

STATISTICS DIRECTORATE  
COMMITTEE ON STATISTICS AND STATISTICAL POLICY

Cancels & replaces the same document of 01 March 2017

**Working Party on International Trade in Goods and Trade in Services Statistics****TOWARDS A HANDBOOK ON LINKING TRADE AND BUSINESS STATISTICS**

22-24 March 2017, OECD Headquarters, Paris

*Many OECD countries are currently developing linked Trade and Business micro datasets from which new, policy relevant statistics on economic globalization are derived. Linking trade and business statistics involves important methodological challenges, such as managing different microdata linking procedures, mitigating incomplete source data, grossing up, as well as dealing with large and complex businesses, and confidentiality issues in data dissemination. To help overcome these, the OECD is currently developing a Handbook on Linking Trade and Business Statistics that aims to address these challenges and to bring together a wide variety of country practices on data linking. This document presents an update on the progress made so far, and includes a detailed annotated table of contents. WPTGS Members are kindly invited to comment on the outline of the handbook. In addition, they are invited to contribute relevant national expertise and examples regarding the compilation issues that are addressed.*

Contact person: Diana DOYLE, E-mail: [diana.doyle@oecd.org](mailto:diana.doyle@oecd.org)

**JT03410737**

Complete document available on OLIS in its original format

*This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.*

## TOWARDS A HANDBOOK ON LINKING TRADE AND BUSINESS STATISTICS

### 1. Introduction

1. Many OECD countries are currently developing linked Trade and Business micro datasets from which new, policy relevant statistics on economic globalisation are derived, for example, on the differences in export intensity between large and small enterprises, or the value added and output of trading and non-trading enterprises. These linked Trade and Business micro datasets also form a vital building block for developing Extended Supply and Use Tables and Integrated International Economic Accounts.

2. Linking trade and business statistics involves using existing data sources in new ways. At its foundation is the knowledge that the different national statistical data sources shed light on different facets of *the same enterprises* operating in the country. However, linking trade and business statistics also involves important methodological challenges. Linking data, either administratively or through surveys, which were originally collected without the intention to engage in such exercises, often means facing challenges, such as managing different microdata linking procedures, mitigating incomplete source data, grossing up, as well as dealing with large and complex businesses, and confidentiality issues in data dissemination. To help overcome these, the OECD, in collaboration with WPTGS members (and in particular those involved in the WPTGS Informal Reflection Group on Linking trade and business statistics, as well as participants to the 2016 WPTGS Workshop on Linking Trade and Business Statistics), is currently developing a **Handbook on Linking Trade and Business Statistics** that aims to address these challenges and to bring together a wide variety of country practices on data linking. In particular, the practices highlighted are those that generate important policy relevant indicators without imposing a sizeable burden on either statistical institutes or survey respondents. Above all, the handbook is directed toward removing barriers to linking trade and business statistics by providing a common language and understanding of the various economic statistics microdata, and offering concrete methodological options being used across national statistical systems and legal infrastructures.

3. This document presents an update on the progress made so far, and includes a detailed annotated table of contents. ***WPTGS Members are kindly invited to comment on the outline of the handbook. In addition, they are invited to contribute relevant national expertise and examples regarding the compilation issues that are addressed.***

## 2. Proposed annotated Table of Contents

### *Chapter 1. Introduction*

4. The Introduction to the Handbook on Linking Trade and Business Statistics explains its main purpose, which is to summarise the infrastructure, key challenges and solutions relating to micro-data linking of international trade and business data sets. The chapters of the Handbook are organised in three main parts, reflecting the initiation, compilation, and dissemination phases. National experiences and examples are provided throughout the publication to demonstrate both first steps as well as more refined approaches taken by many national statistical agencies and central banks in order link trade and business statistics.

