



**STATISTICS DIRECTORATE  
STATISTICAL INFORMATION MANAGEMENT AND SUPPORT DIVISION**

STD/SIMS(2004)4  
Unclassified

**OECD EXPERT GROUP ON STATISTICAL DATA AND METADATA EXCHANGE**

**Joint OECD/UNSD Trade Database Application Project Framework**

**1-2 April  
Château de la Muette**

**JT00158703**

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

English - Or. English

## JOINT OECD/UNSD TRADE DATABASE APPLICATION PROJECT FRAMEWORK

By Trevor Fletcher  
Information Technology and Network Services, OECD

### TABLE OF CONTENTS

JOINT OECD/UNSD TRADE DATABASE APPLICATION PROJECT FRAMEWORK .....	2
1. Project Background and Overview .....	3
2. Project Deliverables.....	4
3. Task List (Work Breakdown structure) .....	5
4. Project Estimates .....	11
5. Project Risks Analysis.....	12
6. User Requirements Change Management .....	12
7. Development Approach.....	12
8. The Role of Other International Organisations in Joint Trade Project .....	13

## 1. Project Background and Overview

The phasing out of the current version of Express software by Oracle has made it necessary to redevelop the OECD Annual and Monthly International Trade by Commodity Statistics (ITCS). Following a review of the options available (upgrade to subsequent version of Oracle software; application re-use; or complete application redevelopment) it was decided to replace this system with a new Trade system developed jointly with the United Nations Statistics Division (UNSD) in New York.

Developments undertaken at UNSD indicate that it is now feasible to manage Trade data in an SQL/NT environment. Feasibility tests were carried out at the OECD during 2002 to evaluate whether the data model would be suitable for OECD Trade data and to monitor performance levels. These tests proved positive and as a result the decision was taken in conjunction with STD to go ahead with the joint project in partnership with UNSD.

### 1.1 Purpose and Scope

The purpose of the joint OECD/UNSD ITCS project is to provide a common, SQL-based repository for trade data, a set of tools for data management and a web-based end-user interface for querying the database.

This application will in turn be fully integrated with other corporate developments currently in progress for the overall statistical production process such as the *OECD.Stat* data dissemination warehouse, the *MetaStore* metadata interface and the new tools for paper and electronic publication.

The scope of the project is to cover the phases of the Trade data production process from data capture through standardisation, aggregation, harmonisation to query and analysis for end users and eventual dissemination via Electronic data Products (EDP) and paper publication.

The joint project will be applicable for annual data only. OECD Monthly trade data will be dealt with separately at a later date.

### 1.2 OECD/UNSD ITCS Project Goals

The goal of this project is to design, build, test and implement a production system for Annual Trade data.

The system will provide tools for the database administrators to carry out the following tasks:

- Database administration;
- Database activity monitoring and management;
- Data collection;
- Data importing;
- Data validation and error reporting;
- Commodity aggregation;
- Conversion of trade data to other classifications;
- Trade data aggregations at World level;

- Trade data aggregations at Total level;
- Data inspection and analysis tools;
- Process summary information;
- Publication and Data dissemination;
- Security management.

The new system should enable the STD/TASS trade database administration and the STD/SIMS publication team to process trade data within the same timeframes and with existing resource levels. The system should provide an interface to carry out the full cycle of data processing with the same levels of ease and efficiency as does the current ITCS application.

The application will cover joint data processing and a joint management system. It will allow for separate data dissemination systems.

Separate data dissemination platforms will permit autonomy for each organization in defining its own output formats.

Each organization will process its own data on site. Thereafter data will be combined onto a single joint processing server (in New York in the first instance) and subsequently be split to the respective dissemination servers. This will ensure validation and integrity checks are carried out on the joint database server prior to distribution.

### ***1.3 Approach and Method.***

#### ***1.3.1 Approach.***

The approach used by this project had been to use the existing standard software platforms of both OECD and UNSD based on the Microsoft SQL Server relational database developed by UNSD. The database application management server-side software will be developed using SQL Stored Procedures, data import procedures using SQL Data Transformation Services (DTS), ASPdotNET for client development and XML for data dissemination.

#### ***1.3.2 Method.***

The method adopted is based on shared development.

