

Statistics and Indicators on the Labour Market in the eEconomy

Mobility in the eEconomy

Definitions and concepts, database development

Peter van der Hallen
Alex Stimpson
Maarten Tielens



Copyright (2003) © STILE project

All right reserved. No part of the report may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording; and may not be quoted or cited, without prior permission in writing from the Project Co-ordinator.

This project is funded under the 'Information Society Technologies Programme of the European Commission' (IST-2000-31099).

The authors are solely responsible for the content of this report. It does not represent the opinion of the Community. The Community is not responsible for any use that might be made of data appearing therein.

Table of contents

Chapter 1 / Introduction	3
Chapter 2 / Definitions and concepts	4
2.1 Capturing the eEconomy: choices and challenges	4
2.1.1 Definition of ICT used with Labour Force Survey data	5
2.1.2 Definition of ICT used with administrative data	6
2.2 Idealism versus reality: what are the different ways of calculating mobility?	8
2.2.1 The Labour Force Survey	8
2.2.2 The Datawarehouse base-application 9	13
2.3 Formalising knowledge to improve comparability	17
2.4 Who is mobile? Ensuring age groups are comparable with existing international indicators	17
Chapter 3 / Database development	18
3.1 The European Union Labour Force Survey	19
3.1.1 Requested data	19
3.1.2 Summary of database development	20
3.2 The Belgian Labour Force Survey	21
3.2.1 Requested data	21
3.3 The Datawarehouse base-application 9	22
3.3.1 Summary of database development	22
3.3.2 Base-applications 8 and 9: database description	24
Chapter 4 / Next steps	26
Annexes	27
Annex 1 / Example of documentation for checking the coherence of, and re-coding, the EU LFS data	28
Annex 2 / Sources and most important variables of the Belgian Datawarehouse Labour Market Data	40
Bibliography	43

Introduction

This report represents the first milestone for workpackage 6 of the STILE project. Workpackage 6, entitled *Mobility in the eEconomy*, seeks to reveal determinants of the mobility of ICT workers, and to do so, two principal sources are being used: the Labour Force Survey and the Belgian Datawarehouse. However, information on ICT workers, their mobility and other socio-economic factors is only telling when it is evaluated in contrast with something else. As such, care needs to be taken to analyse the information, always placing it within a suitable context i.e. the economy as a whole, other manufacturing or service sectors, various age groups, etc.

In refining the methods for exploiting Labour Force Survey and administrative data for measuring mobility in the eEconomy, this workpackage should make a valuable contribution to the IST programme objectives and in particular to the Cross-Programme Action 8 'Statistical tools, methods, indicators and applications for the Information Society' and its focus on 'statistical disclosure control and improvement in quality and in timely and low-cost data production'. It should also be of direct interest and relevance to Objective 2 of the e-Europe Action Plan: Working in the knowledge-based economy, investing in people and skills. It is hoped that this could contribute to a more accurate monitoring of the ICT people shortages.

This first milestone - *Definitions and concepts, database development* - provides an introduction to the issues and challenges confronted in the development of this workpackage. First, it provides an overview of the various methodological questions that should be posed when developing indicators on mobility in the eEconomy, linking in these questions with the necessary reality of the available data and the structure and form in which it comes. As such, key themes are how to measure mobility, the eEconomy and knowledge.

Following this methodological discussion, a summary of the steps taken in the database development is provided, underlining the necessary procedures in the database management process as well as the thematic organisation of the data. Information is provided for each of the databases intended for exploitation and further documentation and examples are provided in annex. Chapter 4 outlines the next steps.

Definitions and concepts

One of the key concerns for a project such as STILE is to provide data, analysis and recommendations that are not just instructive on a stand-alone basis, but can also be viewed in a wider context and in comparison with related research. In order to do so, it is important that the methodological basis used, the concepts developed and the definitions employed are either rooted in or can be broadly applied to an existing set of international standards and references. This has been one of the starting points for *Mobility in the eEconomy*.

That said, most statistical projects face a dilemma between what ideally should be measured and what, in reality, can be. And so while every effort has been made to maintain full comparability with international references, in some cases the level of data availability offered by the source means that certain minor deviations have been necessary.

As well as drawing from the experiences gained through other workpackages in the STILE project, and also feeding into others, efforts have already been made to cluster with another project funded by the European Commission under its Information Society Technologies programme. Agreements between the STILE and NESIS (New Economy Statistical Information System) projects have been reached on consistent methodological approaches to ensure that comparison of results between the two can be more easily achieved.

2.1 Capturing the eEconomy: choices and challenges

Information and communication technology, or ICT, provides the basis for measuring the eEconomy. Though there are a number of different definitions around for ICT, *Mobility in the eEconomy* broadly follows that agreed by the OECD Working Party on Indicators for the Information Society. In its subsequent publication, *Measuring the ICT sector*, the OECD details the definition of ICT as the following, where some ICT subsectors at the very detailed sector of activity level (four digit) are included:

Table 2.1 OECD Definition of ICT by sector of activity (NACE/ISIC)

Description	ISIC Rev. 3	NACE Rev. 1.1
Office, accounting and computing machinery	3000	30
Insulated wire and cable	3130	31.3
Electronic valves and tubes and other electronic components	3210	32.1
Television and radio transmitters and apparatus for line telephony and line telegraphy	3220	32.2
Television and radio receivers, sound or video recording or reproducing apparatus, and associated goods	3230	32.3
Instruments and appliances for measuring, checking, testing, navigating and other purposes, except industrial process equipment	3312	33.2
Industrial process control equipment	3313	33.3
Wholesaling of machinery, equipment and supplies*	5150	51.6
Telecommunications	6420	64.2
Renting of office machinery and equipment (including computers)	7123	71.33
Computer and related activities	72	72

* Where possible, this should be limited to the wholesaling of ICT goods.

Source: Measuring the ICT sector, OECD, 2000

This is based on the principle that products in the manufacturing sector (OECD, 2000):

- must be intended to fulfil the function of information processing and communication including transmission and display;
- must use electronic processing to detect, measure and/or record physical phenomena and communication by electronic means;

and that products in the service sector:

- must be intended to enable the function of information processing and communications by electronic means.

2.1.1 Definition of ICT used with Labour Force Survey data

The definition in Table 2.1 requires information at the three digit and sometimes four digit level of NACE. Labour Force Survey (LFS) data, however, are available at the broader two digit (i.e. NACE 32), and for a minority of countries, three digit level of NACE (i.e. NACE 32.1). To ensure comparable data for a maximum number of countries, the starting point has to be NACE at the two digit level. But this entails choices being made concerning the inclusion of NACE sectors at the two digit level on the basis of the existence of ICT in some of its subsectors. Since LFS data will not be able to provide an exact measurement of ICT according to the OECD definition, this means that care must be taken not to inflate the definition of ICT, while on the other hand ensuring that the size of the population in the approximation does not increase sampling variation to the extent that it loses the representativeness of the population in question.

For this reason, certain subsectors have been excluded from ICT compared to the OECD definition due to the expected difficulty in accurately measuring total ICT employment. To judge on whether a NACE sector should subsequently be included or excluded, an extraction of data was made from Eurostat's Structural Business Statistics (SBS) database, which has data at the three and four digit level of NACE. The share of employed people in the three digit sector (or group of three digit sectors) was then calculated against the two digit level to see to what extent over-representation of individuals would result from the inclusion of the two digit level sector i.e. employment in NACE 33.2 and NACE 33.3 as a share of NACE 33.

On this basis, manufacture of electrical machinery and apparatus n.e.c. (NACE 31) has been excluded from the definition of ICT since manufacture of insulated wire and cable (NACE 31.3) accounted for under 20% of employment in NACE 31 for the majority of reporting countries between 1996 and 2000. Similarly, wholesale of machinery, equipment and supplies (NACE 51.6) only accounted for under 30% of employment in wholesale trade and commission trade, except motor vehicles and motorcycles (NACE 51). To meet the OECD definition, renting of office machinery and equipment (including computers) is required at the

NACE four digit level (NACE 71.33). This level of detail for NACE 71 is infrequently reported in the SBS, but what evidence there is suggests that it accounts for a negligible employment share, and for this reason it has been excluded too.

But while the inclusion of these NACE two digit sectors would have incorporated non-ICT workers into the ICT cohort, at the same time their omission means that certain individuals have been excluded from what, following the OECD definition, would be workers in the ICT sector.

To a certain extent, this is 'corrected' by the inclusion of NACE 33 and NACE 64 at the two digit level. In the first of these, NACE 33, though only NACE 33.2 and 33.3 are in the ICT definition, according to the SBS data, their employment share generally represents between 40 and 60%. Telecommunications (NACE 64.2), meanwhile, is very often the most important ICT sub-sector in the economy and its omission would not only mean excluding a large number of workers, but would also mean that any subsequent analysis would neglect this very significant facet of the ICT cohort.

As a result, though further definitions may still be tested, the starting point for data coming from the Labour Force Survey is the following definition of ICT:

Table 2.2 Definition of ICT used with LFS data

NACE Rev. 1.1	Description
30	Manufacture of office machinery and computers
30.0	Manufacture of office machinery and computers
32	Manufacture of radio, television and communication equipment and apparatus
32.1	Manufacture of electronic valves and tubes and other electronic components
32.2	Manufacture of television and radio transmitters and apparatus for line telephony and line telegraphy
32.3	Manufacture of television and radio receivers, sound or video recording or reproducing apparatus and associated goods
33	Manufacture of medical, precision and optical instruments, watches and clocks
33.1	<i>Manufacture of medical and surgical equipment and orthopaedic appliances</i>
33.2	Manufacture of instruments and appliances for measuring, checking, testing, navigating and other purposes, except industrial process control equipment
33.3	Manufacture of industrial process control equipment
33.4	<i>Manufacture of optical instruments, photographic equipment</i>
33.5	<i>Manufacture of watches and clocks</i>
64	Post and telecommunications
64.1	<i>Post and courier activities</i>
64.2	Telecommunications
72	Computer and related activities
72.1	Hardware consultancy
72.2	Software consultancy and supply
72.3	Data processing
72.4	Data base activities
72.5	Maintenance and repair of office, accounting and computing machinery
72.6	Other computer related activities

Note: NACE two digit level in bold text, ICT sectors in plain text, non-ICT but included in LFS definition of ICT level in italic text.

Source: Camire, Steunpunt WAV, 2003

2.1.2 Definition of ICT used with administrative data

The Belgian Datawarehouse (see description of the Datawarehouse in section 3.3) has data at a NACE five digit level, but for the moment only the three digit level is available (in the database entitled base-application 9). In the exercise below we quantify the employment share (the self-employed are excluded) of the ICT sector using a three digit and a two digit level in order to see how much additional information a three digit classification can provide us with.

Table 2.3 Number of 15-64 year-olds working in the ICT sector (Belgium; 4th quarter of 1999)

NACE Rev 1.1	Number of ICT workers 2 digit level	Number of ICT workers 3 digit level	Employment share of 2 digit level (%)
30.0	522	522	100.0
32.1	8,256	8,256	38.5
32.2	9,838	9,838	45.8
32.3	3,377	3,377	15.7
33.1	2,913	0	40.6
33.2	1,352	1,352	18.9
33.3	1,897	1,897	26.5
33.4	651	0	9.1
33.5	354	0	4.9
64.1	50,066	0	62.9
64.2	29,488	29,488	37.1
72.1	13,113	13,113	39.1
72.2	15,551	15,551	46.4
72.3	2,084	2,084	6.2
72.4	914	914	2.7
72.5	1,755	1,755	5.2
72.6	108	108	0.3
Total	142,239	88,255	

Note: ICT sectors in plain text, non-ICT but included in LFS definition of ICT level in italic text.
Source: KSZ-DWH Labour market data (Processing Steunpunt WAV)

The main advantage of using the three digit level is that it is possible to distinguish within NACE 33 between, on the one hand 33.2 and 33.3 as ICT, and on the other hand 33.1, 33.4 and 33.5 as non-ICT. Similarly one can distinguish within NACE 64 between, 64.2 as ICT and 64.1 as non-ICT. With data available at the three digit level we can exclude 33.1, 33.4, 33.5 and 64.1 which will give us a more accurate classification of the ICT sector. Doing this the total number of employees in ICT in Belgium drops from 142,200 people to 88,300 people, which is a decrease of 38%. The main cause is NACE 64.1 (post) which is almost 63% of the NACE 64 as a whole (or 35% of the whole ICT sector using two digits).