5. The Handbook identifies the main statistical challenges that are encountered in the data linking process and provides suggestions and concrete examples for overcoming these, taking into account the variability in national practices and resources. The methods presented build mainly on the work of national and international experts engaged in the OECD Working Party on Trade in Goods and Services Statistics (WPTGS) and in the various WPTGS workshops organised in recent years on linking data. In addition, the Handbook provides insights into how the results of microdata linking can become more integrated into existing statistical work programs, and can provide the foundation for new statistical developments built around linked databases, providing more insights without increasing response burden for firms.

### *Chapter 2. Initiation: Statistical infrastructure*

6. Chapter 2 describes the prerequisites for linking data, highlighting the importance of a well-functioning Statistical Business Register as the core of the statistical infrastructure required for data linking. The business register is often cited as the backbone of economic statistics, and this is highlighted for microdata linking trade and business statistics. The challenges of maintaining the business register, and the practical decisions taken in order to do so, are important for microdata linking experts to be aware of, regardless of their specific area of expertise, as there are implications for linking trade and business projects. When microdata linking using the statistical business register, there are several recommendations that will be expanded upon in this chapter, in order to increase awareness of how to best leverage the business identifier and statistical unit, the hierarchy, the data sources, the updating process, and the vintages of the statistical business register. This backbone role is best fulfilled if all micro-data linking is conducted with a framework of a common set of units from the business register with a specific reference date, and common vintage (or frozen frame) for all linkage projects in that year..

7. Globalisation-related data sources such as the Customs/Trade Register, Trade in Services Statistics, Structural Business Statistics/Industry Statistics, Foreign Direct Investment, and Foreign Affiliate Trade Statistics will also be discussed in the context of microdata linking. The chapter will also clearly identify the role of various classifications that will be referenced throughout the Handbook.

8. *WPTGS participants that have already contributed through communications and documentation to Chapter 2 include Banque de France and Statistics Netherlands.*

### *Chapter 3. Compilation challenges: data linking practices*

9. Linking data from different data sources is in principle relatively straightforward when a common identifier is available in all relevant sources. However, difficulties are even more pronounced when such a common identifier is absent, or when traders have business identifiers (for tax purposes) but no physical presence in the country, meaning that they are likely absent from the business register. This chapter provides an overview of linking practices across countries, providing examples of both linking with common identifiers, as well as possibilities to use other linking techniques, such as those based on

name standardisation, which can, for example, remove frequent words within firm names that are counterproductive for the search and matching exercise, and probabilistic matching, which assigns a weight to the likelihood of the match, in turn offering a structure in which to accept higher probability matches and review lower ones more extensively.

10. The *Compilation Challenges* chapters serve as important inputs to experts, both in the planning and active stages of linking trade and business statistics, as these highlight that an expert's (or team's) current challenge is one encountered by many others and show how others have approached the issue. In essence, access to these approaches improves the chance of each project pushing forward, and moreover, that each project will further enrich the stock of concrete solutions available.

11. *WPTGS participants that have already contributed through communications and documentation to Chapter 3 include Statistics Canada and Bureau of Economic Analysis.*

#### ***Chapter 4. Compilation challenges: Statistical units and classifications***

12. Chapter 4 highlights the need for consistent statistical unit definitions and classifications across data sources, and provides examples of how to deal with situations where such consistency is absent. In doing so, the chapter identifies the challenges involved in dealing with large and complex businesses, but also the opportunities that can be derived from well-developed business registers that document the structure of enterprise groups. Since many statistical offices and central banks may be linking trade and business data sources that were not originally compiled with this linking in mind, some sources may be establishment-focused and others enterprise-focused or enterprise-group focused. The discussion therefore focuses on the steps necessary to allow maximum flexibility to bring in information at establishment, enterprise and enterprise group levels but to query and compile the final dataset for a common statistical unit. Particular challenges of consolidating variables such as employment, and assigning trade to an entity will be addressed ('wholesalers'), as well as classification issues (firm size in terms of number of employees may not be available but 'large traders' may be identified through revenues, for example).

