

**UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

Work Session on Statistical Data Editing
(Paris, France, 28-30 April 2014)

Report of the 2014 Work Session on Statistical Data Editing

Prepared by the UNECE secretariat

1. The Work Session on Statistical Data Editing was held in Paris, France, from 28 to 30 April 2014 at the invitation of National Institute of Statistics and Economic Studies (INSEE). It was attended by participants from: Austria, Canada, Denmark, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Mexico, Netherlands, Norway, Republic of Moldova, Russian Federation, Slovenia, Spain, Sweden, Switzerland, United Kingdom, and the United States of America. The European Commission was represented by Eurostat. Representatives from the following international organizations also attended: Eurasian Economic Commission (EURASEC), International Labour Organization (ILO), Organisation for Economic Co-operation and Development (OECD).
2. The agenda contained the following substantive topics:
 - Getting the support of all people when implementing data editing;
 - Selective editing / macro editing;
 - New & emerging methods;
 - Editing of Census and social data;
 - International collaboration and software & tools;
3. Mr. Steven Vale (UNECE) opened the meeting by stressing the changes in official statistics, with much interest in the official statistical community on concepts such as modernisation, big data and data revolution. Data editing is at the heart of many of these changes, as there remains a need to ensure data are fit for use, however data editing needs to evolve, and this would be a common theme throughout the Work Session.
4. Ms. Fabienne Rosenwald, Director of Business Statistics, welcomed participants on behalf of INSEE. She spoke about the need for standardisation of editing processes across statistical domains, and the need to respond to new data sources. She wished the participants fruitful discussions during the Work Session.
5. Mr. Claude Poirier (Canada) was elected Chairman of the Work Session. He stressed quality of the papers, and that the whole statistical community would benefit from that. He outlined the need for a future work programme that meets the needs of participating organisations as far as possible. He invited volunteers to propose a list of topics for future work. Mr. Emmanuel Gros (France), Ms. Sarah Giessing (Germany), Mr. Gergely Horvath (Hungary), Mr. Marco Di Zio (Italy) and Mr. Sander Scholtus (Netherlands) agreed to take on this role.
6. The following persons participated in the Organizing Committee and acted as Discussants/Session Organizers: Topic (iii) – Messrs. Li-Chun Zhang (Norway) and Philippe Brion (France); Topic (i) – Ms. Orietta Luzi (Italy), Messrs. Rudi Seljak (Slovenia) and Pedro Revilla (Spain); Topic (ii) – Ms. Maria Garcia and Mr. Jeroen Pannekoek (Netherlands); Topic (iv) – Mr. Daniel Kilchmann and Ms. Felipa Zabala (New Zealand); Topic (v) - Messrs. Claude Poirier (Canada), Gergely Horváth (Hungary) and Pedro Revilla (Spain).

7. The proposals for future work included:
- Selective and macro-editing
 - Quality assessment of selective editing
 - Unification of selective editing approaches
 - Selective editing for classification variables
 - Treatment of issues relating to statistical units in business statistics
 - Highlight complementarity between selective editing and macro-editing
 - Case studies
 - Managing and supporting changes related to editing and imputation
 - Managing/monitoring the editing and imputation process
 - Managing changes to the editing and imputation process
 - Managing cultural changes
 - Getting new methods implemented
 - Standardization/harmonization
 - Software tools and international collaboration
 - Software development
 - Collaboration with information technology specialists
 - SAS vs R
 - Standardization/harmonization on an international level
 - Evaluation and feedback
 - Indicators for editing and imputation
 - Using information from editing and imputation to improve the statistical process
 - Estimation uncertainty due to editing and imputation
 - Editing and imputation with non-traditional data sources
 - Editing and imputation for administrative data
 - Editing and imputation in combination with Big Data
 - Editing and imputation with multi-source data
 - Editing during electronic data collection
 - New and emerging methods
 - Methods related to editing and imputation that are either new or emerging (or both)
 - Innovative applications of existing methods
 - Development of a common, generic process framework for statistical data editing. This could be done by a task team under the High-Level Group for the Modernisation of Statistical Production and Services, and presented at the next Work Session.
8. Mr. Gergely Horvath, on behalf of the Hungarian Central Statistical Office, offered to host the next Work Session on Statistical Data Editing in Budapest in Autumn 2015.

FURTHER INFORMATION

9. The conclusions reached during the discussion of the substantive items of the agenda are contained in the Annex. All background documents and presentations for the meeting are available on the website of the UNECE Statistical Division (<http://www.unece.org/stats/documents/2014.04.sde.html>).