In the near future it will be possible to use data at the NACE five digit level. This will enable us to test further definitions by looking at the weight of employment of subsectors in their aggregate sector on a higher digit level, such as for renting of office machinery and equipment (including computers).

A comparison of the ICT sector at a two digit level using the Datawarehouse on the one hand and the Belgian LFS on the other provides some interesting results (bearing in mind that the Datawarehouse data are from the 4th quarter of 1999 while the Belgian LFS data are an annual average for the same year).

Table 2.4 Number of 15-64 year olds working in the ICT sector at the two digit level - comparison using the LFS and the Datawarehouse (Belgium, 1999)

NACE Rev 1.1	Datawarehouse	Labour Force Survey	Difference
30	522	7,231	+6,709
32	21,471	15,483	-5,988
33	7,167	9,243	+2,076
64	79,554	85,393	+5,839
72	33,525	30,232	-3,293
Total	142,239	147,582	+5,343

Source: LFS, KSZ-DWH Labour market data (Processing Steunpunt WAV)

On the whole, the total number of employees (15-64 year-olds) does not differ a lot using one source or the other. ICT using the LFS is less than 4% larger than using the DWH. But when we look at the two digit subsectors there are greater differences. NACE 30 and 32, especially, differ a lot. A possible explanation is that the respondents, the interviewers and/or the coders of the LFS have difficulties distinguishing between both sub-sectors as they are closely related. On the whole, this variation seems to balance itself out. The differences concerning the other subsectors are rather small.

The NACE codes on a two digit level of the LFS seem to be quite a good reflection of the register data at first sight. Detailed comparisons of results using both sources but introducing a sex and age-breakdown will be tested in the context of the STILE project.

2.2 Idealism versus reality: what are the different ways of calculating mobility?

At the heart of mobility is change. Change between two different states and at two different points in time. In order to be able to calculate mobility, therefore, you need a source that is able to provide you with information on two different points in time. However, secondly, you need a source that is able to provide you with the type of information on which you want to focus your study. So what kind of mobility is it that we want to measure? This section will outline the various different ways of calculating the mobility in which workpackage 6 is interested, what they measure and, crucially, what they miss.

2.2.1 The Labour Force Survey

Unless proxy measurements are used (i.e. changes in the number of people in the ICT sector from one year to the next as a net inflow/outflow), then it is essential to have data for two different points in time. One way of doing this is to make two extractions of data and compare for the same group of sampled individuals key information such as working status or sector of activity. The disadvantage of this method is that if a measurement of mobility between one year and the next is desired, then because respondents only participate in the survey for a limited period of time, the number of people for which data are available at both points will be small and not necessarily representative of the economy as a whole.

Another way of calculating mobility is by using information from the retrospective questions that are included in the European Union Labour Force Survey questionnaire. The wording of these questions is such that the information can be easily compared with the current year. In combining the indicators with other information relevant to the eEconomy, a number of different labour market developments can be measured. One is the total number of inflows into ICT sectors, providing an indication of the attractiveness of these sectors. As well as previously employed people, this includes inflows of people that in the previous year were either unemployed or inactive (i.e. a student). Another, more focused measurement, looks at job-to-job mobility, which is of someone that changes employer between one year and the next. Yet another can look at changes between employee and self-employed status, which can provide an indication of the level of entrepreneurship in each country. Mobility between two *same* states can be controlled for by using responses on when the person started working for this employer or as self-employed. The problem with retrospective questioning is that it relies on respondents correctly remembering a situation that was twelve months before. The ability of respondents to do so will depend partly on the type and level of detail of the question. Retrospective questioning may also pose a particular problem in the case of proxy responses, that is, if someone is answering on behalf of someone else in the household.

There may be also be differences in the way the questions are posed in the EU LFS and, for example the Belgian national questionnaire. For the purposes of measuring mobility, these differences are however minimal, and their effect on the final result is likely to be negligible.

In the context of workpackage 6, although not for this milestone, investigations will be made to identify to what extent and with what level of reliability the various different measurements can report on mobility in the eEconomy.

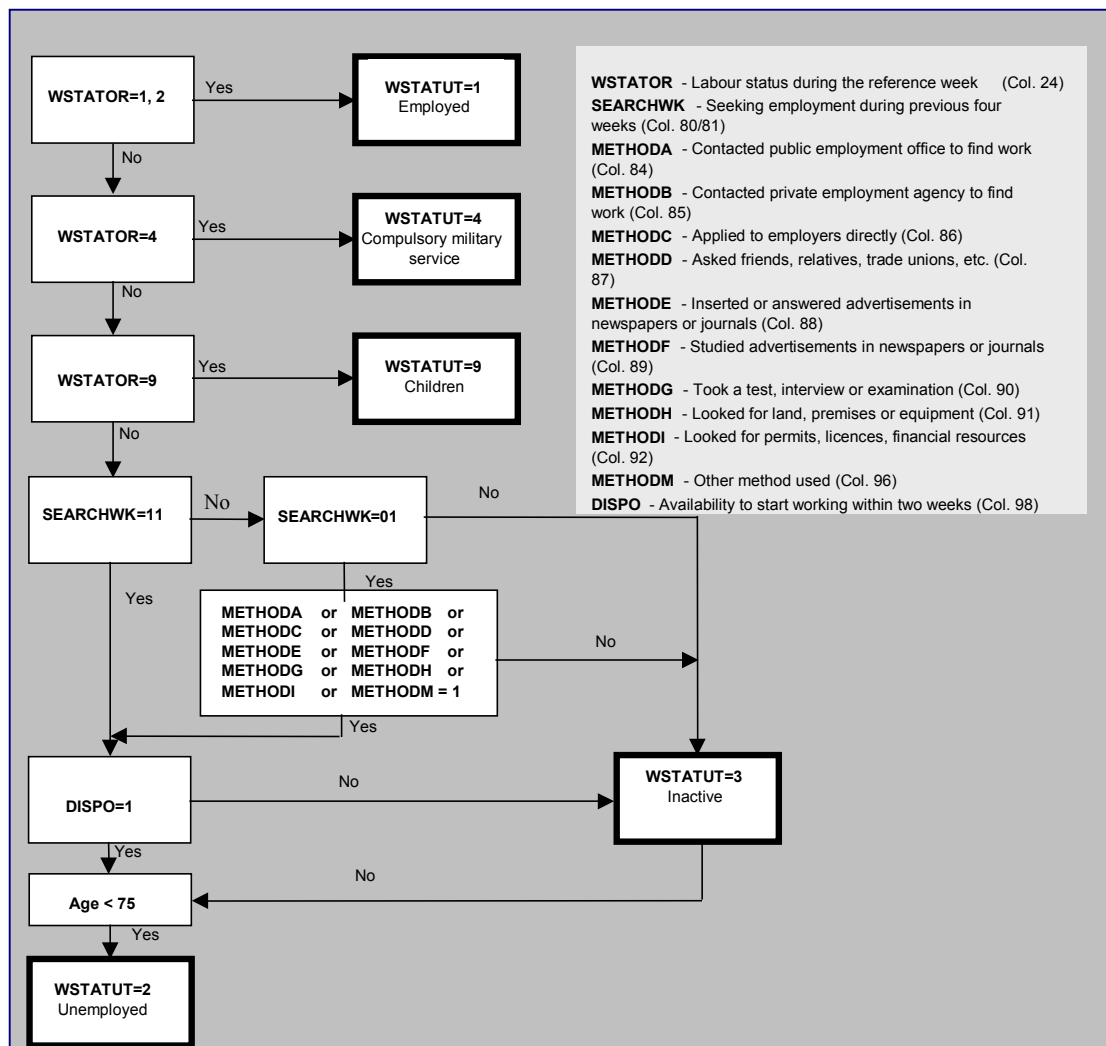
2.2.1.1 Measuring labour status mobility - inflows and outflows

A person's labour status refers to their situation vis-à-vis employment. In other words, are they employed, unemployed or inactive? Making use of the retrospective questioning in the EU Labour Force Survey, there are two main ways of calculating labour status mobility: using the variable main status or the variable working status. Details on each of the three variables are provided in Table 2.5. The variable on labour status (WSTATUT) is more complicated and is derived from a series of other variables. Box 2.1 shows how it is calculated.

Table 2.5 Questions for labour status mobility in the EU LFS

Column	Code	Description	Filter/remarks
101	Mainstat	<i>Main status</i>	Everybody aged 15 years or more
	1	Carries out a job or profession, including unpaid work for a family business or holding, including an apprenticeship or paid traineeship, etc.	
	2	Unemployed	
	3	Pupil, student, further training, unpaid work experience	
	4	In retirement or early retirement or has given up business	
	5	Permanently disabled	
	6	In compulsory military service	
	7	Fulfilling domestic tasks	
	8	Other inactive person	
	9	Not applicable (child less than 15 years)	
	Blank	No answer	
116	Wstat1Y	<i>Situation with regard to activity one year before survey</i>	Everybody aged 15 years or more
	1	Carries out a job or profession, including unpaid work for a family business or holding, including an apprenticeship or paid traineeship, etc.	
	2	Unemployed	
	3	Pupil, student, further training, unpaid work experience	
	4	In retirement or early retirement or has given up business	
	5	Permanently disabled	
	6	In compulsory military service	
	7	Fulfilling domestic tasks	
	8	Other inactive person	
	9	Not applicable (child less than 15 years)	
	Blank	No answer	

Source: The European Union Labour Force Survey, Methods and definitions, 2001

Box 2.1 Calculation of labour status variable (WSTATUT) in EU LFS

Source: The European Union Labour Force Survey, Methods and definitions, 2001

There are a few main differences between the variables WSTATUT and Mainstat. WSTATUT refers specifically to a reference week and uses responses from other questions to determine whether people fall into the ILO definitions of employed, unemployed or inactive. Mainstat, on the other hand, simply relies on the individual's subjective opinion on their main status over a reference period of three months prior to and including the reference week of the survey. It is only natural therefore to expect some differences to exist when comparing the two.

The variable used to calculate mobility in this case is Wstat1Y, referring to the individual's working status one year prior to the survey reference period. A measure of mobility relies on a changing status of the individual. With the responses to these questions, it is theoretically possible to measure such flows as labour market inflows (unemployment or inactive status such as student \Rightarrow employment), or labour market outflows (employment \Rightarrow unemployment or retirement).

The algorithm for calculating labour market inflows would be the following:

$$\frac{\text{WSTATUT} = 1, \text{Wstat1Y} = 2 \text{ to } 8}{\text{WSTATUT} = 1, \text{Wstat1Y} = 1 \text{ to } 8} \times 100$$

Note that it is important to exclude not applicable and non-response answers from the both the numerator and the denominator so as to include only those people who have been able to report on their situation one year before. Non-response could just as easily apply to inactives as to employed or unemployed individuals.

2.2.1.2 Measuring job-to-job mobility

As well as the variables WSTATUT and Wstat1Y explained in the previous section, when measuring job-job-mobility a variable is necessary that distinguishes individuals employed in both years, but with the same employer, from those that have changed between two different situations of employment. The variable referring to the time at which the person started working for this employer or as self-employed (TOW) is detailed in Table 2.6.

