

Seminar on International Trade Statistics

Implementation of IMTS 2010 and Follow –up to the Global Forum

26-30 September 2011

Comext overview (Eurostat Unit G5)

Benoit Faes benoit.faes@ec.europa.eu

Eurostat - Unit G5 – September 2011



Eurostat - G5 mission

- Directorate G : Business Statistics
- Unit G.5 : International trade – production
HoU : Ani Todorova (ani.todorova@ec.europa.eu)
- Mission : To provide the European Union with a quality statistical information service on the Intra- and Extra-EU trade.
- Main data flow : Detailed Intra- and Extra-EU trade data

Eurostat - Unit G5 – September 2011



Short history

- First version in 1988.
- Second version in 1992-1993, base of the current system (version 8 is standard and version 9 is coming)
- Comext is part of XT-NET
 - XT-Net Edit
 - Comext
 - XT-Net Assist
 - XT-Net Publish
 - NPS
 - Others
- Self-describing systems (structure and characteristics of data contained in the system).

Table of Contents

Production

- Main tasks
- Data Flow
- Detailed MS Data Flow
- Volumes
- Validation
- Confidentiality
- Camouflage

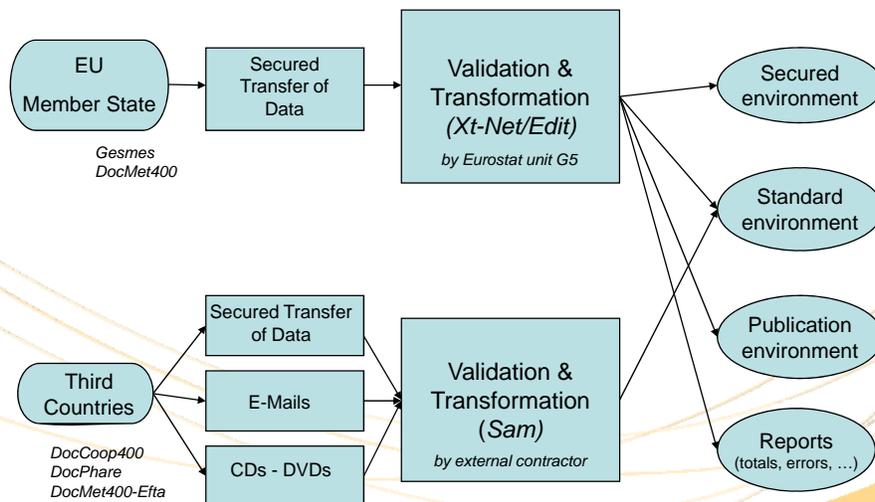
Dissemination

- Main tasks
- Data
- Data Flow – Comext Online
- Data Flow – Easy Comext
- Quality Controls
- Means and Volumes

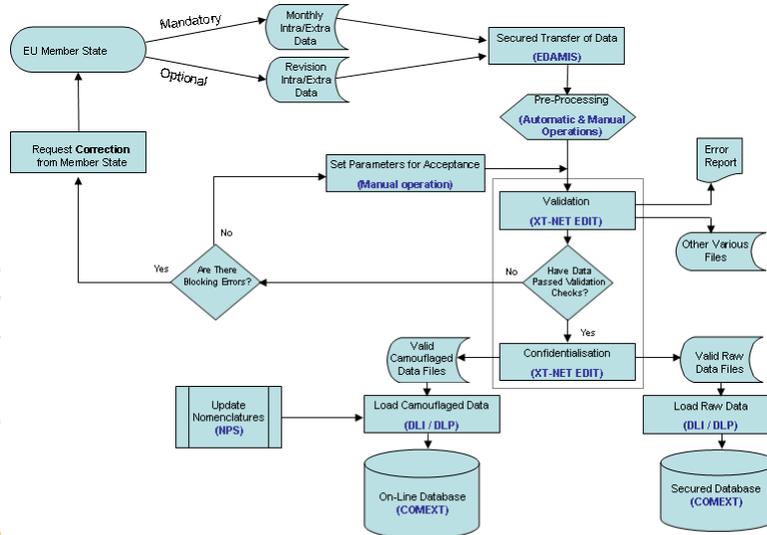
Production – Main tasks

- Data collection
 - from Member States and Third Countries
 - including revisions
 - contact Countries when data not received or over target date
- Data validation
 - detect syntax errors and some outliers
 - contact countries to confirm or not
 - manual decisions (not 100% automated)
- Data transformation
 - e.g. perform camouflage, generate aggregates
- Data Loading (Comext)
 - into standard and confidential environments
- A posteriori quality management
 - manage reported problems (contact countries, reprocess corrected data)
- Reporting
 - generate statistics such as : respect of legal delays
 - standard quality reports per Member State

Production - Data Flow



Production - Detailed MS Data Flow



Eurostat - Unit G5 - September 2011

eurostat  7

Production - Volumes

Data processed for 2009 reference periods (MSs only) :

2009	Intrastat	Extrastat	Total
# monthly periods	2500	2027	4527
first deliveries	(27x12) = 324	(27x12) = 324	648
revisions	2176	1703	3879
⇒ Average nb files per MS per month	7.7	6.3	
⇒ Average nb revisions per MS per month	6.7	5.3	
# records (millions)	429	572	1001
# Gbytes	50	66.9	116.9
# errors and warnings	83636	70304	153940
	= 1 every 5130 records	= 1 every 8136 records	

Eurostat - Unit G5 - September 2011

eurostat  8

Production - Validation

- Huge volumes (1 billion records in 2009)
- Time constraints : cannot block the loading of a monthly file if not 100% correct.
- Complexity/difficulty of validation of detailed data : cannot look manually at all errors (153940 errors in 2009)
- Examples of validation rules :
 - *checks against nomenclatures*
 - *checks for completeness (mandatory fields)*
 - *validity intervals for some numerical data / ratios,*
 - *price / net mass, price / supplementary unit, ...*
 - *amount of confidential data*
 - *amount of data with “QU / QV / QW” countries*

Thus validation performance is a balance between

- *Validating as much as possible (high data quality)*
- *Validating as quickly as possible (quick data availability)*

Production - Confidentiality

- Passive confidentiality :

MSs take appropriate measures only at the request of importers or exporters who feel that their interests would be harmed by the dissemination of the data.