Following on from initial feasibility tests, a set of functional requirements were provided by the OECD based on the existing Foreign Trade System (**see attached Trade Description.doc**) and these were merged with the UNSD requirements to produce an agreed, joint set of processing requirements as the basis for the project.

## **2. Project Deliverables**

### ***2.1 Data and Database Structures***

Complete set of Annual International Trade by Commodities data migrated from the existing EXPRESS production environment.

## ***2.2 Application Interface for Managing the Following Tasks:***

- Security and permissions;
- Database Administration;
- Data Importing;
- Data Standardisation;
- Commodity aggregation;
- Conversion of trade data to other classifications;
- Aggregation at World level;
- Aggregation at Total Level;
- Data Inspection/Analysis;
- Data Correction;
- Process Summary;
- Data Query and Analysis;
- Data Reporting;
- Publishing data via OECD.stat.

## ***2.3 Application Outputs***

- Electronic Data Product (EDP);
- Published data via XML;
- Standard Reports (To be defined).

## ***2.4 Reports and Documentation***

- System Design documents;
- User guide;
- Technical Documentation.

## **3. Task List (Work Breakdown structure)**

### ***3.1. Initial Data Loading.***

Annual trade data will be transferred from the EXPRESS production environment to the new SQL database as a one-off exercise on a Country/Classification basis. Once the full set of data has been loaded then the two datasets will be synchronised by parallel updates as new data is provided by member countries.

#### ***3.1.1 Prepare data extracts from EXPRESS FTS database (MB).***

Schedule and run extraction procedures for EXPRESS databases to generate input files for loading into SQL database.

### **3.1.2 Import data**

Schedule and run data loading routines to transform input files to SQL format.

### **3.1.3 Validate data content. Identify any differences old/new.**

Develop validation routines to ensure coherence between EXPRESS and SQL databases.

*3.1.4 Develop mechanism to ensure that once full set of data has been transferred, subsequent EXPRESS database country data updates are reflected in SQL database in order that the two datasets are fully synchronised.*

## **3.2 Security and Permissions**

Establish a security mechanism based on NT user id to ensure specified data access roles can be defined according to agreed user profiles (administrator, guest etc).

## **3.3 Database Administration**

Interface to carry out database administrator functions (not comprehensive) including:

- Modification of Zone definitions;
- Modification of dimension elements;
- Management of classification product conversion elements;
- Management of data to be published.

## **3.4 Data Importing**

Creation of standard input files derived from the data sent by the member countries.

Development of procedures (one per country) using Data Transformation Services (DTS) to load input files into production database.

Reports will be developed to provide a log of all data importing activities.

## **3.5 Data Standardisation**

Development of procedures to ensure country data meets standards for the following dimension values:

- Commodity code;
- Quantity unit;
- Country.

Correspondence tables (one per country) will map any country-specific codes to standard values.

Value conversion will be from national currency to US\$ based on the latest available exchange rates.

Quantity units will be standardised and any conversions indicated in automatically generated footnotes.

Reports will be developed to provide a log of all data standardisation activities.

### ***3.6 Basic Level Commodity Aggregation***

Tariff Line data (original national data) are aggregated to obtain Basic Level Commodity (which is six digits in HS). Unit Aggregation includes Netweight/Qty Estimation. The rules for estimation are as follows:

#### Quantity units and quantity estimation

(a) The database will contain two quantity fields, one being the net weight and the second one - quantity in terms of the WCO recommended standard unit of quantity;

(b) For estimation of quantities, the following priority of steps will be adopted:

(i) if a direct conversion from the reported quantity unit to net weight and the WCO units is possible than this will be done;

(ii) if that is not possible then the 20% or 50% rule (depending on the development of the country - respective lists are to be prepared and agreed upon by mid 2004) will be chosen to estimate the quantity;

(iii) if that is not possible then a standard unit value will be used to estimate the missing quantity (procedures to calculate standard unit values are to be developed jointly by mid 2004).

(c) These rules should be used in succession and in relation to the aggregation from one level of the classification to the next;

(d) Quantity conversion factors will reviewed and agreed with understanding that the factors provided by the FAO will be taken to the maximum extent possible;

(e) 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.