Table 2.6 Additional question for job-to-job mobility in the EU LFS

Column	Code	Description	Filter/remarks
40/43	TOW	<i>Year in which person started working for this employer or as self-employed</i> Enter the 4 digits of the year concerned	Col.24=1,2
	9999	Not applicable (col.24=3-5,9)	
	Blank	No answer	
44/45	TOW	<i>Month in which person started working for this employer or as self-employed</i> Enter the number of the month concerned	Col.40/43≠9999, blank & col.164/167-col.40/43≤2
	01/12	Enter the number of the month concerned	
	99	Not applicable (col.40/43=9999,blank or col.164/167-col.40/43>2)	
	Blank	No answer	

Source: The European Union Labour Force Survey, Methods and definitions, 2001

The variable TOW is calculated at Eurostat and is derived from the information contained in the above questions. In effect it separates individuals into groups according to the time they have been with their current employer or as self-employed. The groups are TOW1: <12 months; TOW2: 12-23 months; TOW3: 24 months and over.

If someone is reported in the first category, they have been working with their current employer or as a self-employed person for under one year, thereby providing information on the change necessary to distinguish them from individuals that have remained in the same job. The algorithm for calculating labour market inflows, where again it is important to exclude non-reporting, would be the following:

$$\frac{\text{WSTATUT} = 1, \text{Wstat1Y} = 1, \text{TOW1}}{\text{WSTATUT} = 1, \text{Wstat1Y} = 1, \text{TOW1 to 3}} \times 100$$

One problem with this measurement is that the data does not provide any information on the length of time that the person who has changed jobs spent in their former occupation. It will therefore be difficult to gain much understanding as to what degree knowledge diffusion in ICT will have occurred as a result of a change in employer.

Another limitation is that it refers to the status at the time of the two reference periods in year t and t-1 and so therefore does not identify other occurrences: a long period of unemployment between jobs or numerous changes of employment within the same year.

2.2.1.3 Measuring sectoral mobility - inflows, outflows and intra flows

There are two ways of measuring sectoral mobility using the EU LFS, one providing more focus than the other. The simplest, but least focused, uses information on NACE sector of

activity in both the current and the previous year to inform on whether there have been structural shifts into or out of the ICT sector. The problem with this method is that workers are often mobile within the same sector. That is to say when individuals change job, often they will merely change employer, and will carry out a similar occupation within the same sector. As a result, though suitable for measuring inflows and outflows, only using information on the NACE sector of activity will not be sufficient for capturing an accurate level of mobility *within* the ICT sector. To capture mobility within the ICT sector, for example, another method is necessary. This supplements information on NACE with that from the variable TOW.

Information on sector of economic activity is requested in the following way in the EU LFS:

Table 2.7 Questions for sectoral mobility in the EU LFS

Column	Code	Description	Filter/remarks
27/29	NACE2D/ NACE3D	<i>Economic activity of the local unit</i>	Col.24=1,2
	000	NACE Rev. 1 coded 2 or if possible 3 digits	
	Blank	Not applicable (col.24=3-5,9) No answer	
118/119	NACE1Y 2D	<i>Economic activity of local unit in which person was working one year before survey</i>	Col.116=1
	00	NACE Rev. 1	
	Blank	Not applicable (col.116=2-9 blank) No answer	

Source: The European Union Labour Force Survey, Methods and definitions, 2001

To calculate inflows and outflows, first it is necessary to select the ICT sectors as specified in section 2.1. The following algorithm can be used to calculate the number of inflows to the ICT sector (where t = time [year]).

$$NACE_{t=ICT}, NACE_{t-1 \neq ICT}$$

Furthermore, non-ICT sectors can theoretically be split/aggregated into groups of sending sectors. Indicators can be developed to show the proportion of all inflows that have gone into the ICT sector or the proportion of all ICT that are inflows. Should, in practice this provides meaningful results, the sending sectors will need to be aggregated significantly, for example into manufacturing and services.

2.2.1.4 Measuring mobility between professional statuses

A fourth measurement of mobility concerns the professional status of the individual. This differs from the working status in the sense that, for employed individuals, it provides additional focus on the type of employment from which they benefit: self-employed, employee, family worker, etc. Since a parallel question exists, providing the same breakdown of information but for the year prior to the survey, flows between employee status and self-employed could be used as an indicator for entrepreneurship and compared in the ICT sector to the rest of the economy. The detail of the information requested in the EU LFS can be seen in Table 2.8.

Table 2.8 Questions for levels of entrepreneurship in the EU LFS

Column	Code	Description	Filter/remarks
26	STAPRO	<i>Professional status</i>	Col.24=1,2
	1	Self-employed with employees	
	2	Self-employed without employees	
	3	Employee	
	4	Family worker	
	9	Not applicable (col.24=3-5,9)	
	Blank	No answer	
117	STAPRO1Y	<i>Professional status one year before survey</i>	Col.116=1
	1	Self-employed with employees	
	2	Self-employed without employees	
	3	Employee	
	4	Family-worker	
	9	Not applicable (col.116=2-9 blank)	
	Blank	No answer	

Source: The European Union Labour Force Survey, Methods and definitions, 2001

The following algorithm provides a measure of the number of individuals that previously had an employee status but have since become self-employed (with or without employees):

$$\text{STAPRO} = 1 \text{ or } 2, \text{ STAPRO1Y} = 3$$

2.2.1.5 Net flows - using changes over time as a proxy for overall mobility

The final method of measuring change in the eEconomy is the simplest. This takes the difference between, for example, the number of female ICT workers at two different points in time as a proxy measurement for mobility inflows or outflows. It should be underlined that this is a net measurement, that is the number of inflows minus the number of outflows, and so to that end will underestimate the overall level of mobility. However, with a longer time series, it will nevertheless give a good impression of the changing composition of the eEconomy and its workers.

2.2.2 The Datawarehouse base-application 9

The Datawarehouse is a database in which a series of social data from a number of social security institutions are linked permanently (for a more complete description of the Datawarehouse see section 3.3 on Database development). The data files of these institutions are related to one another via the (coded) personal identification number, which every person has for the social security (INSZ).

As each individual's socio-economic position and/or employer is known for the end of each quarter in the Datawarehouse, there are various possibilities to analyse mobility. Changes in position or in employer in between two moments can be studied; so quarterly as well as yearly comparisons are possible. This can be done by looking ahead in time (outflow: shift between t and $t+1$) as well as back in time (inflow: shift between $t-1$ and t). But just as a comparison of positions at two points in time is possible, (individual and collective) so too can career paths be constructed. It is perfectly possible, for example, to follow up a particular cohort's future career path, or to look at the employment history of a certain cohort.

The mobility, which we thus measure, is crossed with various other variables, such as the type of employment (full-time/part-time), or the employer's NACE code. In this way many questions on mobility can be answered. How many employees in ICT changed employer between 31 March and 30 June 2000? In which sectors can these people now be found? How are the youngsters who in the first quarter of 1998 were active as self-employed: how many had to stop their business by the end of 1998? Do senior unemployed find a new job easily? etc.

The example below illustrates the output and the possibilities of the Datawarehouse. The example provides an extract of a table concerning the distribution of the population by age (3 classes: 15 to 24 years; 25-49 years; 50-64 years) and the labour status every quarter (1 digit level). If desired, other variables can be added (per quarter), such as the type of employment, the NACE code, etc.

Table 2.9 Labour status progression using a datawarehouse example

Age	Labour status							Number
	1998			1999				
	2nd quarter	3rd quarter	4th quarter	1st quarter	2nd quarter	3rd quarter	4th quarter	
15-24	1	5	5	1	1	2	2	2
15-24	1	5	5	1	1	5	2	1
15-24	1	1	3	1	1	1	1	128
15-24	1	1	3	1	1	3	1	4
15-24	1	1	5	1	1	5	1	4
15-24	1	1	5	3	1	1	1	4
15-24	1	2	2	5	5	1	1	1
15-24	1	5	2	5	5	1	1	2
15-24	1	1	1					10
15-24	1	1	1	1				5
25-49	4	4	4	4	4	3	3	239
25-49	4	4	4	4	4	3	4	32
25-49	5	3	1	1	1	1	2	1
25-49	5	5	1	1	1	2	2	2
25-49	2	1	1	1	1	1	1	1
25-49	2	1	1	4	1	5	1	1
25-49	1	1	1	3	4	3	1	3
25-49	1	1	1	3	4	4	1	8
50-64	5	5	1	5	1	1	5	1
50-64	5	1	1	1	1	2	2	1
50-64	5	5	1	1	1	1	1	15
50-64	5	5	1	1	1	5	1	1
50-64	1	5	5	5	4	4	4	6
50-64	1	5	5	5	5			1

1 = employee; 2 = self-employed; 3 = unemployed; 4 = inactive; 5 = other.

Source: KZS-DWH Labour Market Data (Processing Steunpunt WAV)

There are two people in the 15-24 year old age group who were employees in the second quarter of 1998 and who were in another labour status during the third and fourth quarters only to become an employee again in the first and second quarters of 1999 and self-employed in the third and fourth quarters of 1999. On the basis of these tables short-term careers can be analysed so as to uncover career patterns. But such tables could, for example, be transformed into a matrix to quantify collective mobility flows on the labour market for two particular moments. If, for example, the calculation of mobility between the second quarter of 1998 and the second quarter of 1999 is desired, it results in the following matrix:

Table 2.10 Mobility matrix using a Datawarehouse example

Labour status 2nd quarter 1998	Labour status 3rd quarter 1998					Blank	Total
	1	2	3	4	5		
1	3,003,589	6,046	36,455	8,173	65,011	22	3,119,296
2	4,297	595,269	536	58	2,240	6	602,406
3	36,646	1,342	306,741	12,426	17,867	2	375,024
4	7,153	199	3,381	258,998	7,387	1	277,119
5	97,848	5,108	24,597	2,705	1,768,958	45	1,899,261
Total	3,149,533	607,964	371,710	282,360	1,861,463	76	6,273,106

1 = employee; 2 = self-employed; 3 = unemployed; 4 = inactive; 5 = other.

Source: KZS-DWH Labour Market Data (Processing Steunpunt WAV)

Table 2.10 shows that there were 36,455 people who were employees at the end of the 2nd quarter of 1998 and who were unemployed at the end of the 3rd quarter of 1998. The code 'blank' refers to persons who in the second quarter 1998 were registered in the Datawarehouse, but who in the second quarter 1999 disappeared from the Datawarehouse. The link with one of six social security institutions is broken for these, such as people who are no longer entitled to an unemployment benefit, people who have passed away, but particularly people who are retiring.¹

It is on the basis of such matrix mobility indicators can be calculated.

2.2.2.1 Measuring labour status mobility - inflows and outflows

Labour status mobility refers to a *change in socio-economic position/labour status* (employed, unemployed, inactive; with further possible subdivisions), but also a change in employment regime (full-time, part-time) could be looked upon as labour status mobility.

Using a mobility matrix as in Table 2.10 it is possible to quantify labour status mobility. In the example the mobility between two consecutive quarters is shown (e.g. 36,455 people who moved from employee to unemployed). But with such a matrix it is also possible to measure mobility between two moments in time on a larger time scale, e.g. between the 2nd quarter of 1998 and the 2nd quarter of 1999. In that case we have to be aware that we lose information on the 'in between' mobility, that is people who make more than one transition between these two measured moments.

