactions or special commodities that could not be allocated to a standard code of the official nomenclature. Most of these codes come from the Eurostat Combined Nomenclature (C.N.). In accordance with UNSD, it has been decided that non standard codes would be retained in the common database as Memorandum Items (MIs).

The number of OECD non standard codes has been significantly reduced:

- From 586 to 223 for HS 2002;
- From 529 to 231 for HS 1996;
- From 526 to 231 for HS 1988;
- From 152 to 10 for SITC Rev.3;
- 6 for SITC Rev.2.

2.1 Selection process of MIs

1) In HS classifications

Each code of the Eurostat CN has a validity date. Starting with the HS 2002, when a special code was no more valid in 2002, it has not been retained as a MI. For simplification needs (especially for conversion needs between classification) and taking into account the use of the special codes within the OECD database, almost the same list of MIs has been retained for the different versions of the HS.

2) In SITC classifications

Since data are converted from the HS into the SITC classifications, special codes have not been retained as MIs in the SITC Rev.3 when the original HS code was suppressed.

Only three special codes referring to the Eurostat classification could have been maintained as MIs. These codes have not been retained because of their small number and also because of the difficulty to convert them from HS.

For harmonization with UNSD, some 5-digit codes have not been retained when they already existed at the 4-digit level.

In accordance with UNSD, all SITC Rev.2 special codes already existing in the OECD database have been retained as MIs.

2.2 Structure of MIs

1) In HS classifications

MIs concern three main categories of special transactions:

- Estimations of missing declarations of the Intrastat system;
- Components of complete industrial plants;
- Confidential trade.

Estimations of missing declarations have not been retained as MIs.

When the MIs refer to the CN, the structure is HSEUXX where:

- HS refers to the HS chapter where MIs should be allocated;
- EU refers to the Eurostat CN;
- XX refers to the previous OECD classification.

French and English labels of the CN have been used for MIs.

The MIs structure is HSCF00 for confidential trade. One confidential MI per HS chapter has been kept or created. The structure of MIs referring to petroleum products is 27PPXX.

2) In SITC classifications

All MIs in SITC 3 refer to confidential trade. The structure is the same as in HS: SCF00 where S refers to the SITC section.

In SITC 2, the 6 codes and labels currently used in the OECD database have been kept identical.

2.3 Impact of the new procedure

For HS classifications it has been decided in December 2003 in accordance with UNSD that wherever possible, data shown under MIs would be added in the appropriate HS 2-digit. Most of OECD special codes were previously included into 98 or 99 chapters. This change significantly affects HS chapters' values (e.g. 19.3% of 2001 Australian exports are confidential).

In order to avoid break in series in the OECD database, it has been decided that HS historical series back to 1988 would be re-processed in accordance with the new rule of allocation to an HS 2-digit. SITC Rev.3 series will also be re-processed for harmonization.

2.4 Conversion of special codes into MIs

Old OECD special codes will be allocated to MIs using specific correspondence tables for each classification. Only values are allocated to avoid quantity conversions. Consequently, quantities referring to old special codes will be lost. The rules of conversion are as follow:

- If there is a correspondence with a MI, the special code is allocated to it. If there is no simple correspondence with a MI, the special code is allocated to an aggregated level of the official classification. In order to simplify conversions calculations, allocations are given at the highest level of the classification (2-digit in HS and 1-digit in SITC Rev.3);

- Aggregated and lower levels of old special codes cannot be both allocated because of double counting. Preference is given to lower levels of the classification to keep as much detail as possible. But it also implies a loss of information when higher level values are substantially different than the sum of the lower levels;

- When a special code has only one component, the highest level is allocated. Since all SITC Rev.2 special codes have been maintained as MIs, there is no need of conversion table for this classification.

2.5 Special cases

Codes 999999, 9999 and 99 will be created as non standard HS codes but they will not be considered as MIs. They are the total value minus the sum of the values of all official codes.

2.6 Conversions of MIs to older classifications

1) Between HS classifications: MIs are identical from one classification to another one except for oil products which only represent 8 codes. MIs are converted using “one to one” conversion keys.

2) From HS into SITC classifications: conversion tables already exist. Four conversions referring to a new code or referring to codes which have changed of scope are created or reviewed.

2.7 UNSD-OECD harmonization

UNSD will use a simplified version of MIs. Data will be processed using the detailed OECD list of MIs. A correspondence table will be created to convert OECD MIs into UNSD MIs.