### ***3.7 Conversion of Trade Data to Other Classifications***

The following classification types are maintained within the database:

- HS 2002
- HS 96
- HS 92
- SITC 3
- SITC 2
- SITC 1

There are two possible approaches for converting data from one classification code to another: cascading (A->B, B->C, C->D etc) or sequential (A->B, A-> C, A->D etc). It is recommended that the sequential method be used and the cascade approach would amplify imprecise representations from one classification to another whereas a sequence would limit an imprecision only to the next step and not carry it forward.

An example (Textile fibres commodity codes) would be where a good match occurred between A and D but not A and B. In this case the link between A and D would be retained using sequential conversion but lost in the cascade approach

Reports will be developed to provide a detailed log of all commodity conversion activities.

### ***3.8 Aggregation at World Level***

Aggregate partner to World totals by commodity from 6-digit -> 4-digit -> 2-digit level and the values stored in the database. The reported data must be kept and excluded from aggregation.

Any differences between aggregated totals and country reported World-level totals will be indicated in automatically generated footnotes.

Reports will be developed to provide a log of all World-level aggregation activities

### ***3.9 Aggregation to Total Level***

Totals will be aggregated from 6-digit -> 4-digit -> 2-digit level and the values stored in the database. The reported data must be kept and excluded from aggregation.

Reports will be developed to provide a log of all Total-level aggregation activities.

Any differences between the aggregated 2-digit level data and the data provided by the country will be indicated in automatically generated footnotes.

Weights aggregation will be from 6-digit -> 4-digit level and involve some estimation of missing values. To help determine estimations a threshold of dominant quantity will be established. Any estimation will be indicated in automatically generated footnotes.

Reports will be developed to provide a log of all commodity aggregation activities.

### ***3.10 Data Inspection/Analysis***

Provide means for database administrators to carry out the following checks on the processes from data importing through to Aggregation at Total level:

- Checking of values and quantities;
- Comparison of values with the previous years data to highlight any possible anomalies;
- Data reports (expand);
- Decision rules (expand);
- Comparison on unit and secondary value/quantity (clarify).



### ***3.11 Data Correction***

Provide means for correction of estimated quantities (expand).

### ***3.12 Process Summary***

Provide a report giving a summary of the whole process to allow database administrator/supervisor to review database updates.

In addition, each step of the process will be linked to a management process workflow to allow monitoring of the process from import through publishing.

### ***3.13 Data Reporting***

Detailed reports will be generated at each step of the process and be available to the database administrators.

### ***3.14 Data Query and Analysis***

A web-based client will be developed for users of the system to access the database. This will include the following features (to expand – base on UNSD client).

### ***3.15 Publishing Data Via OECD.stat***

A ‘gateway’ will be developed to pass certified data via XML into the Corporate OECDdotstat data warehouse environment for publication and online dissemination to a wider audience (expand).

This will be integrated with the wider corporate strategy for publishing currently being developed in conjunction between PAC, STD and ITN.

#### ***3.15.1 EDP***

- i. Specify EDP requirements;
- ii. Develop interface to select published data;
- iv. Develop EDP export routines;
- v. Develop EDP (Beyond2020 or other).

### ***3.16 Security***

Database administrators will have the means to manage their user profiles and data access rights to ensure that necessary security levels are in place to give staff valid permissions to update or view data:

- Define access profiles;
- Implement security database structures;
- Develop application security access.

### ***3.17 Database Administration***

Develop interface to maintain database structures.

### ***3.18 The SDMX Schema for Trade Data Exchange.***

The SDMX schema for Trade data was presented which is based on a normalised XML design. The file will use the sequential SAX parser for validation purposes.

The SDMX schema is considered an essential component of the data transfer process and efforts will be made from both sides to ensure the structure is as generic as possible to allow reuse for other applications.

It was agreed that the role of SDMX in Trade data exchange between OECD should be stressed in particular with reference to the incorporation of metadata in the existing generic schema.