The formula for calculating labour market inflows for example would be the following:

$$(\text{Labour status Q2} = 3 \text{ or } 4 \text{ or } 5, \text{ Labour status Q3} = 1 \text{ or } 2) / \text{Total Labour status Q3} = 1 \text{ or } 2 * 100$$

or

$$(36,646 + 7,153 + 97,848 + 1,342 + 199 + 5,108) / (3,149,533 + 607,964) * 100$$

2.2.2.2 Measuring job-to-job mobility

Job-to-job mobility is commonly conceived of as a change of employer. In the Datawarehouse only data about the individual's legal employer (enterprise) is included and not about the establishment. We measure mobility as a change of the relation between an individual's unique ID number and his/her employer's unique ID number. A shortcoming of this method is that we miss the job mobility within one enterprise: a clerk becoming head of his department is not regarded as job mobile. Nor is a change of establishment within the enterprise regarded as job mobile. A change in professional status (employee, self-employed or combined) too could be regarded as job-to-job mobility (an employee who becomes self-employed changes employer).

When analysing job-to-job mobility we come to a specific problem. An important cause of mobility is entry and exit of enterprises: a significant share of mobility is the result of enterprises going out of business or being restructured in such a way that they change their identification number in the registers upon which the definition of mobility is based (Nås, 2001). The identification number of enterprises can change due to administrative and economic reasons (e.g. take-over, split-off, etc.) resulting in 'false' mobility.

¹ At the moment the first steps have been made to link the databases of the social security institutions of the pensions with the Datawarehouse. This is to result in the designation of people now coded 'blank' by the code 'pension'.

To a certain degree, this 'false' mobility is corrected for in base-application 9 of the Datawarehouse. We developed an algorithm in order to make a distinction between 'individual' (I), 'collective' (C) and 'no' (Q) mobility. This algorithm has been applied to all employees who then receive a mobility code I, C, Q or blank. Only people with a 'special' regime were excluded and receive no mobility code (blank).²

The idea is that if a certain transition from employer X to employer Y between two quarters is being made by 20 employees or more, this change of employer identification number is considered to be a statistical artefact. The employee making this transition gets a code 'C' and is not considered to have been mobile.

If the employee has been 'individually' mobile (I) during one of the considered quarters the resulting mobility code is 'I'. If the employee has worked in a special regime during one of the quarters than he will receive no code (blank), unless he has been 'individually' mobile as well (then he will have code 'I'). If the employee has made a 'collective' transition during one of the quarters he will receive code 'C', unless he also worked in a special regime (then he gets no code/blank) and/or has been 'individual' mobile (then he gets code 'I'). If the employee hasn't made a transition he gets code 'Q', unless he worked in a special regime (then he gets no code/blank).

2.2.2.3 Measuring sectoral mobility - inflows, outflows and intra flows

If we select all persons with code 'I' and put them in a mobility matrix with on the axes the NACE codes of the two considered quarters we can quantify the mobility between sectors.

Table 2.11 shows that there are 36 people who has been job mobile between the 2nd and 3rd quarter of 1998 and who moved from NACE 72.1 to 74.1.

Table 2.11 Sectoral job mobility - extraction from the Datawarehouse (Belgium, 1998)

	NACE 3rd quarter 1998						Total	
	...	74.1	74.2	74.3	74.4	74.5		74.6
NACE 2nd quarter 1998								
...								
72.1		36	6		6	7		499
72.2		20	5		3	3		391
72.3		2				2		47
72.4		2				1		18
72.5		2						35
72.6								3
...								
Total		1,932	608	101	407	973	305	61,057

Source: KSE-DWH Labour Market Data (Processing Steunpunt WAV)

In the tables concerning yearly job-to-job mobility where the variable 'NACE code' is included we have to be careful drawing conclusions on shifts between two sectors since in these tables there is no information on other transitions (individual or collective) within this year. We only have the NACE code of the first and the last (the fifth) quarter. It is possible that an employee made a transition from sector X to Y, and from Y to Z. In the table we only have X and Z. To analyse shifts of employees between sectors it is better to use the tables concerning quarterly job-to-job mobility.

² There are 4 types of regime: full-time, part-time, special and indefinite. The special regime refers to seasonal work and temporary work (literally 'at short intervals'). These workers are believed to be mobile by definition and they would distort the mobility rates of the total population.

2.2.2.4 Measuring mobility between professional statuses

The professional status (self-employed, employee, ...) is a further breakdown of the labour status on one digit (working people). If we measure labour status mobility on a higher digit level we can measure professional status mobility (see nomenclature in Table 2.15).

2.2.2.5 Net flows - using changes over time as a proxy for overall mobility

A description of measuring net flows is provided in the section concerning Labour Force Survey data.

2.3 Formalising knowledge to improve comparability

Knowledge and skills are most commonly measured using data on formal education, where formal education corresponds to the years, level and type of schooling received. Though there are many other forms of knowledge that individuals acquire as a result of their personal and professional development, these are much more difficult to measure in a way that will provide meaningful and internationally comparable indicators. Formal education, on the other hand, can be measured and translated to the International Standard Classification of Education.

In the context of this workpackage, the level of education provided using Labour Force Survey data will at the very least be divided into the groups listed in Table 2.12.

Table 2.12 Education levels to be investigated

Group	ISCED level	Description
Low: lower secondary	1	Primary education or first stage of basic education
	2	Lower secondary education of second stage of basic education
Medium: upper secondary	3	Upper secondary education
	4	Post-secondary non tertiary education
High: third level	5A	Tertiary programmes that are largely theoretically based and are intended to provide sufficient qualifications for gaining entry into advanced research programmes and professions with high skills requirements
	5B	Tertiary programmes that are practically oriented/occupationally specific
	6	Second stage of tertiary education (leading to an advanced research qualification)

Source: International Standard Classification of Education, UNESCO, 1997

With regard to the Datawarehouse, the possibilities of measuring knowledge are rather limited. The major problem is that neither information about the educational attainment (ISCED) nor about the occupation (ISCO) of the individuals is available. Such an analysis will only become possible when LFS data or the 2001 census are connected to the Datawarehouse, as only these sources contain information about the educational attainment and/or the occupational code.

2.4 Who is mobile? Ensuring age groups are comparable with existing international indicators

The age groups chosen in the context of this workpackage are comparable with a host of international statistics and indicators. Providing additional focus for lower age groups, they are 15-24, 25-34, 35-44, 45-64 and 65 and over. These are the data that will be available with the Labour Force Survey. Unfortunately the Datawarehouse base-application 9 only has a distinction between 15-24, 25-49 and 50-64.

Database development

Making choices concerning the concepts and definitions is just one aspect of workpackage 6. One of the most laborious tasks though is the database development. Figure 3.1 summarises the database management procedures necessary in order to subsequently build up and analyse the desired data and indicators.

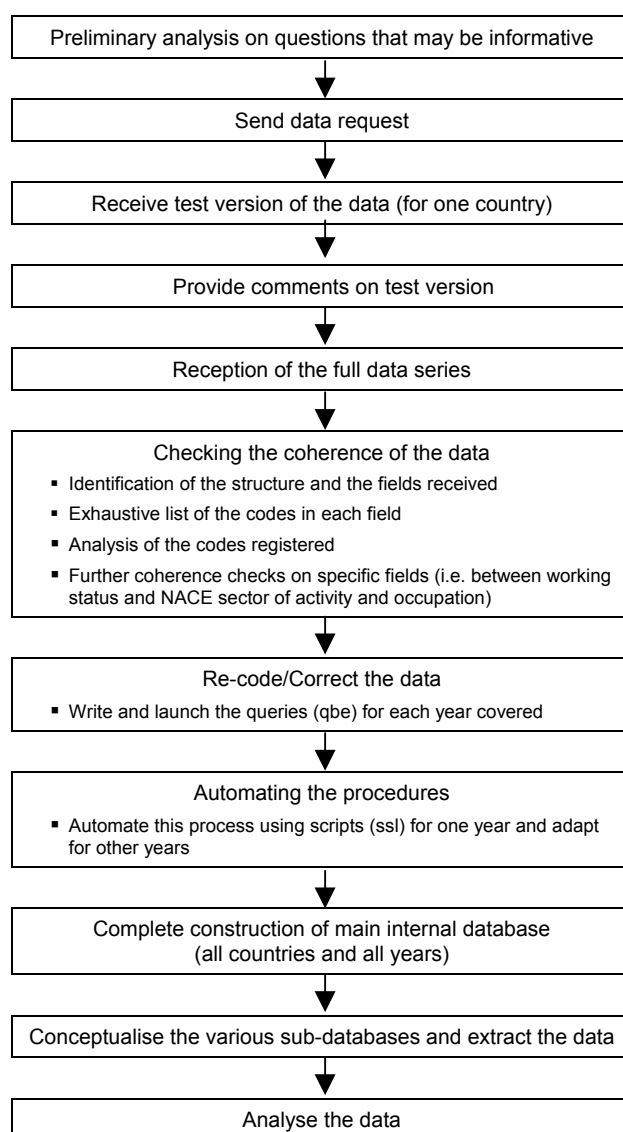


Figure 3.1 Database Management Procedures

An integral part of database management is documentation, and with a database as large as that provided by Eurostat, it becomes absolutely crucial. An example of the necessary documentation for checking the coherence of the data and the re-coding steps for the EU LFS data is provided in annex. With these documents being quite long, only an example for the year 1994 is provided.

3.1 The European Union Labour Force Survey

The European Union Labour Force Survey³ is essentially a centralised way of collecting harmonised and comparable data. The main statistical objective of the EU Labour Force Survey is to divide the population of working-age (15 years and above) into three mutually exclusive and exhaustive groups (persons in employment, unemployed persons and inactive persons) and to provide descriptive and explanatory data on each of these categories. The definitions of employment and unemployment used in the CLFS closely follow those adopted by the 13th International Conference of Labour Statisticians.

The technical aspects of the implementation of the survey are laid down in agreement with the National Statistical Institutes. On the basis of proposals from Eurostat, the Working Party on the Labour Force Sample Survey determines the content of the survey, the Community list of questions and the common coding of individual replies, as well as the principal definitions to be applied for the analyses of the results.

The National Statistical Institutes are responsible for selecting the sample, preparing the questionnaires, conducting the direct interviews among households, and forwarding the results to Eurostat in accordance with the common coding scheme.

The labour force characteristics of each person interviewed refer to their situation in a particular week. The survey is intended to cover the whole of the resident population, i.e. all persons whose usual place of residence is in the territory of the Member States of the Community, the EEA or the various candidate countries.

For technical and methodological reasons, however, it is not possible in all countries to include the population living in collective households, i.e. persons living in homes, boarding schools, hospitals, religious institutions, workers' hostels, etc.

Consequently, for the purposes of harmonising the field of the survey, results are compiled for the population of private households only. The main units of measurement for which results are obtained from the survey are individuals and households. The definition of a household varies somewhat from country to country but these differences are unlikely in the majority of cases to have a significant effect on the comparability of the results.

3.1.1 Requested data

The list of questions provided in *The European Union Labour Force Survey, Methods and definitions* was the starting point for the potential analysis. Each question was assessed for its pertinence to measuring mobility in the eEconomy. The final data request made is summarised in Table 3.1.

³ This information and more can be found in *The European Labour Force Survey, Methods and definitions*, 2001.

