Preface

This volume includes papers that have been presented at the DIECOFIS Workshop "Data Integration and Record Matching" that took place in Vienna, Austria, November, 13-14, 2003. DIECOFIS (Development of a System of Indicators on Competitiveness and Fiscal Impact on Enterprises Performance, http://www.istat.it/diecofis) is an EU-funded international research project that aims at the development of a system of indicators on enterprise competitiveness and fiscal impact on enterprise performance. The project was sponsored within FP5-IST and coordinated by the Italian national statistical institute ISTAT.

A core element of DIECOFIS is the creation of a multi-source database, consisting of real or "virtual" businesses. On this basis, micro-simulations can be run and used for assessing the effects that public policy may have on companies and on their performance. Hence, dataset integration was an issue of main emphasis, and statistical integration methodology and quality indicators for the assessment of different approaches played a major role. Two broad classes of integration procedures were distinguished: Micro procedures integrating datasets at record level by combining records representing the same (or a similar) real-world entity in different datasets, and macro procedures where the main interest is on aggregates of the integrated data. The DIECOFIS partners from national statistical offices and research institutions have gained experience both in methodological questions concerning database integration and in the application of such techniques. Theoretical aspects primarily include statistical issues, such as the assessment of different approaches and statistical indicators of multi-source database quality.

Integration of data files is not only an issue for the project DIECOFIS. It is of major interest for official statistical agencies as a means of using available information more efficiently. By so doing, the "value added" that can be extracted from the existing stock of information is greatly augmented. Various national statistical organizations apply corresponding techniques and actively work on their enhancement; some have developed their own software tools. Academics and statistical journals publish on theoretical issues of data set integration methods such as probabilistic record matching, statistical matching or data fusion.

Given this background, the workshop "Data Integration and Record Matching" was organized in order to bring together DIECOFIS partners with practitioners, researchers, and developers from leading institutions. The participants had the opportunity to learn from their experiences, particularly in applications, and to exchange their views on new ideas and developments. The proceedings contain three sections: (1) *Applications* are considered in the papers by Martha Fair, Alois Haslinger, Frank Linder, and Felix Ritchie; these papers report on experiences with data integration made in national statistical agencies of Canada, Austria, The Netherlands, and UK, respectively. (2) Various *methodological issues* are subjects of papers by Leicester Gill, Rainer Schnell, Michaela Denk and Peter Hackl, Susanne Raessler, and Karl A. Froeschl. (3) *The case DIECOFIS* contains papers that report on the specific role that data integration plays within the project DIECOFIS. The editors think that the collection of papers is indeed of interest for practitioners, researchers and developers in the area.

4

The editors are most grateful to the contributors for their high quality papers and to the referees for their committed and timely work. A list of both can be found at the end of the volume. Furthermore, the editors would like to thank Ewald Kutzenberger, Director-General of Statistics Austria, for hosting the workshop, Elisabeth Sachs, International Relations/Statistics Austria, for organisational support, and Sonja Steffek, Vienna University of Economics and Business Administration, for secretarial aid.

M. Denk, K. Froeschl, P. Hackl, N. Rainer   Vienna, May 2004