### ***3.19 User Acceptance Testing***

### ***3.20 User Training***

#### 4. Project Estimates

Project Start date: 01-09-03

End date. 31-07-04

Task	Estimate (days)	Completion date	Responsible.
<b>4.0 Database Design.</b>			
Database Structure and Design		July 2003	OECD/UNSD
<b>4.1 Data Migration</b>			
Initial data load EXPRESS->SQL		Dec 2003	OECD
Data validation		Mar 2003	OECD
Data synchronisation for parallel tests.		July 2003	OECD
<b>4.2 Server Development.</b>			
1) Transform country data with XML to standard data		October 2003	UNSD
2) Import standard data into standard SQL table		October 2003	UNSD
3) Do completeness check		October 2003	OECD/UNSD
4) Normalize reference codes		October 2003	OECD
5) Map rejected code		October 2003	
6) Aggregate to Basic Level (including Net Weight and Supplementary quantity estimation)		December 2003	OECD
7) Do Net Weight, Quantity and Value checks		December 2003	OECD
8) Aggregate to Higher Levels		December 2003	OECD
9) Write and Review Processing Report		February 2004	
10) Convert to other classifications		February 2004	
Data Distribution for ITC		December 2004	UNSD
Data Inspection/Analysis		December 2003	UNSD/OECD
Data Correction		December 2003	UNSD/OECD
Data Publication (including data distribution)		March 2004	UNSD
Data exchange architecture and procedure		October 2004	UNSD
User Interface design and development		December 2003	UNSD/OECD
Report development		February 2004	OECD/UNSD
Footnotes & metadata integration		February 2004	OECD/UNSD
Data Query & Analysis deployment/modification		December 2003	UNSD/OECD
Middleware design and development (Web Services)		December 2003	UNSD

## 5. Project Risks Analysis

Risks to the delivery of the project in the agreed timetable have been identified in the following areas:

Task	Potential Risk	Impact	Risk mitigation
Data Exchange between Servers	Firewall and/or Network security issues could hinder large volume data exchange between OECD and UNSD servers.	Delay in updates between OECD and UNSD.	
Data Transfer across internet.	Delays in transfer of large volume files across internet.	Delay in updates between OECD and UNSD.	Ensure suitable bandwidth in place.
Processing flow between OECD and UNSD	Unknown incompatibility	Delay in data processing and updates.	
Multi languages interface for existing Dissemination Module	OECD may need to have other language.		
Business Logic/ Interface of existing dissemination Module	OECD may want to have different interface		
Server Processing phase.	Dual processing of country data using OECD and UNSD rules could add to processing time.	Delay in data processing and updates.	Minimise differences in processing rules.
Project development.	Insufficient development resources at disposal of project team	Delay in delivery of completed application.	

## 6. User Requirements Change Management

Additional user requirements identified during the course of the project will be assessed individually to view their impact on the project timetable. If the changes have no impact on the existing timetable then the plan will be modified accordingly.

If it is agreed that the requirements will delay the delivery of the application or cannot be absorbed by the project budget then implementation will be postponed to the next release of the application.

## 7. Development Approach

The project will be developed by ITN and UNSD staff in collaboration with external consultants. The development team will consist of Sandrine Phelipot (SP) and Monique Bouvet (MB) from ITN, Valerie Thielemans (VT) and Lynda Hawe (LH) from STD plus a number of consultants from InfoCubed(I<sup>3</sup>), a Paris-based software house. **The user team will be coordinated by Gregory Legoff from STD/TASS.**

UNSD will lead the development based on the core database management and client functionality that they have already developed during 2002-3.

MB will be responsible for transfer of existing trade data from the EXPRESS database to the SQL Server environment.

SP will develop with UNSD the stored procedures from data importation through to finalised data.

UNSD will develop the client interface and business logic for managing the whole process.

LH and VT will adapt the existing UNSD Trade client for OECD use.

Development will be shared with the initial stages developed jointly at UNSD. Subsequent phases will be developed at both sites with SP tasks co-ordinated by UNSD. First phase testing will be based on putting a single country data through the full process.

Overall project development will be managed on the OECD side by Trevor Fletcher (TF) and on the UNSD side by Ronald Jansen (RJ).

## **8. The Role of Other International Organisations in Joint Trade Project**

Other International Organisations (ITC, UNCTAD and FAO) have expressed interest in participating in the Trade project. It was agreed that any such role should be peripheral at least until the OECD/UNSD development has been stable and running in production for 1-2 years in order to avoid over-complicating the project while the core software modules are being developed.