Table 3.1 Summary of data requested from Eurostat

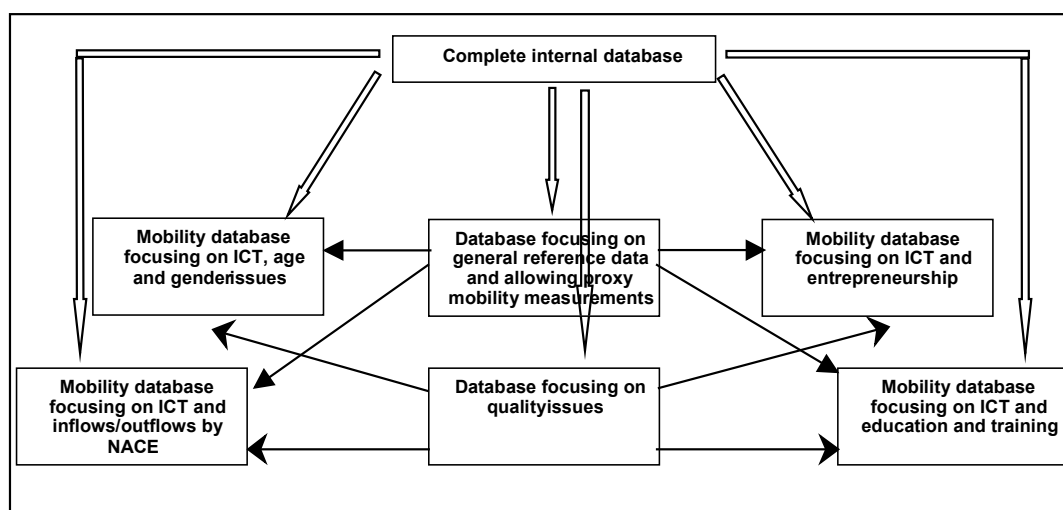
Variable	Column	Breakdown description
Countries		EU-15, plus EFTA countries, plus candidate countries where available
Time scale		1994 - 2002 where available
Regional level		NUTS 2 level
Sex	10	Male/Female
Age	11/14	Derived from year of birth (0-14, 15-19, 20-24, 25-29, 30-34 and at 5 year intervals until 65+)
Civil status	16	Single, Married, etc.
Qualification	109/110	ISCED L, M, H, blank for years 1998-2000, also ISCED H disaggregated to 5a, 5b and 6
Year in which highest level attained	112/115	
Occupation	30/33	ISCO 3 digit level
NACE sector of activity in year t	27/29	2 digit level
NACE sector of activity in year t	27/29	3 digit level for the following 2 digit sectors: NACE Rev.1 codes 24, 29 to 35, 51, 61, 62, 64 to 67, 70 to 74, 80, 85 and 92
Working status in year t	Derived 24a	Employed, unemployed, inactive
Main status in year t	101	Employed, unemployed, inactive
Working status in year t-1	116	Employed, unemployed, inactive
Year/month in which person started working for this employer or as self-employed	40/43 44/45	Derived <12, 12-23, >24 months
By NACE sector of activity in year t-1	118/119	2 digit level
Professional status in year t	26	Self-employed, employee, etc
Professional status in year t-1	117	Self-employed, employee, etc
Full-time/Part-time distinction	46	Full-time, part-time and why
Permanency of the job	47	Permanent/temporary contract and why
Working at home	58	Usually, sometimes, never
Education or training received during previous four weeks	102	None/some
Level of this education or training	104	ISCED L, M, H, blank
Purpose of this education or training	105	Skills, vocational, specific, etc.
Nature of participation in the survey	23	Direct, via another member of household

Source: The European Union Labour Force Survey, Methods and definitions, 2001

3.1.2 Summary of database development

Though the main internal database contains all the data necessary to conduct the analyses, they are not necessarily in the form that is desired. For example, age groups may be in five year bands whereas an analysis in ten year bands provides more accessible results. Or, otherwise, NACE sectors of activity are at too detailed a level to provide reliable information on the number of employed individuals and need to be grouped at a broader level. Because of the way the database is provided, however, with an estimated value representative of the sum of the different permutations (i.e. female, 25-29, highly educated, working in profession X, employed, in NACE sector X, unemployed last year, working at home with frequency X, etc.) choices need to be made as to which data it is desirable to have at a detailed level and which data can be aggregated.

An effective way of doing this is to devise sub-databases according to theme, each having a different focal point, as summarised in Figure 3.2. This allows the creation of databases of a size manageable to conduct the descriptive statistical analysis necessary for an insight into *Mobility in the eEconomy*.



Source: Camire

Figure 3.2 Diagram of databases

3.2 The Belgian Labour Force Survey

3.2.1 Requested data

The data requested from the Belgian National Statistical Institute is detailed in Table 3.2.

Table 3.2 Summary of data requested from Belgium NIS

Variable	Column	Breakdown description
Time scale		1999-2002
Regional level		Flanders, Wallonia, Brussels
Sex		Male/female
Age		15-64y divided in 5-yearclasses (15-19, etc.)
Qualification	78	ISCED 1 digit: 0, 1, 2a, 3a, 3c, 4, 5a, 5b, 6
Working status in year t	Derived from 1-6	Employed, unemployed, inactive
Main status in year t	81	
Main status in year t-1	59 and 65	
Professional status in year t	7	Self-employed, employee, etc.
Professional status in year t-1	68	Self-employed, employee, etc.
NACE sector of activity year t	11a	2 digit level aggregated (starting STILE classification)
NACE sector of activity year t-1	71	2 digit level aggregated (starting STILE classification)
Occupation current job	9	ISCO 1 digit level
Occupation last job	69	ISCO 1 digit level
Address of current job	10	Province
Address of last job	70	Province
Year in which person started working for this employer or as self-employed	13	Year
Year in which last job was ended	61-66	Year
Most important reason to leave last job	67	Retirement, dismissal, end of temporary contract, etc.

Source: Belgian LFS questionnaire, NIS Belgium

3.3 The Datawarehouse base-application 9

3.3.1 Summary of database development

3.3.1.1 Introduction to the Datawarehouse

The Datawarehouse 'Labour Market Data' was set up by request of a number of social security institutions and scientists. The objective was to create a database in which a series of (quarterly) social data which these institutions have at their disposal, were to be linked permanently so that these could be accessed better for scientific research.⁴ The linked data go back to the 2nd quarter of 1998.

This Datawarehouse has a double advantage. On the one hand, it has become possible to access in an easy manner and at a minimal cost data on the socio-economic position of the population. On the other hand, very diverse and detailed statistics can be generated about the functioning of the labour market.

The basis for the linking is the '*social security identification number*', a unique identification number which all persons known to the Belgian social security institutions possess (this number is coded and made anonymous in the Datawarehouse). The individuals are therefore the most important statistical unit. For persons in employment the characteristics of the employment regime, the employer and the number of hours worked are also integrated in the Datawarehouse, a fact which makes statistics about jobs, employers and employment volume possible.

The population of the Datawarehouse consists of all persons who during one quarter were known to one of the institutions involved. Added to these are the individuals' family members (in as far as these are not known themselves by one of the participating institutions). Of the latter group we merely know the sex, the age and the home address.

In terms of the labour force, the Datawarehouse therefore comprises the majority of working inhabitants of Belgium and a large part of the unemployed. Missing from the group of employed are especially the employees who are working for an employer that is not required to contribute to the Belgian social security, such as frontier workers employed abroad. Of the total population of unemployed, the individuals who are not - be it directly or indirectly - entitled to an unemployment benefit remain beyond the scope of the Datawarehouse.

Each participating social security institution offers an extensive list of variables in the Datawarehouse. Beside this, a number of derived variables are created, such as the socio-economic position or the number of jobs held by one person.

3.3.1.2 Nomenclature of socio-economic positions

On the basis of the information from the participating social security institutions, a detailed division of the population by labour status is drawn up in the Datawarehouse. Thereby the situation on the last day of the quarter is systematically taken into consideration.⁵

The nomenclature of these socio-economic positions is built up hierarchically and can be broken down to a five digit level. This allows for a far-reaching division within the four major cate-

⁴ The social security institutions involved are Rijksinstituut voor Ziekte en Invaliditeitsverzekering (RIZIV), Rijksdienst voor Kinderbijslag van Werknemers (RKW), Rijksinstituut voor de Sociale Verzekering van Zelfstandigen (RSVZ), Rijksdienst voor Sociale Zekerheid (RSZ), Rijksdienst voor Sociale Zekerheid voor Plaatselijke en Provinciale Overheden (RSZPPO), and Rijksdienst voor Arbeidsvoorziening (RVA). See Annex. The linking is done by Kruispuntbank Sociale Zekerheid (KSZ); Maatschappij voor Meconografie (SmalS-MvM) is responsible for the IT-technical support, for the scientific support the project relies on Steunpunt Werkgelegenheid, Arbeid en Vorming (Resource Centre for Labour Market Research, Belgium) and Point d'Appui Travail Emploi Formation.

⁵ Not all social security institutions share this approach. For the RVA, it suffices that a person was unemployed for 1 day to be counted as unemployed, whereas for the Datawarehouse a person needs to be unemployed on the last day of the quarter to be considered a member of the unemployed population.

gories. The new possibilities of the Datawarehouse are exploited to the full by mapping positions for which different social security institutions provide the information (e.g. persons who are both employees and self-employed). As an illustration this is shown in Table 3.3 at the three digit level.

Table 3.3 Datawarehouse Labour Market. Overview of the socio-economic nomenclature, age group 18-64 years (Belgium, 2nd quarter 1998)

Code and description of the socio-economic position	Number	%
1. Employed	3,704,142	62.7
1.1. Employee	2,990,375	50.6
1.1.1. In a single employment	2,873,600	48.7
1.1.2. In several jobs	116,775	2.0
1.2. Self-employed	547,846	9.3
1.2.1. Main occupation	513,550	8.7
1.2.2. Secondary occupation	20,884	0.4
1.2.3. Self-employed after retirement age	13,412	0.2
1.3. Helping an employer who is self-employed	43,141	0.7
1.3.1. Employed as helper as main occupation	42,295	0.7
1.3.2. Employed as helper as secondary occupation	354	0.0
1.3.3. Employed as helper after retirement age	492	0.0
1.4. Employee and self-employed	122,780	2.1
1.4.1. Mainly employee	111,615	1.9
1.4.2. Mainly self-employed	11,165	0.2
2. Unemployed with RVA benefits	374,613	6.3
2.0.1. Unemployed after full-time employment	250,493	4.2
2.0.2. Unemployed after studies, qualifying for 'waiting benefit'	101,447	1.7
2.0.3. Unemployed after a voluntary part-time job	22,554	0.4
2.0.4. Unemployed after graduation, entitled to a 'bridging benefit'	119	0.0
2.0.5. Unemployed without benefit (new status)	0	0.0
3. Inactive (with RVA benefits)	276,794	4.7
3.0.1. Full-time early retirement	117,438	2.0
3.0.2. Full-time career break	17,163	0.3
3.0.3. Exemption from registration as unemployed	142,193	2.4
4. Unknown	1,549,727	26.2
4.0.1. Suspended unemployed	1,191	0.0
4.0.2. Other	1,548,536	26.2
Total	5,905,276	100.0

Source: KSZ-DWH Labour market data (Processing Steunpunt WAV)

3.3.1.3 Possibilities and limitations

The data files of the participating social security institutions are related to one another via the (coded) personal identification number which every person has for the social security (INSZ).

In this manner a double counting of persons who appear in several files is prevented. A person who, for example, is both self-employed and an employee is counted only once and is put in the category 'Employee and self-employed'. Thus a more exact determination of the size of the labour force and its segments (employed, self-employed and unemployed) can be obtained.

By linking the data over time, new frontiers of the labour market sciences may be explored. We thus acquire a (more) precise view on the nature and the scale of mobility on the labour market, and studies of the dynamic processes of inflow and outflow in the labour market are more readily accessible.

Finally, the geographic division is based on a person's address, allowing for data on a very detailed geographic level (neighbourhood level).

A first important hiatus relates to the population. The Datawarehouse is based on the administrative data of the Belgian institutions for Social security. This means that, concerning the employee population, the Datawarehouse comprises solely data about the persons who are employed by an employer obliged to make declarations to the Belgian social security. Outbound frontier workers are missing, incoming ones are included. Naval staff too is missing. Apart from these, there are also self-employed who need not report to the social security. Furthermore, of the unemployed only those individuals that directly or indirectly receive an unemployment benefit are known.