ADOPTION OF THE REPORT

10. The participants adopted the present report before the Work Session adjourned.
11. The Chairman thanked INSEE for the excellent facilities, the Organising Committee for preparing the content of the Work Session, the participants and paper authors for their contributions, and the UNECE Secretariat for their support.

Annex

Summary of the Main Conclusions Reached at the Work Session

Topic III. Getting the support of all people when implementing data editing

Session Organizers/Discussant: Li-Chun Zhang (Norway) and Philippe Brion (France)

Documentation: Papers by Canada, Sweden, Finland, Norway, France and New Zealand (presented by Claude Poirier)

1. The papers presented under this topic described the way how different national statistical organisations have implemented new and standardised data editing systems. This kind of implementation often raises broader technical, infrastructural and even organizational issues that seem to lie beyond the department responsible for data editing in a given statistical domain.
2. Statistics Canada is currently undertaking a major integration project for its Business Statistics surveys, the Integrated Business Statistics Program (IBSP). The IBSP will provide a common processing framework, a vertically integrated data editing approach and common methodologies for the various business surveys conducted at Statistics Canada. From 2014 to 2017, around 150 surveys will migrate into this new integrated and harmonized framework. The presentation gave a brief overview of the key aspects of the IBSP program and its new vertically integrated data editing approach developed and presented the elements that have contributed to get the buy-in of the partners involved.
3. Sweden presented on applying process indicators to monitor editing. They have tried to establish usable process indicators and have computed them for a selected survey. Indicators can primarily be used to evaluate the editing process itself and secondarily to evaluate the data collection process. Also, they would like to use the indicators as a quality measure of the survey, concerning measurement errors. The goal is to decide on a set of indicators that can be produced regularly and used in most surveys, in conjunction with the SELEKT tool.
4. Finland presented lessons learned from their project on standardisation of production processes for statistics. Following the recommendations of the project, the implementation of selective editing was set as a strategic goal of Statistics Finland. Training of the staff to use the new methods and tools may help to clarify the benefits from and enhance engagement into the renewal work.
5. Norway presented preliminary experiences from implementing a new editing system. With the potential benefits of a more standardized editing framework, significantly more efficient working procedures and improved quality, the concept was well received at management level. The barriers to change seemed to stem from the work culture rather than any system-based barrier. To achieve a change in the approach to editing it was recognized that editing staff needed to be involved and engaged in the process, some complimentary approaches are needed instead of a singular approach using courses and workshops.
6. The presentation of France described how a new approach to data editing for structural business statistics has been implemented. The new procedures helped INSEE assess a list of businesses to control according to their relative contribution to statistics of main interest. The concrete implementation of these methods needed some time, some adjustments were necessary during the first campaigns, and led, in the beginning, to a lack of confidence from the clerks. Therefore annual process reviews have been put into place on different parts of the process, involving clerks and methodologists. These meetings have proved to be very useful, and helped to improve the system.
7. The presentation of New Zealand described their experiences on implementing more efficient data editing processes as part of a major business transformation programme. They focused on the role of leadership, changing organisational culture with respect to data editing, and on the lessons learned from their experience. Future work will include measurement of culture change.

8. To introduce the discussion, the session organisers identified four common themes:
 - Motivations for change – senior management requirements, budget cuts, standardisation etc.
 - Demonstration of process and results
 - Measuring and documenting benefits
 - Culture, knowledge, skills and training

9. Points raised in the general discussion included:
 - The importance of involving editing staff from the start of any change, and to focus criticisms on the old editing systems, and not the people responsible for them
 - Common systems should be versatile enough to adapt to new methods
 - How to develop the new skills needed by editing staff, in particular the greater analytical skills needed for selective editing approaches, particularly to correctly design and apply parameters
 - The extent to which training of editing staff should include consideration of methodological issues. This can help to build trust between the two groups.
 - Training on data editing concepts and definitions can help to facilitate communication
 - A proposal for international collaboration to develop a common framework for statistical data editing. This could cover how to design an editing process comprising components such as micro-editing and selective editing, and could build on the work done in the European Statistical System “Edimbus” project. This could be overseen by the High-Level Group for the Modernisation of Statistical Production and Services, or the Statistical Network
 - How to measure the costs and benefits of change? No solutions were proposed.
 - Moving to common editing approaches can be seen as trading flexibility associated with local solutions for the increased robustness of a generalised system. There is no reason to keep domain specific approaches, as the differences are mostly due to culture and history.
 - Standardisation should not be seen a “de-humanising” editing work
 - Visualisation techniques may help editing staff understand methodologies
 - The balance between micro and macro editing may be shifting, but it can be useful to calculate aggregates at several stages during the production process to help to identify editing issues.