- Processus :

- *MSs send to Eurostat :*
 - Full Real data (= confidential data)
 - Instructions on what can be disclosed and to whom
- *Eurostat :*
 - Uses Full Real data for internal purposes (e.g. compute TOTAL aggregates)
 - Apply camouflage according to MSs instructions
=> disseminated data

Production - Camouflage

- Type of camouflage
 - *camouflage by product*
 - *camouflage by partner*
 - *camouflage by product & partner*
- Indicators concerned
 - *all indicators (value, quantity expressed in net mass, quantity expressed in supplementary units) can be declared confidential, independently ones from the others, by using a distinct flag ("confidentiality flag") for each indicator.*
 - *values for indicators declared confidential will be associated to camouflaged data*
 - *other values will be associated to real (non camouflaged) data*

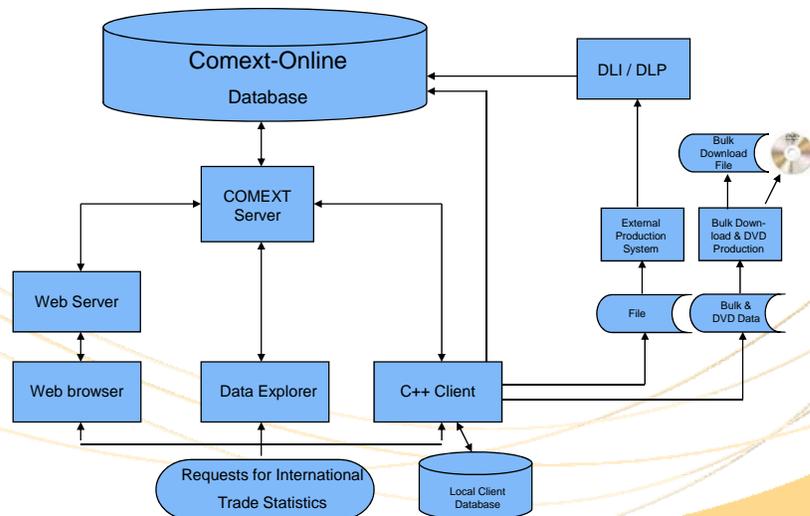
Dissemination – Main Tasks

- Disseminate detailed trade data and metadata through Comext to the Commission, EU institutions, National Administrations and general public
- Provide a continuous support to the users
- Organise co-operation activities related to data collection, validation and dissemination

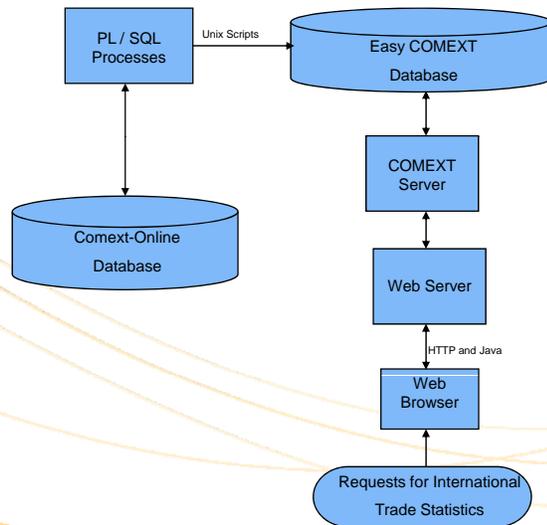
Dissemination - Data

- Public
 - EU data, EFTA, BALKAN monthly data by:
 - CN (8 digits) and HS product classifications
 - SITC up to 6 digits
 - CPA up to 4 digits (EU data only)
 - BEC (EU data only)
 - By tariff regime
 - By mode of transport
- Restricted to Commission only
 - EU data by TARIC10
 - UN and IMF databases
 - Data from cooperating third countries

Dissemination – Data Flow – Comext Online



Dissemination – Data Flow – Easy Comext



Dissemination – Quality Controls

- Visual (about 120 graphics)
- Outliers (Berthelot and Hidioglou method)
- K-Clustering
- Mirror outliers and analysis (EU only)

Dissemination – Means and Volumes

- Comext On-Line and Easy Comext
 - One terabyte database (Oracle)
 - 2-3 billions of records loaded annually
 - 800 institutional and more than 13000 public registered users
 - About 0.8 million extractions per year
 - About 40 redistributors of Comext data
- DVD – FTP
 - DVD copies distributed each month
 - Direct download 276 Gb/year
- Helpdesk
 - Calls / e-mail / XT-Net-Assist
 - Data analysis: Users, SIFs
- Courses
 - 3 courses/month in Brussels or Luxembourg

Software

- Comext (data loader, metadata manager, server, 3 user interfaces, Nomenclature Production System)
- XT-NET/Edit (validation, editing)
- XT-NET/Assist (workflow of user support)
- Nomenclature Production System (NPS)
- WEB FORMS and off-line tools
- EUROTRACE (production and dissemination system)
- SAM (validation)
- Message (XML, SDMX) standardisation

Thanks for your attention...