All this implies that we do not know the total active population in the Datawarehouse. But the larger part of this missing population belongs to the inactive population. In particular, great many pensioners remain unknown. In addition, the people living on social assistance⁶ are missing. In total the Datawarehouse contains some 9 million persons (from a total of 10 million Belgians). From these, 7.5 million are registered by one of the social security institutions. Of the remaining people, we only know that they are a member of the family of a person known to one of the institutions. In the future the number of persons known in the Datawarehouse will rise due to, among other things, the linking of the databases about pensions to the Datawarehouse.

Moreover, not all variables in the Datawarehouse meet the needs of labour market research. After all, the data are collected for administrative purposes and therefore are at times insufficient for statistical and research purposes. A striking example of this is the information about the *local business units (establishments)*, which is not available because there is only a centralised databank to work on. Moreover, each social security institution follows its own logic and concepts, a fact that often renders comparisons between data of different institutions all but self-evident. These problems will be resolved in the future by two forthcoming projects. The project Multifunctionele Aangifte (Multifunctional Registration) will result in every employer making one integrated (electronic) declaration, which is used by all social security institutions together. Therefore all concepts like 'wage', 'working day', etc. will be harmonised. The project Kruispuntbank Enterprises ensures that every establishment obtains a unique identification number. In this way statistics will become possible for both enterprises and local establishments.

Lastly, a number of important socio-economic data are simply not covered by the Datawarehouse. The Datawarehouse comprises, for example, no information about the educational attainment of the persons concerned. This could partly be solved by integrating the educational data (and perhaps also other valuable information) of the Census 2001 on the basis of the national identification number in the Datawarehouse. Also linking the information with that from the Labour Force Survey can provide additional information about the educational attainment of individuals (and perhaps about other themes as well).

Other data sources which can signify an important surplus value to the Datawarehouse are the 'pensioenkadaster' ('register of pensions') and the data about whoever claims a social assistance.

Despite the enormous possibilities, which the current design of the Datawarehouse already offers for labour market research, from the perspective of broader socio-scientific research it is merely a first building block for a larger Datawarehouse of socio-economic data.

3.3.2 Base-applications 8 and 9: database description

To render the Datawarehouse smoothly accessible a number of base-applications were produced, illustrating what is possible with the Datawarehouse and comprising the most wanted variables. These base applications are accessible to all. Specific statistics can be requested. There are two base applications (8 and 9) on mobility on the labour market: base-application 8 considers labour status mobility and base-application 9 considers job-to-job mobility. Each

⁶ This 'social assistance' is a social benefit which guarantees a minimum income to people with no or few means of existence; a subsistence level allowance.

base-applications is built up with a number of tables including some relevant variables. These tables are described below.

8-I Quarterly mobility (2 consecutive quarters)

Distribution of the population according to sex, age, socio-economic position (three digits) and regime on the last day of quarter x, and according to socio-economic position and regime on the last day of quarter x+1.

8-II Quarterly mobility (7 consecutive quarters)

Distribution of the population according to sex, age, socio-economic position (one digit) and regime on the last day of quarter x, and according to socio-economic position and regime on the last day of the 6 following quarters.

8-III Quarterly mobility (7 consecutive quarters)

Distribution of the population according to sex, age, socio-economic position and regime on the last day of quarter x, and according to socio-economic position and regime on the last day of the six previous quarters.

Not yet available.

9-I Quarterly mobility (2 consecutive quarters)

Distribution of the employee population according to sex, age, NACE code (three digits) and regime on the last day of quarter x and according to NACE code (three digits), regime and mobility code on the last day of quarter x+1. Only persons who are employees at both points in time are included.

9-II Yearly mobility (2 quarters)

Distribution of the employee population according to sex, age, NACE code (three digits) and regime on the last day of quarter x and according to NACE code (three digits), regime and mobility code on the last day of quarter x+4 (the same quarter one year later). Only persons who are employees during all five quarters are included.

9-III Quarterly mobility (5 consecutive quarters)

Distribution of the employee population according to sex, age and regime on the last day of quarter x and according to regime and mobility code on the last day of the four following quarters. Only persons who are employees during all five quarters are included.

Next steps

This document has underlined the methodological, definitional and conceptual issues faced in order to measure mobility in the eEconomy. The different types of mobility have been underlined - unemployed to employed, job-to-job, sectoral, etc. - and it is clear that there are quite a number of different ways in which it will be possible to measure mobility. Each method has its merits, but also its shortcomings. In the next phase of the project it will be necessary to test each of these different methods in a qualitative sense and also to ascertain quite what effect this has on the mobility indicators themselves.

The section on database development has highlighted the variety and far-reaching nature of the accompanying socio-economic variables. The EU and Belgian Labour Force Surveys provide a multitude of variables which allow an analysis of the determinants of mobility in the eEconomy: age groups, educational attainment, marital status, permanency of the job, etc; although it is likely that these variables will often need to be analysed independently of one another due to sampling variation in the statistics. As such, the Belgian Datawarehouse, as well as providing valuable detailed information on sector of activity that cannot be provided by the Labour Force Survey, can act as a control mechanism to the more aggregated Labour Force Survey indicators. Furthermore, correcting the administrative data for take-overs and split-offs should have a considerable positive effect on the accuracy of the administrative data as far as measuring mobility is concerned.

The next phase of the project, therefore, will concentrate on the analysis of data and indicators, seeking to reveal determinants of mobility in the eEconomy and, at the same time, comparing these indicators - or placing them in context - with other sectors, other countries and over time.

Annexes

Example of documentation for checking the coherence of, and re-coding, the EU LFS data

Réception des fichiers STILE en date du 10/04/2003 ALL_1994

Réception

Les fichiers contenant les données originales ainsi que les documents méthodologiques si rapportant et les mails se trouvent dans:

1-database\02-R&D_2002-2003\03-HRST_ALL\05-HRST_2003\03-HRST_STILE\01-DATA_ORI\02-reception_fichier_STILE_10-04-03

Principe de mise en format des fichiers reçus:

- Ouvrir les fichiers CSV avec Word
- Dans Word, faire 'Save as ...' en text only
- Fermer fichier
- Importer ce nouveau fichier txt en paradox
- Créer sous Paradox une structure de table vide délimitant les champs des variables
- Ajouter à cette structure vide le fichier txt
- Fusionner les fichiers EU15 et Other obtenus

Préparation des données originales

Au final, on obtient une table: **ALL_sti30401_1994.db soit 659 064 enregistrements**

Structure des tables avant recodification

CHAMPS	sti30401_1994MS	Remarques
COUNTRY	01.B - 02.DK - 03.D - 04.EL - 05.E - 06.F - 07.IRL - 08.I - 09.L - 10.NL - 12.P - 15.UK	Pas de blancs
YEAR	1994	Pas de blancs
REGIONOR	Voir impressions des codes	Pas de blancs
SEX	Female, male	Pas de blancs
AGE	00-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65+	Pas de blancs
ISCED1D	1. Low, 2. Medium, 3. High, No Answer, Not Appl.	Pas de blancs
ISCED2001	9, L, M	Existence de blancs pour tous les pays
ISCO3D	Voir impressions des codes	Existence de blancs pour BE, DK, FR, IE, IT, LU, NL, UK
NACE2D	Voir impressions des codes	Existence de blancs pour DK, FR, IE, LU, NL, UK
NACE3D	Voir impressions des codes	Pas de blancs
WSTATUT	1. Employed, 2. Unempl., 3. Inactive, Not Appl.	Pas de blancs

MAINSTAT	Champ non renseigné	Existence de blancs pour tous les pays
WSTAT1Y	1, 2, 3, 6, 8, 9	Existence de blancs pour DK, DE, ES, FR, IE, LU, UK
NACE1Y2D	Voir impressions des codes	Existence de blancs pour BE, DK, DE, FR, IE, LU, NL, UK
STAPRO	1, 2, 3, 4, 9	Existence de blancs pour DK, ES, IE, PT, UK
STAPRO1Y	1, 2, 3, 4, 9	Existence de blancs pour BE, DK, ES, IE, UK
ISCEDYEAR	No answer	Pas de blancs
MSTATUS	1, 2, 3, 4	Existence de blanc pour DK, FR, IE, NL
FTPT	1, 2, 3, 4, 5, 6, 7, 9	Existence de blancs pour DK, ES, FR, IE, PT, UK
PERM	1, 2, 3, 4, 5, 6, 9	Existence de blancs pour DK, DE, ES, FR, IE, PT, UK
HOMEWK	1, 2, 3, 9	Existence de blancs pour DK, DE, ES, FR, IE, LU, UK
EDUC4WN	0, 1, 9	Existence de blancs pour DK, DE, FR, IE, LU, PT, UK
EDUCLEVE	Champ non renseigné	Existence de blancs pour tous les pays
PURP4WN	Champ non renseigné	Existence de blancs pour tous les pays
NATPART	1, 2, 9	Existence de blancs pour DK, DE, FR, IE, IT
TOW	<12 months, 12-23 months, (<24-59 months, >60 months), No answer	Pas de blancs
POP	Valeurs numériques	Pas de blancs
RELIAB	'a', 'b'	Le blanc signifie que la donnée est fiable

Vérification de cohérence des données

- COUNTRY & REGIONOR
Regarder si pas d'incohérences (voir impression)
- AGE & ISCED1D
Pour ISCED1D="NotAppl.", le résultat doit être égal à AGE="00-14"
- ISCO3D & WSTATUT
Pour ISCO3D="999", le résultat ne doit pas être égal à WSTATUT="1.Employed"
Pour WSTATUT="1.Employed", le résultat ne doit pas être égal à ISCO3D=999
Cas ou ISCO3D=999 ET WSTATUT="1.Employed" (DE, NL, PT)
- NACE2D & WSTATUT
Pour NACE2D="00", le résultat ne doit pas être égal à WSTATUT="1.Employed"
Et inversement.
- NACE2D & NACE3D
Pour NACE2D=blank, le résultat ne doit pas être incohérent (exemple: NACE2D=blank → NACE3D="X" => OK, mais NACE2D=blank → NACE3D=XX not OK) (voir impression)
- AGE & MAINSTAT
Pour MAINSTAT="9", le résultat doit être égal à AGE="00-14"
MAINSTAT non renseignée
- AGE & WSTAT1Y
Pour WSTAT1Y="9", le résultat doit être égal à AGE="00-14"

- WSTATUT & FTPT
Pour FTPT= (not 9, not blank), le résultat doit être égal à WSTATUT="1.Employed"
Pour WSTATUT=(2.Unempl. or 3.Inactive or Not Appl.), le résultat doit être égal à FTPT="9"
- ISCED1D & ISCED2001
Cohérence des données (voir impression)
Les données concernant ISCED2001 ne sont pas renseignées intégralement avant 1998.
- STAPRO & WSTATUT
Pour STAPRO=not blank, not 9, WSTATUT="1.Employed" ou "2.Unempl."
- STAPRO1Y & WSTATUT1Y
Pour STAPRO1Y=not blank, not 9, WSTATUT1Y= 1 ou 2
- HOMEWK & WSTATUT
Pour HOMEWK=not blank, not 9, WSTATUT=1 ou 2
- EDUC4WN & AGE
Pour EDUC4W=not blank, not 9, AGE ne doit pas être égal à 00-14.
- EDUCLEVE & AGE
Pour EDUCLEVE=not blank, not 9, AGE ne doit pas être égal à 00-14.
EDUCLEVE non renseignée
- PURP4WN & EDUC4WN
Pour PURP4WN=not blank, not 9, EDUC4WN doit être égal à 1
Pour EDUC4WN =1, PURP4WN ne doit pas être égal à blanc
PURP4WN non renseignée

Recodification des fichiers pour harmonisation

Copie de sti30301_ALL.db = ALL_1994-2.db

COUNTRY:

