



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Statistical Office FSO

Establishing a Recommended Practices Manual for Editing and Imputation: The EDIMBUS Project

Daniel Kilchmann

16.11.2006 Swiss Statistics Meeting, Lugano





Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Statistical Office FSO

Contents

- Introduction
- The EDIMBUS project
 - Aim
 - Organisation
 - Work packages of the project
 - State of the project





Introduction

A fictitious example of business survey data:

Id	Number of employees	Research activities	Turnover (in 1'000 Fr.)	Research costs (in 1'000 Fr.)	...
...
9834327	23	N	4650000		...
9834	44	Y	9000	145	...
42342	125	Y		528	...
675411					...
345645	2		-862	90	...
82726	81	N	17500	93	...
...





Introduction

- Raw survey data of official statistics have usually missing and inconsistent data.
- Such data may lead to inconsistent and biased results.
- Analysis of data with missing values are cumbersome, especially for time series.
- A huge amount of methods for the treatment of missing and inconsistent data was developed during the last 30 years.
- All methods have advantages and disadvantages. A general evaluation of their performances and impact is difficult.





Introduction

- What is editing? No standard definition.
- The English word 'editing' means control and change.
- **Editing** is used here exclusively in the sense of controlling data.
- This is the only definition which ensures the management of the editing and imputation (E&I) process and the evaluation of the incoming data quality which are the main objectives of E&I.





Introduction

- What is imputation? No standard definition but usually:
 - **Imputation** stays for setting a value where before there was none or a value one would like to replace.
- Sometimes, imputation means setting a value where before there was none and a replacement is called correction.
- However, ‘correction’ should **only** be used if the imputed value comes from a callback (new original information) because ‘correction’ may suggest that the ‘true’ value could be found!
- Editing and Imputation form a process which is built up by applications of several methods and contains usually loops.





The EDIMBUS project: aim

- Harmonisation of European statistics is one of the major objectives of the European Statistical System.
- The Leadership Expert Group on Quality recommended the development of recommended practices manuals (RPMs) for statistics production in order to support survey managers.
- The RPM is a set of proven good statistical methods and recommendations for their application.
- The EDIMBUS project focuses on the development of the RPM for data **ED**iting and **IM**putation (E&I) in cross sectional **BU**siness **S**urveys.





The EDIMBUS project: aim

- The E&I process is methodologically not well structured, not well researched and lacks of harmonisation between National Statistical Institutes (NSIs) and also inside most NSIs.
- The process of editing and imputation is one of the most time and resources consuming survey processes (~40%).

Recommendations

- Selection of methods depending on the characteristics of the survey and the data
- Measuring the performance of E&I methods
- Design and monitoring of the E&I process
- Documentation and archiving of the E&I process





The EDIMBUS project: organisation

Project partners

- ISTAT Italy (coordinator)
 - O. Luzi, M. Di Zio, U. Guarnera, A. Manzari
- CBS Netherlands
 - T. De Waal, J. Pannekoek, C. Templeman, J. Hoogland
- SFISO Switzerland
 - B. Hulliger, D. Kilchmann

Partially fund by EUROSTAT: 40'000 €. Bilateral II: Switzerland renounces grants from EUROSTAT. Total costs ~140'000 €, ~300 Pd.

Period of the EDIMBUS project: 1.1.2006 – 30.6.2007





The EDIMBUS project: Work packages (WP)

1. **State-of-the-art survey** in the EDIMBUS field: finished
2. Development of a **draft version** of the RPM: by 1.07
3. **Validation** of the draft RPM in different national settings by referee reports: 11.06 – 4.07
4. Revision of the draft RPM and production of the **final version** of the RPM: 3.07 – **6.07**

EUROSTAT is responsible for dissemination of the RPM and training in the field of EDIMBUS.



WP 1. State-of-the-art survey

- Questionnaire mailed to the NSIs of the EDIMBUS project, the European NSIs, the Australian Bureau of Statistics, Statistics Canada and the US Bureau of the Census.
- 63 questionnaires returned, ~80% of the 29 countries involved.
- Canada, Finland, Norway and UK have developed some kind of guidelines or manuals for E&I in business surveys.
 - No general framework.
 - E&I is not described as a process, no recommendations about assessing data quality (indicators), management (problems?), documentation and archiving.
- Canada has also general quality guidelines -> too general for the purpose of an RPM on E&I.





WP 1. State-of-the-art survey

- Few countries have an established procedure for obtaining approval for the design of the E&I process of a survey.
-> no clear idea how an E&I process should look like.
- High heterogeneity in definitions and concepts not only at international, but also at national level:
 - editing, 'validation', imputation.
 - where starts and where ends E&I in the survey process.
 - minimal changes to achieve statistical accuracy vs. accuracy at unit level.





WP 1. State-of-the-art survey

- While methods and practices for particular problems are more or less known, the communication about E&I seems still under-developed (documentation, quality indicators).
- It is quite common that E&I methods are only tested during their application in the E&I process. -> No design and no testing, i.e. probably no management of the E&I process.
- Most of these problems are also reflected in the different views of the EDIMBUS project partners.





Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Statistical Office FSO

State of the project

- The project team met at the end of October in Rom and discussed the first draft version of the RPM.
- Difficulties to structure the manual because of different views.
- E.g. agreement about the definition of a general process flow of editing and imputation is difficult and leads to a different structuring of the manual -> compromise.

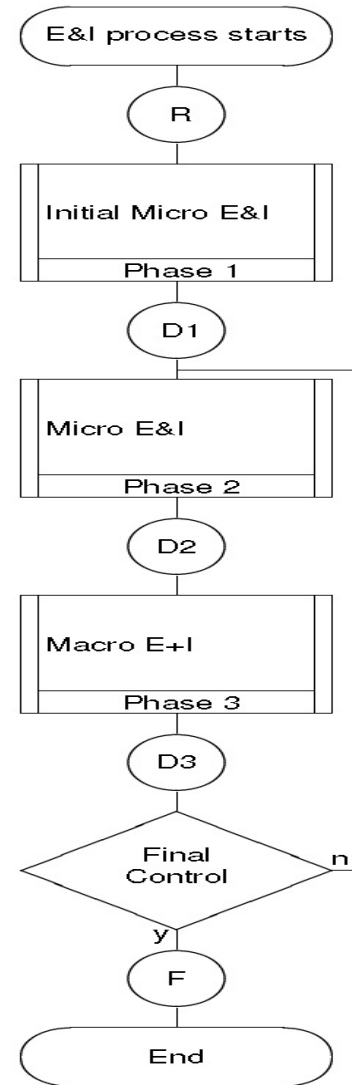




Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Statistical Office FSO





State of the project

- Appraisal of advantages and disadvantages of manual revision is difficult.
 - Manual revision often defines the gold standard without objective criteria -> automatic procedures are only used as a last resort or with the aim to save costs.
 - Manual revision can also be seen as a bad method
 - subjective method
 - not repeatable
 - mix of editing and imputation -> compromises the assessment of data quality, the overall management of the E&I process and documentation.





State of the project

Future action

- Second draft version of the RPM by Jan. 07
- External and internal refereeing.
- Final revision of the RPM of EDIMBUS will be submitted by 30.6.2007
- Further Information about the RPM of EDIMBUS in the public area of <http://edimbus.istat.it>