3. Reported data at higher commodity level

If reported data at a higher level of the commodity classification is substantially different in value than the sum of the aggregated lower level, then it is concluded that the reporting country added on purpose information to this level and this information should be processed. The value and quantity are added at the immediate higher level of the classification and the difference is footnoted

4. Conversion from one classification to another classification

Four cases have been envisaged:

- commodity code A correlates perfectly one-to-one with commodity code B of another classification, then code A will be converted to code B;

- commodity code A is fully included in the heading of commodity code B, then code A will be converted to code B;

- commodity code A is almost fully included in the heading of commodity code B, then code A will be converted to code B;

- commodity code A is only partially included in the heading of commodity code B, then code A will be converted to the immediate higher level of the classification and the difference is footnoted. OECD and UNSD have agreed to undertake a study to determine which correlations should be qualified as “almost fully included” and which as “only partially included”.

UNSD currently uses conversion tables between all classifications. With this conversion rules, values and quantities can be added at the HS 4-digit level between HS 2002 and HS 1988 without loss of information. The OECD will use UNSD’s conversion tables.

5. Quantity unit and estimations

The database will contain **two quantity fields**, one is net weight and the second one is quantity in terms of WCO recommended standard unit of quantity. These quantity unit codes are the following:

- 1 = not reported or not available
- 2 = Area in square meters
- 3 = Electricity in 1000 Kwh
- 4 = Length in meters
- 5 = Number of items
- 6 = Number of pairs
- 7 = Volume in liters
- 8 = Weight in kilograms
- 9 = Thousands of items
- 10 = Number of packages
- 11 = Dozens of items
- 12 = Volume in cubic meters
- 13 = Weight in carats

For estimation of quantities the following sequence of priorities has been adopted:

- If a direct conversion from the reported quantity to net weight or WCO unit is possible, then this will be done; quantity conversion factors recommended by FAO will be used as much as possible;

- If direct conversion is not an option then quantity (or net weight) will be estimated if at least 20% of the value has the desired quantity unit; a threshold of 50% will be applied for developing countries;

- If both options, (a) and (b), are not available then a standard unit value will be used to estimate the missing quantity; UNSD has improved its procedure for the calculation of standard unit values (see annex 2).

Estimation of quantities will be necessary in the process of aggregating from one level to the next; quantities will not be shown at the 2-digit level or higher. These rules should be used in succession and in relation to the aggregation from one level of the classification to the next. Quantity conversion factors will be reviewed and agreed upon with the understanding that the factors provided by the FAO will be taken to the maximum extent possible. Finally, it was decided that quantities will only be shown at the 4- and 6-digit levels of the HS and the 3-, 4- and 5-digit levels of the SITC.

6. Common strategy to process confidential trade data reported at the HS 2-digit level

Processing confidential data reported at the HS 2-digit level can create discrepancies between chapters and lower levels of the classification. Confidential trade included into the country code “secrets and differences” can be lower at the 2-digit level than at the 6-digit level. The difference comes from a

different partner country breakdown at the HS 2-digit level and at the detailed level when the reporter country includes less confidential trade at aggregated levels of the product classification. OECD and UNSD have agreed to include a general methodological note explaining these apparent discrepancies to trade statistics users.

The same discrepancies appear with commodities when the confidential trade breakdown is different at HS 2-digit level and at detailed levels of the classification. Consequently reported data referring to confidential chapter 99 can be lower than computed data from HS 6-digit sectors. Discrepancies between aggregated and detailed levels of the HS classification create additional inconsistencies when the data are converted into SITC classifications. The current practice of the OECD is to maintain the consistency at the HS 2-digit level by including all reported data and to ignore reported detailed data referring to confidential trade. This practice allows keeping a global data coherence at the aggregated level of the classification. Considering that confidential trade is already included in the difference between reported HS 2-digit data and computed data from lower levels of the classification, it is not reported again at the detailed levels.

7. Metadata and Metastore

MetaStore is a general toolkit for accessing and managing reference metadata for statistics. It has been designed to improve the efficiency of metadata preparation, storage, access, management and dissemination for statistical products across the OECD. Improvements in these areas will enhance both the transparency and accessibility of OECD statistics.

Trade metadata are currently located in a number of files according to different formats, the result being some duplication of effort in metadata preparation. The OECD is currently loading trade metadata into Metastore making available methodological information from the trade application client (see paper 3 on processing system issues). Once loaded into Metastore, metadata will also be available for paper and electronic publications as well as for the OECD Internet site update.

Metastore's purpose is the management of reference metadata documenting characteristics of the data within a dataset. For trade statistics it concerns trade definitions, recommendations, sources, country notes, information about classifications used, ... Metastore remains separated from the datasets' data production system and does not allow the management of structural metadata including codes and names of dimensions and corresponding dimension members. Therefore, control codes (e.g. estimated data) will not be managed by Metastore.

As a common repository, MetaStore manages reference metadata for datasets of different structures. Other international organizations like UNSD are studying the possibility to implement Metastore as a management tool for their metadata.

8. Exchange rates

OECD and UNSD use the same exchange rates initially sourced from IMF. Slight differences could exist between both organizations because of exchange rates revisions.