001-recodif_ALL_1994-2.qbe

01.B	→	BE
02.DK	→	DK
03.D	→	DE
04.EL	→	GR
05.E	→	ES
06.F	→	FR
07.IRL	→	IE
08.I	→	IT
09.L	→	LU
10.NL	→	NL
11.A	→	AT
12.P	→	PT
13.FIN	→	FI
14.S	→	SE
15.UK	→	UK
16.IS	→	IS
17.NO	→	NO
18.CH	→	CH
19.BG	→	BG
20.CY	→	CY
21.CZ	→	CZ
22.EE	→	EE

23.HU	→	HU
24.LT	→	LT
25.LV	→	LV
27.PL	→	PL
28.RO	→	RO
29.SI	→	SI
30.SK	→	SK

YEAR

Pas de modifications

REGIONOR

002-recodif_REGIONOR_ALL_1994-2.qbe

BE 11	→	21
BE 12	→	24
BE 13	→	24
BE 15	→	22
BE 18	→	23
BE 19	→	25
BE 22	→	31
BE 23	→	32
BE 24	→	33
BE 26	→	34
BE 27	→	35
BE 30	→	1
DE 10	→	F
DE 20	→	6
DE 10	→	F
DE 3A	→	91
DE 3B	→	92
DE 3C	→	93
DE 3D	→	94
DE 40	→	5
DE 51	→	A1
DE 52	→	A2
DE 53	→	A3
DE 54	→	A4
DE 55	→	A5
DE 6A	→	71
DE 6B	→	72
DE 6C	→	73
DE 71	→	B1
DE 72	→	B2
DE 73	→	B3
DE 81	→	11
DE 82	→	12
DE 83	→	13
DE 84	→	14
DE 91	→	21
DE 92	→	22
DE 93	→	23
DE 94	→	24
DE 95	→	25
DE 96	→	26
DE 97	→	27
DE A0	→	C
DE B1	→	3
DE B2	→	3
DE C0	→	4

DE	D0	→	8
DE	E1	→	D1
DE	E2	→	D2
DE	E3	→	D3
DE	F1	→	E1
DE	F2	→	E2
DE	F3	→	E3
DE	G0	→	G
GR	30	→	3
ES	30	→	3
ES	70	→	7
FR	10	→	1
FR	30	→	3
IE	02	→	01
IE	03	→	01
IE	04	→	01
IE	05	→	02
IE	06	→	02
IE	07	→	02
IE	08	→	02
IT	20	→	2
IT	40	→	4
IT	60	→	6
IT	70	→	71
IT	81	→	72
IT	82	→	78
IT	A0	→	A
IT	B0	→	B
NL	23	→	21
NL	24	→	22
NL	25	→	23
NL	51	→	41
NL	52	→	42
NL	71	→	31
NL	72	→	32
NL	73	→	33
NL	74	→	34
PT	20	→	2
PT	30	→	3
UK	10	→	C
UK	20	→	E
UK	30	→	F
UK	40	→	H
UK	50	→	7,777
UK	60	→	K
UK	70	→	G
UK	80	→	D
UK	90	→	L
UK	A0	→	M
UK	B0	→	N

003-concat_REGIONOR_ALL_1994-2.qbe
Concaténation de la variable REGIONOR en fonction de la variable COUNTRY
Quelque soit a b → (a+b)

SEX
004-recodif_SEX_ALL_1994.qbe
Male → 1
Female → 2

AGE
005-codif_AGE_ALL_1994-2.qbe
"00-14" → A1
"15-19" → A2
"20-24" → A3
"25-29" → A4
"30-34" → A5
"35-39" → A6
"40-44" → A7
"45-49" → A8
"50-54" → A9
"55-59" → A10
"60-64" → A11
"65+" → A12

ISCED1D
006-codif_ISCED1D_ALL_1994-2.qbe
"1. Low" → 1
"2. Medium" → 2
"3. High" → 3
"No answer" → 8,888
"Not appl." → 9,999

ISCED2001
007-codif_ISCED2001_ALL_1994-2.qbe
L → 1
M → 2
9 → 9,999
Blank → 8,888
No answer → 8,888

ISCO3D
008-codif_ISCO3D_ALL_1994-2.qbe
999 → 9,999
010 ou 011 → 000
Blank → 8,888
299 → 200
329 → 320
729 → 720
799 → 700
899 → 800

NACE2D
010-codif_NACE2D_ALL_1994-2.qbe
00 → 9,999
Blank → 8,888

NACE3D
011-codif_NACE3D_ALL_1994-2.qbe
000 → 9,999
X or blank → 8,888
No answer → 8,888

WSTATUT
012-codif_WSTATUT_ALL_1994-2.qbe
Blank → 8,888
1.Employed → 1
2.Unempl. → 2

3.Inactive	→	3		
Not appl.	→	9,999		
No answer	→	8,888		
MAINSTAT				
013-codif_MAINSTAT_ALL_1994-2.qbe				
Blank	→	8888		
9	→	9,999		
No answer	→	8,888		
WSTAT1Y				
014-codif_WSTAT1Y_ALL_1994-2.qbe				
Blank	→	8,888		
9	→	9,999		
No answer	→	8,888		
NACE1Y2D				
015-codif_NACE1Y2D_ALL_1994-2.qbe				
00	→	9,999		
Blank	→	8,888		
STAPRO				
016-codif_STAPRO_ALL_1994-2.qbe				
Blank	→	8,888		
9	→	9,999		
No Answer	→	8,888		
STAPRO1Y				
017-codif_STAPRO1Y_ALL_1994-2.qbe				
Blank	→	8,888		
9	→	9,999		
No Answer	→	8,888		
ISCEDYEAR				
018-codif_ISCEDYEAR_ALL_1994-2.qbe				
Blank	→	8,888	6 years	→ 6
No Answer	→	8,888	7 years	→ 7
0 year	→	0	8 years	→ 8
1 year	→	1	9 years	→ 9
2 years	→	2	10 years	→ 10
3 years	→	3	>11 years	→ "11+"
4 years	→	4		
5 years	→	5		
MSTATUS				
019-codif_MSTATUS_ALL_1994-2.qbe				
Blank	→	8,888		
9	→	9,999		
No Answer	→	8,888		
FTPT				
020-codif_FTPT_ALL_1994-2.qbe				
Blank	→	8,888		
9	→	9,999		
No Answer	→	8,888		
PERM				
021-codif_PERM_ALL_1994-2.qbe				
Blank	→	8,888		
9	→	9,999		
No Answer	→	8,888		
HOMEWK				
022-codif_HOMEWK_ALL_199-2.qbe				
Blank	→	8,888		
9	→	9,999		
No Answer	→	8,888		

EDUC4WN

023-codif_EDUC4WN_ALL_1994-2.qbe

Blank → 8,888

9 → 9,999

No Answer → 8,888

EDUCLEVE

024-codif_EDUCLEVE_ALL_1994-2.qbe

Blank → 8,888

9 → 9,999

No answer → 8,888

PURP4WN

025-codif_PURP4WN_ALL_1994-2.qbe

Blank → 8,888

9 → 9,999

No answer → 8,888

NATPART

026-codif_NATPART_ALL_1994-2.qbe

Blank → 8,888

9 → 9,999

TOW

027-recodif_TOW_ALL_1994-2.qbe

<12 months → T1

12-23 months → T2

24-59 & 60+ months → T3

No answer → 8,888

ISCO3D

009-calcsum_ISCO3D_ALL_1994-2.qbe

Calc sum as pop

On obtient ainsi une table :**1994.db = 638 608 enregistrements**

Pour les besoins du projet STILE, il est indispensable de procéder à une recodification des champs "NACE2D" et "AGE".

Afin de ne pas travailler sur les tables originales, on réalise une copie de la table 1994.db.

Copie de "1994.db" en "STILE_1994.db"**NACE_GROUP**

028-codif_NACE_GROUP_STILE_1994.qbe

01-14 → "AGRI"

15-29/31/34-37 → "N-IM"

30/32/33 → "IM"

40-45 → "UTIL"

50-63/65-71/73-99 → "N-IS"

64/72 → "IS"

G1-G9 → 8,888

NACE_GROUP

029-codif_NACE_GROUP_STILE_1994.qbe

01-14 → "AGRI"

15-29/31/34-37 → "N-IM"

30/32/33 → "IM"

40-45 → "UTIL"

50-63/65-71/73-99 → "N-IS"

64/72 → "IS"

G1-G9 → 8,888

AGE

030-codif_AGE_STILE_1994.qbe

A1 → 00-14

A2+A3 → 15-24

A4+A5 → 25-34

A6+A7 → 35-44

A8+A9+A10+A11 → 45-64

A12 → 65+

NACE_GROUP

031-calcsun_NACE_GROUP_STILE_1994.qbe

Calc sum as pop

Au final, on dispose d'une table :

STILE_1994.db = 542 763 enregistrements.

Après recodification

- **ISCO3D & NACE2D**

Pour ISCO3D=(9,999 ou 8,888) et NACE2D=(not 9,999, not 8,888), le résultat ne doit pas être égal à 9,999 (exception pour ISCO3D=9,999 & NACE2D=9,999)

Pour ISCO3D=(not 9,999, not 8,888) et NACE2D=(9,999 ou 8,888), le résultat ne doit pas être égal à 9,999 (exception pour ISCO3D=9,999 & NACE2D=9,999)

Récapitulatif

CHAMPS	CONTENU											
COUNTRY	BE	DK	DE	GR	ES	FR	IE	IT	LU	NL	PT	UK
YEAR	1994											
REGIONOR	Codes régions, niveau NUTS2											
SEX	Male Female											
AGE	A1	00-14		A7	40-44							
	A2	15-19		A8	45-49							
	A3	20-24		A9	50-54							
	A4	25-29		A10	55-59							
	A5	30-34		A11	60-64							
	A6	35-35		A12	65+							
ISCED1D	1	Low		8,888	No answer							
	2	Medium		9,999	Not applicable							
	3	High										
ISCED2001	1	Low		360	High—6							
	2	Medium		8,888	No answer							
		High—5a		9,999	Not applicable							
		High—5b										
ISCO3D	Codes 3 chiffres correspondant à la classification ISCO											
NACE2D	Codes 2 chiffres correspondant à la classification NACE Rév. 1.											
NACE3D	Codes 3 chiffres correspondant à la classification NACE Rév. 1.											
WSTATUT	1	Employed		9,999	Not applicable							
	2	Unemployed										
	3	Inactive										
MAINSTAT	1	Employed		6	In compulsory military service							
	2	Unemployed		7	Fulfilling domestic tasks							
	3	Student, pupil, ...		8	Other inactive person							
	4	In retirement		9,999	Not applicable							
	5	Permanently disabled		8,888	No answer							
WSTAT1Y	1	Employed		6	In compulsory military service							
	2	Unemployed		7	Fulfilling domestic tasks							
	3	Student, pupil...		8	Other inactive person							
	4	In retirement		9,999	Not applicable							
	5	Permanently disabled		8,888	No answer							