13. *WPTGS participants that have already contributed through communications and documentation to Chapter 4 include Banque de France, Bureau of Economic Analysis and Statistics Canada.*

#### ***Chapter 5. Compilation challenges: Imputations and grossing up***

14. One particular challenge of integrating administrative and survey data sources is the comparison of administrative data sources, which tend to cover the universe of firms, and sample-oriented survey data sources, which focus on obtaining data directly from a representative group (or from representative groups, if a stratified sample), requiring subsequent grossing up. However, the grossing up factors (and sample strata) in most enterprise surveys tend to be based on enterprise industry and size class, but not on other characteristics relevant for analysing globalization, such as trading status or foreign ownership. Chapter 5 focuses on national experts' solutions on grossing up and imputation methods, so as to allow universal representation as an outcome of the linking of Trade and Business Statistics. Examples include post-hoc recalibration of grossing up factors, employing logistics and other regressions in order to assign the characteristics of one surveyed firm to like un-surveyed firms, or shrinking the universe to compiling all of the variables but for large firms only, if all large firms in all industries are surveyed as a point of practice for all major surveys. Not only bottom-up, but also 'top-down' methods of deriving Trade by Enterprise Characteristics (TEC) and Services Trade by Enterprise Characteristics (STEC) data are discussed, whereby for example STEC data are calculated residually from total exports (from business surveys) and TEC data.

15. *WPTGS participants that have already contributed through communications and documentation to Chapter 5 include Banque de France, Turkish Statistical Institute, and Statistics Denmark.*

***Chapter 6. Compilation challenges: Designing statistical systems with data linking in mind***

16. With large scale investments related to micro-data linking, it is important to ensure that the infrastructure built today is as able to respond to potential new policy demands as it is current demands. Chapter 6 therefore highlights the importance of developing a data infrastructure that achieves is able to achieve this aim. The chapter focuses on how new micro data sets can produce valuable metadata that can be re-integrated into existing official data sources, securing this statistical innovation. Contributions range from new survey frames (*e.g.* survey Independent SME exporters), new stratification variables (*e.g.* Manufacturing into MNEs and non-MNEs) and monitoring international trade programs with new enterprise group-focused views.

17. *WPTGS participants that have already contributed through communications and documentation to Chapter 6 include Statistics Netherlands, Bureau of Economic Analysis and United States Census Bureau.*

***Chapter 7. Data dissemination: confidentiality concerns***

18. When developing new, more granular statistics on linked trade and business data, confidentiality requirements as well as data quality concerns present an important impediment to the dissemination of results. This chapter reviews these issues and provides an overview of existing creative approaches to overcome such constraints, such as obtaining explicit or implicit consent from respondents, the production of coefficients of variation, aggregations, as well as using the size of the sample frame instead of the sample to establish risk of individual disclosure.

19. *WPTGS participants that have already contributed through communications and documentation to Chapter 7 include Banque de France and Instituto Nacional de Estadística.*

***Chapter 8. Data dissemination: Policy relevant indicators derived from linked trade and business statistics***

20. Linked Trade and Business Statistics enable an integrated analysis of the trade-investment-production nexus and of the role of different types of firms (SMEs, MNEs) in international trade and investment. This chapter will provide compilers with concrete examples of tables that can be compiled and policy messages that can be derived from these new data, building on national examples. Breakdowns that will be discussed include firm size, firm ownership (including foreign owned firms and domestic MNEs), and trading status (used in many countries), which are among the main globalisation-related dimensions of firm heterogeneity.

21. *WPTGS participants that have already contributed through communications and documentation to Chapter 8 include Statistics Denmark and Banque de France.*

***Chapter 9. Data dissemination: Using linked trade and business data to enhance statistics on firm heterogeneity in Global Value Chains***

22. This chapter describes how linked trade and business data is not only an important new source of policy relevant information on globalisation in and of itself, but can be used as well to refine the existing accounting frameworks to better account for firm heterogeneity in Global Value Chains, and as a building block to develop Extended Supply and Use tables. Examples of various ongoing empirical projects involving OECD and member states will be highlighted.

23. *WPTGS participants that have already contributed through communications and documentation to Chapter 9 include Statistics Denmark and Statistics Canada.*