In fact, trade statistics usually refer to hundreds of billions of USD and a slight revision on exchange rates has a significant impact on converted figures. OECD and UNSD have agreed to check automatically the significance of exchange rates revisions on a quarterly basis and to perform automatic recalculations if necessary. The proposed dates for automatic checks are 1st January, 1st April, 1st July and 1st October. A track of historical exchange rates should be kept for reference. The threshold for the necessity of revisions will have to be defined.

UNSD and OECD trade databases will use synchronized and up to date exchange rates. Consequently, differences due to revisions of exchange rates will disappear.

9. OECD-UNSD databases synchronization

In order to provide one single dataset to trade statistics' users, both UNSD and OECD databases will be permanently synchronized. The synchronization process will be launched every night to update OECD and UNSD mirror databases. The data will be transferred in XML format using SDMX cross-sectional schema. Further details about technical aspects are given in the presentation "Processing System issues – OECD/ITN". Therefore, data differences between UNSD and OECD will completely disappear starting from 2005 data.

Slight differences will remain for historical series before 2005 data. A work of data reconciliation is ongoing particularly in the framework of the Common Data Set handled by WTO involving UNSD and OECD (see document 8: "The Inter Agency Common Data Set in Trade – WTO"). In 2004 the OECD has revised its US historical series back to 1991 to match UN concepts and definition. The OECD is currently revising its data for New Zealand in order to include re-exports into exports, in line with UN practices and recommendations.

The intensive and increasing cooperation between UNSD and OECD started several years ago has progressively permitted the harmonization of statistical and data processing practices reducing gradually differences of trade statistics between both organizations. This work is about to be completed with the joint trade data collection and processing system.

ANNEX 1
UNSD – OECD JOINT TRADE DATABASE
TECHNICAL NOTE ABOUT COUNTRY CODES

UNSD and OECD currently use different country codes and labels to process and disseminate trade statistics.

On the one hand, UN Concepts and Definitions recommend that the statistical territory as described by the country itself should be used for the definition of the statistical territory of that country as a reporter and as a partner. On the other hand, the OECD needs to keep the coherence with the Eurostat nomenclature and the OECD Trade in Services database.

1. General principles

OECD and UNSD have agreed to maintain the most detailed level of country codes currently existing in each database and to maintain (or create when necessary) aggregated countries/territories to match the different needs of both organisations. All aggregated countries/territories having a numerical code are processed like other partner countries in order to convert quantities properly. Each organization will have the choice to publish the aggregated or the detailed countries/territories according to its own definitions. UNSD numerical codes are used for data processing. The OECD uses ISO alpha 3-digit for dissemination.

2. Country codes list

The detailed country codes list “UNSD-OECD detail.xls” is available for both organizations in the Sharepoint tool. For each country/territory, the following fields are available:

- the UNSD numerical code which is the key used for data processing;
- the corresponding 3-digit ISO code. A 3-digit code has been created when no ISO code exists. Special territories are coded on 5-digit basis: ISO_A, ISO_B, ... when it only refers to a component of the ISO country and ISO_Z, ISO_Y, ... when it includes the main part of the ISO country.
- English label;
- French label;
- in UNSD: a cross indicates that the country/territory already existed in the UNSD database before 2005. This field has been left empty when there was no need to precise the difference between UNSD and OECD;
- in OECD: a cross indicates that the country/territory already existed in the UNSD database before 2005. This field has been left empty when there was no need to precise the difference between UNSD and OECD;
- historical: a cross indicates that the country/territory will not be used after 2005. Historical codes are only used to load historical series;

- world calculation: a cross indicates that the country/territory is included into the world aggregate calculation. Only detailed countries/territories are taken into account when they are also included into other entities;

- comments;
- starting year;
- ending year.

3. Different definition of statistical territories

In order to match the different definitions of both organizations, two or three codes are used for the 5 following countries: Switzerland, United States, France, Italy and Norway. For these countries, both components and aggregates have to be processed for quantity conversion needs. For example, the three codes Switzerland including Liechtenstein, Switzerland excluding Liechtenstein and Liechtenstein are processed like any other partner country code.

Aggregates are calculated by adding all components as stated in the UNSD-OECD detail.xls file. If a reporter country does not report any data with components, then a null value is added and the calculated aggregate remains correct. For example, if no data have been reported for Liechtenstein because they have already been included into Switzerland, then the calculated aggregate Switzerland including Liechtenstein is still correct. A methodological note should precise that depending on reporter countries some components of the countries mentioned above are not always reported separately.

4. ISO code changes

The OECD is changing its ISO codes for:

- Myanmar MMR instead of the outdated code BUR. The two codes MMR and BUR currently exist. Since no aggregation is carried out on country codes, BUR becomes an historical code and MMR will be used starting from 2005 data;

- Timor-Leste TLS instead of TMP;

- Romania RUM instead of ROM,

- Serbia and Montenegro SCG instead of XAA;

- the Occupied Palestinian Territory PSE instead of XAB and XAC. No aggregation is done for the Occupied Palestinian Territory which is currently broken down on two codes.