NACE1Y2D	Codes 2 chiffres correspondant à la classification NACE Rév. 1. (année précédente)			
STAPRO	1	Self employed with employees4	Family worker	
	2	Self employees without employees	9,999	Not applicable
	3	Employee	8,888	No Answer
STAPRO1Y	1	Self employed with employees4	Family worker	
	2	Self employees without employees	9,999	Not applicable
	3	Employee	8,888	No Answer
ISCEDYEAR	0 year, 1 year, 2 years, 3 years, 4 years, 5years, 6 years, 7 years, 8 years, 9 years, 10 years, 11+ years, No answer			
MSTATUS	1	Single	4	Divorced or legally separated
	2	Married		
		Widowed		
FTPT	1	Full-time	6	Part-time because other reasons
	2	Part-time because school	7	Part-time but giving no reason
	3	Part-time because ill or disability	8	Looking after children or incapacitated adults
	4	Part-time because no full-time job	8,888	No answer
	5	Part-time because not want full-time job	9,999	Not applicable
PERM	1	Permanent job	5	No reason given
	2	Period of training	6	Probationary job
	3	Could not find permanent job	9,999	Not applicable
	4	Did not want permanent job	8,888	No answer
HOMEWK		Person usually works at home	9,999	Not Applicable
		Person sometimes works at home	8,888	No Answer
		Person never works at home		
EDUC4WN		Received no education or training	8,888	No Answer
		Received some education or training		
	9	Not Applicable		
EDUCLEVE		General Education		
	1	Primary or lower secondary (ISCED 1-2)		
		Upper secondary or post-secondary not tertiary (ISCED3-4)		
		Pre-vocational or vocational education or training		
		Lower secondary (ISCED 2)		
		Upper secondary or post-secondary not tertiary (ISCED 3-4)		
		Tertiary education		
		ISCED 5a		
		ISCED 5b		
		ISCED 6		
		Training that is not allocated to the ISCED classification		
	9,999	Not Applicable		
	8,888	No Answer		

PURP4WN	Initial education or training aiming at getting the skills for a job (except under a specific employment measure) Continuing vocational training (except under a specific employment measure) 5 Training under a specific employment measure General interest 9,999 Not Applicable 8,888 No Answer			
NATPART	1	Direct participation	9,999	Not applicable
	2	Indirect participation	8,888	No answer
TOW	T1	<12 months	T3	>24
	T2	12-23 months	8,888	No answer
AGE	00-14, 15-24, 25-34, 35-44, 45-64, 65+			
NACE_GROUP	01-14:	"AGRI"	AGRICULTURE	Agriculture & Pêche & Industries extractives
	15-29/31/34-37:	"N_IM"	Non ICT_Manufact.	Industries manufacturières (hors secteur télécommunications)
	30/32/33:	"IM"	ICT_Manufact.	Industries manufacturières (secteur télécommunications)
	40-45:	"UTIL"	UTILITIES	Electricité, gaz, construction
	50-63/65-71/73-99:	"N_IS"	Non ICT_Services	Services (hors secteur télécommunications)
	64/72:	"IS"	ICT_Services	Services (secteur télécommunications)
NACE1Y_GROUP	01-14:	"AGRI"	AGRICULTURE	Agriculture & Pêche & Industries extractives
	15-29/31/34-37:	"N_IM"	Non ICT_Manufact.	Industries manufacturières (hors secteur télécommunications)
	30/32/33:	"IM"	ICT_Manufact.	Industries manufacturières (secteur télécommunications)
	40-45:	"UTIL"	UTILITIES	Electricité, gaz, construction
	50-63/65-71/73-99:	"N_IS"	Non ICT_Services	Services (hors secteur télécommunications)
	64/72:	"IS"	ICT_Services	Services (secteur télécommunications)
RELIAB	"a", "b"			

Sources and most important variables of the Belgian Datawarehouse Labour Market Data

1. The sources

The majority of the data in the Datawarehouse originate from the participating social security institutions: RIZIV, RKW, RSVZ, RSZ, RSZPPO and RVA. For each person with a State's Register Number the individual's characteristics are provided by the Nationaal Instituut voor de Statistiek ('National Institute for Statistics', in short NIS), on the basis of the data which the NIS obtains from the State's Register each year. For the persons who do not have a State's Register Number, the State's Register Bis or Ter is used.

Below a brief description is provided of the way each institution collects the various data.

- *Nationaal Instituut voor de Statistiek ('National Institute for Statistics'): Data of the State's Register*

The NIS produces annual statistics about the demographic condition and the evolution of the population during that year. For this the NIS obtains an annual copy with a number of data from the State's Register, the body which centralises the data about the natural persons in the Register of births, deaths and marriages of the Belgian municipalities. So the municipal registers constitute the actual source of the personal data.

In the Datawarehouse the personal characteristics (address, age, sex) of persons with a State's Register Number are completed on the basis of the copy which the NIS obtains from the State's Register annually.
- *Kruispuntbank van de Sociale Zekerheid ('Crossroads Bank of the Social Security'): Data from the State's Register Bis and Ter*

The personal characteristics for the Bis and Ter Numbers are taken from the State's Register Bis. Included in the State's Register Bis are the data of persons who are known to the Belgian social security institutions, but who are not registered in a Belgian municipality (and who therefore remain unknown to the State's Register). It is especially frontier workers and persons who no longer live at their official address and who are neither registered with another municipality. The State's Register Bis is less complete than the State's Register and the data are less reliable. Finally, the State's Register Ter comprises the persons whose crucial data are missing.
- *Rijksinstituut Ziekte- en Invaliditeitsverzekering: GRI data ('State Institute Health- and Disablement Insurance')*

The RIZIV manages the file with the data about all persons who have submitted an application before the Geneeskundige Raad voor Invaliditeit ('Medical Council for Disablement', or GRI). An application for recognition by the GRI can be filed from the moment that a person has been unable to work for one year. During the first year of inability to work (the primary disability) the person remains registered with the RSZ, RSZPPO or the RSVZ.

For the moment the GRI file is not yet fully integrated in the Datawarehouse, since an INSZ (unique id number Social Security) has not been retrieved for all persons.

- *Rijksinstituut voor de Sociale Verzekering van Zelfstandigen: ARZA data ('State Institute for the Social Insurance of Self-Employed')*
The RSVZ collects the data, which it is provided by the social insurance companies in a data file, ARZA. It covers the data on December 31st of each year and comprises all those required to be insured who in a particular year were insured for at least one quarter. Required to be insured are the persons whose social status is self-employed because they hold an occupation as self-employed or helper, and persons who are granted equal status on the basis of the law (due to disease or invalidity, studies or articles of apprenticeship or continued insurance).
- *Rijksdienst voor de Sociale Zekerheid: LATG data ('National Agency for the Social Security')*
Every quarter the RSZ collects the data about the staff files (personnel register) declared by the employers in the wage- and working hours database. The observations are limited to the employees who need to be declared to the RSZ and therefore does not cover the employees working for local authorities (see RSZPPO) and the see(wo)men in the merchant navy (whose data are collected by HVKZ⁷ and who thus are not included in the Datawarehouse).
The LATG data are centralised data, which means that there are no data about the organisations,⁸ but that the employer is the point of departure. If the employer possesses several plants and/or practises more than one activity, the geographical location of the principal plant and/or the primary activity is used.
- *Rijksdienst voor Sociale Zekerheid voor Plaatselijke en Provinciale Overheden: STATPLUS data ('National Agency for Social Security for Local and Provincial Authorities')*
The employees employed by local authorities (municipalities and provinces) are collected by the RSZPPO in a database called STATPLUS.
- *Rijksdienst voor Kinderbijslag voor Werknemers ('National Agency for Child Benefit for Employees')*
The RKW has at its disposal data relating to the child benefits that are paid to employees. The basic unit of these statistics is every child for whom child benefit is paid.
- *Rijksdienst voor Arbeidsvoorziening - Paydays*
The RVA has at its disposal the monthly data regarding the payments, which are at the expense of the unemployment budget. These data cover both those fully entitled to benefits and other categories of benefit recipients: the early retired, the persons taking a career break, the part-time employed who receive an income guarantee allowance, the persons participating in employment programmes.

2. The most important variables in the Datawarehouse

- Region/District/Municipality/Statistical sector
The statistical sectors (at neighbourhood level) are included for all municipalities, which on December 31st, 1998 count at least 30,000 inhabitants, the provincial capitals and the 19 municipalities in Brussels.
- Country: Germany, France, Luxembourg, the Netherlands, other countries, unknown. Persons who live abroad are included in a separate table.

⁷ Hulp- en Voorzorgskas voor Zeevarenden ('Assistance and Precautionary Fund for Seamen').

⁸ For employers who possess several separate establishments or who practise several activities, the RSZ counts each establishment and each activity as one separate organisation. Different establishments with an identical activity lying within the same municipalities, count as one. The characteristics concerning these establishments are presented in the decentralised statistics.

- Sex.
- Age category.
- Socio-economic position: nomenclature of socio-economic positions.
- Size of enterprise
For persons who combine several jobs the size of the enterprise is given of the employer for whom the primary job is practised.
- Activity sector: Nace-Bel-code (5 digits) or RSVZ occupational code
For persons who combine several jobs the activity sector/occupational code of the employer, for whom the most important job is practised, is stated.
- Employment regime: full-time, part-time, special, unknown
The employment regime is only applicable to wage earners. Persons who are *self-employed* are attributed the code RSVZ. For persons who hold several paid jobs the employment regime of the main job is taken into account, unless stated otherwise.
- Percentage category part-time labour
The part-time percentages are regrouped in the table into 9 classes: 0-10%, 11-30%, 31-45%, 46-55%, 56-65%, 66-75%, 76-85%, 86-95%, +95%. Unless mentioned otherwise it is the percentage of part-time labour of the main job that is looked at when people have more than one employment.
- Accumulated percentage class part-time labour
The part-time percentages are regrouped in the table into 9 classes: 0-10%, 11-30%, 31-45%, 46-55%, 56-65%, 66-75%, 76-85%, 86-95%, +95%.
- Reduced SZ contribution ('Social Security contribution')
Included are only the target group-oriented reductions of SZ contribution or those advancing the redistribution of labour. For persons who combine several employments the reduction of SZ contribution of the principal job is stated.
- Number of jobs.
- Number of jobs for one and the same employer.
- Wages: each gross pay (without prejudice to the taxes), discharge and other contributions not included.

Bibliography

- Åkerblom (1999), *Mobility of highly qualified manpower: a feasibility study on the possibilities to construct internationally comparable indicators*, report for Camire, Estadística y Analisis, S.L, Eurostat and the OECD, Eurostat, Luxembourg.
- Åkerblom & Virtaharju (2001), *Measuring mobility: some methodological issues*, Statistics Finland.
- Colecchia A. (2001), *Reviewing the ICT sector definition: issues for discussion*, Voorburg group on services statistics.
- Communication from the Commission to the Council (2001), *New European labour markets, open to all, with access for all*, COM(2001)116 final, Brussels, 28 February 2001.
- Communication from the Commission to the Council, the European Parliament, the Economic and Social Committee and the Committee of the Regions (2002), *Commission's action plan for skills and mobility*, COM(2002)72, Brussels, 18 February 2002.
- Eurostat (1996), *NACE Rev. 1 - Statistical classification of economic activities in the European Community*, Eurostat, Luxembourg.
- Eurostat (1998), *The European Union Labour Force Survey - Methods and definitions*, Eurostat, Luxembourg.
- Eurostat (2001), *The European Union Labour Force Survey - Methods and definitions*, Eurostat, Luxembourg.
- ICT skills monitoring group (2002), *E-Business and ICT skills in Europe*, available at <http://europa.eu.int/comm/enterprise/ict/policy/ict-skills/wshop/synthesis-report-v1.pdf>.
- ILO (1988), *International standard classification of occupation*, International Labour Organisation, Geneva.
- NIS Belgium (2000), *Doorlopende enquête naar arbeidskrachten, Individuele vragenlijst*, NIS, Brussel.
- OECD (2000), *Measuring the ICT sector*, OECD, Paris.
- OECD and European Commission (1995), *The measurement of scientific and technological activities: manual on the measurement of human resources devoted to S&T*, Canberra Manual, OCDE/GD(95)77, OECD, Paris.
- Presidency Conclusions, Barcelona European Council, 15 and 16 March 2002.
- Statistics Denmark, Statistics Finland, Statistics Iceland, Statistics Norway, Statistics Sweden (2000), *The ICT sector in the Nordic countries*, Statistics Denmark, Copenhagen.
- UNESCO (1976), *International standard classification of education*, UNESCO, Paris.
- UNESCO (1997), *International standard classification of education*, UNESCO, Paris.