

## Routine documentation

### ‘Importing individual CIS 3 datasets and basic data processing’

Names:	<table style="width: 100%; border: none;"> <tr> <td style="width: 50%;">CIS3 Belgium</td> <td style="width: 50%;">CIS3 Latvia</td> </tr> <tr> <td>CIS3 Bulgaria</td> <td>CIS3 Lithuania</td> </tr> <tr> <td>CIS3 Czech Republic</td> <td>CIS3 Norway</td> </tr> <tr> <td>CIS3 Estonia</td> <td>CIS3 Portugal</td> </tr> <tr> <td>CIS3 Germany</td> <td>CIS3 Romania</td> </tr> <tr> <td>CIS3 Greece</td> <td>CIS3 Slovakia</td> </tr> <tr> <td>CIS3 Hungary</td> <td>CIS3 Spain</td> </tr> <tr> <td>CIS3 Iceland</td> <td></td> </tr> </table>	CIS3 Belgium	CIS3 Latvia	CIS3 Bulgaria	CIS3 Lithuania	CIS3 Czech Republic	CIS3 Norway	CIS3 Estonia	CIS3 Portugal	CIS3 Germany	CIS3 Romania	CIS3 Greece	CIS3 Slovakia	CIS3 Hungary	CIS3 Spain	CIS3 Iceland	
CIS3 Belgium	CIS3 Latvia																
CIS3 Bulgaria	CIS3 Lithuania																
CIS3 Czech Republic	CIS3 Norway																
CIS3 Estonia	CIS3 Portugal																
CIS3 Germany	CIS3 Romania																
CIS3 Greece	CIS3 Slovakia																
CIS3 Hungary	CIS3 Spain																
CIS3 Iceland																	
Dataset:	Importing and processing all CIS 3 individual country files (BE, BG, CZ, DE, EE, GR, ES, HU, IS, LT, LV, NO, PT, RO, SK).																
Variables required:	All available variables in the individual datasets																
Programme:	SPSS 22.0 Stata 12.0 SAS 9.2 R 3.2																
Abstract:	<p>Routines for 15 countries and for 4 software packages listed above (60 files altogether), are programmed to perform the following:</p> <ol style="list-style-type: none"> <li>1.) Transformation of CIS 3 country data from csv into sav (SPSS), into dta (Stata), into sas7dbta (SAS) and importing them into R.</li> <li>2.) Recoding of certain alphanumeric variables into numeric variables</li> <li>3.) Definition of variable and value labels (by executing the label syntax)</li> <li>4.) Managing missing values</li> <li>5.) Saving datasets in selected software formats</li> </ol>																
Country specifics and deviations:	<p><u>Size of enterprise:</u> Countries used different coding for <i>size98_N</i> and <i>size00_N</i> variables. The reason for that is that the size groups were collapsed during the anonymization procedure (the variables were recoded in a way that for each combination at least three enterprises exist). For year 1998 (<i>size98_N</i>), 4 initial size groups were used for some countries (10-19, 20-49, 50-249, 250+). For 2000 (<i>size00_N</i>), size groups 10-19 and 20-49 are collapsed into one group, 10-49 - small sized enterprises.</p> <p><u>Main activity of enterprise:</u> Countries used different coding for <i>Nace2_N</i></p>																

	<p>variable. The reason for that is that the main activity NACE 2-digit level groups were collapsed during the anonymization procedure and it was done differently in every country. Counting all final recoded values in all individual datasets, there are 77 different NACE activity codes.</p> <p><u>Czech Republic</u>: alternative coding is used for CO36 variable (unexplained value 3 is added to the harmonized range of values 0-1).</p> <p><u>Greece</u>: no valid values for CO cooperation partners set of variables.</p>
Sources:	All CIS 3 data and metadata documentation, available on Eurostat CIS DVDs
Author(s):	<p>Sebastian Kočar, Social Science Data Archives (ADP), University of Ljubljana</p> <p><a href="mailto:sebastian.kocar@fdv.uni-lj.si">sebastian.kocar@fdv.uni-lj.si</a>; <a href="mailto:arhiv.podatkov@fdv.uni-lj.si">arhiv.podatkov@fdv.uni-lj.si</a></p>
Software used for production:	DDIwR package for R 3.2 (programmed by Adrian Dusa)