In order to upload historical series in the new SQL database, old codes will be used at a first stage. At a second stage, outdated codes will be replaced by the new ones. BUR, XAB and XAC will be kept as historical codes.

5. Areas not elsewhere specified

Trade with areas not elsewhere specified are broken down into 7 different zones. The previous area “America not elsewhere specified” has been split into “South America, not elsewhere specified” and “North America, the Caribbean and Central America not elsewhere specified”. UNSD and OECD databases code “America not elsewhere specified” is kept as an historical code. The area code “Other

areas, not elsewhere specified” concerns trade that cannot be allocated to another partner country/territory during the data processing.

6. Confidential information and differences

This code already existed in the OECD database. It contains confidential partner countries plus the difference between the world (when reported) and the sum of partner countries.

7. OECD Economic and geographic zones

Since grouping definitions are specific to each organization, differences will remain between UNSD and OECD. A grouping code cannot be used as an input code for processing; it is an aggregate of several countries/territories. As a second step, groupings could be as well subject to harmonization between both organizations.

OECD groups are available in the “UNSD-OECD detail.xls” file. Geographic zones have been extracted from the dissemination .ivt files since they are not available in the Express database.

Groupings are calculated on the fly.

8. United States Miscellaneous Pacific Islands

This territory contains Midway and Wake Islands. It is not clear if UNSD wants to keep the two codes for Midway and Wake. If yes, reported data with these two islands will be included into the corresponding codes and then added to the United States Miscellaneous Pacific Islands. Since all (or almost all) OECD countries use this code and do not report data with Midway and Wake, it is necessary to use the aggregate United States Miscellaneous Pacific Islands for the world calculation.

ANNEX 2

UNITED NATIONS
DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS

STATISTICS DIVISION
INTERNATIONAL TRADE STATISTICS BRANCH

STATISTICAL ISSUES OF UNSD-OECD DATA PROCESSING PROJECT
9 May 2005

UNSD Report on Standard Unit Values

Introduction

For the purposes of data checking and estimation of missing quantities UNSD wishes to use the most appropriate unit values available for any given commodity. Since it does matter if we are considering imports or exports and for which year we are using unit values, they will be calculated per flow and year, and - of course - separately for different units of quantities. The units of quantities which are used in the UN Comtrade database are weight (in kilograms) and the WCO recommended unit for the given 6-digit HS commodity. Other units of quantity will not be considered.

Methodology

The steps to arrive at the best estimate for a unit value (per commodity, unit, year and flow) are the following:

- i. Extract for a given commodity, unit, year and flow all reporter by partner data from UN Comtrade.
- ii. Eliminate records with a value less than US\$ 25,000.
- iii. Log transform the data and sort them in ascending order.
- iv. Take the 25th (Q1) and 75th percentile (Q3).
- v. Calculate the lower acceptance limit as $Q1 - 1.5 * (Q3 - Q1)$
- vi. Calculate the upper acceptance limit as $Q3 + 1.5 * (Q3 - Q1)$
- vii. Eliminate records outside of the acceptance range.
- viii. Calculate the mean of the accepted log transformed unit values.
- ix. Take the re-transformed unit value of step eight.

The best estimate for the standard unit value is the unit value of step nine. For a given commodity, trade flow, year and unit it is claimed that this unit value is the best available for any reporter.

For the most recent year the standard unit value may not yet have been calculated due to lack of data. In such the previous year's SUV will be taken adjusted for inflation (multiplied by total world trade current year divided by total world trade of the previous year).

Furthermore, the acceptance range can be used in the data checking procedures. For a specific commodity, flow, year and unit the observed unit value should be within the acceptance range. If not, checking needs to be done.

Data checking procedure

Checking of unit value for each commodity, year, flow and unit with partner world can be done against the acceptance levels mentioned in the first part of this note.

Secondly, unit values can be checked against the unit values of previous 5 years. In this case it is suggested to take as acceptance range the smallest SUV divided by 10 and the largest SUV multiplied by 10.

Finally, absolute values can be checked against levels of previous years. In this case it is advised to check only if in any of the previous 5 years the value exceeded 1 million US\$. Then take as acceptance range from the smallest value divided by 100 up to the largest value times 100.

Summary of decisions

- With given methodology calculate Standard Unit Values (SUV) per commodity, year, flow and unit, and use it as best estimate for unknown quantities.
- Use the acceptance range of the SUV calculation as checking range in data verification process.
- Use unit values of the reporting country for previous years' data to verify current year.
- Use absolute commodity values of the reporting country for previous years' data to verify current year.