Topic I. Selective editing / macro editing

Session Organizers/Discussant: Ms. Orietta Luzi (Italy), Mr. Rudi Seljak (Slovenia) and Pedro Revilla (Spain)

Documentation: Papers by Italy, Spain, United Kingdom, Sweden (2), France (2), United States and Japan.

10. Selective data editing approaches can significantly improve the efficiency of the data editing process. The topic covered new and innovative implementation strategies inside the statistical process, including practical implementation of local and global score functions, integration of selective editing/macro editing techniques within the statistical process, tools to support selective editing / macro editing, validation, and editing of unstructured data.

11. Italy described the selective editing procedure using SeleMix, a tool developed in house using the R programming language. It has been applied to Structural Business Surveys (SBS) data on investment. An assessment of the procedure was performed by analyzing a pre-defined subset of non-selected units. A result of the validation analysis is that it is difficult to select an observation with an erroneous investment which it is not atypical with respect to auxiliary information which is used to build the model, i.e. historical values.

12. The presentation from Spain explained how score functions can be obtained using an optimization approach. It introduced a theoretical framework to deal with the problem of selective editing, including the search for an adequate selection strategy as a generic optimization problem. Some case studies with data from Spanish surveys were shown.

13. United Kingdom presented the work to develop methods and a strategy for carrying out regular maintenance reviews of each of the ONS surveys using selective editing. The maintenance strategy also includes a process for dealing with any unforeseen data issues arising from selective editing. Cultural issues such as the need for cooperation with statisticians and constant training were also addressed.

14. Statistics Sweden reported on the implementation and use of selective data editing in eleven “big” surveys the last five years. A first prototype of a generic tool, SELEKT, was used in the Wage and Salary Structures Survey in 2008. As a result, Statistics Sweden has seen improved resource allocation and more efficient edits as a result of review of the existing edits rules. Editing staff consider the work more effective, interesting and less stressful. Late macro editing has moved towards early micro editing which implies that follow-ups can be done closer to the data delivery.

15. INSEE presented the application of selective editing techniques in the monthly production of short-term economic indicators. Four years ago, INSEE launched a large redesign of its short-term statistic production system. One of the main objectives of this project is to define and share approved methodologies. This work was mainly done in the GSBPM framework and, as far as statistical methodology is concerned, following the Edimbus manual recommendations.

16. In their second presentation, France explained the need for prioritization techniques to select the households that need to be investigated in order to reduce as far as possible the non-response bias and loss of precision due to the fall in response rates in such units. INSEE use R-indicators to analyse the representativeness of the sample during the survey.

17. Japan described their automatic data editing process, which is divided into two steps. The first step is error localization, which is performed using the SeleMix software. The second step is imputation, where errors are corrected by some multiple imputation methods. The aim in this project is to propose a way to automate part of the editing process for economic surveys. To attain this goal, they used as an example, data from the Economic Census for Business Activities.

18. The U.S. Bureau of Labor Statistics focused on unstructured and semi-structured text fields in survey data, which are typically underutilized in most processing procedures and systems. They highlighted approaches in text analysis that could be used to extract information from these fields. Accident reports from the U.S. Department of Labor, Occupational Safety and Health Administration were used to illustrate the concepts. Examples were given of how these text analysis tools could be used to assist humans in applying labels to the accident reports, to verify existing coding made by humans, and to edit data.

19. The second Swedish paper provided an argument for a model based approach in the context of selective editing of sample survey data, and a methodology to predict measurement errors in unedited units and their associated bias. This approach has been tested on data from the Swedish survey on establishments and their salary payments to employees.

20. The points raised by the discussants included:

- Theoretical formalizations of the selective editing paradigm are proposed, in a situation where no accepted theory has been developed yet
- The need for a common theoretical evaluation framework for quality assessment of selectively edited data. This requires further work at the international level.
- Issues related to the development and use of generalized tools for selective editing within statistical organisations, such as applicability, maintenance and optimization of the implemented methods and models
- New application contexts where selective / macro editing principles can contribute to most efficient statistical production processes

21. The general discussion included the following points:
- The costs and benefits of automatic editing of non-selected units. This choice can have a positive impact on secondary uses of the data. High potential re-use of data would counter-balance deviating from efficiency savings from using purely selective editing.
 - Probabilistic and deterministic approaches to assessment of selective editing can be combined, by examining a sample of units not selected under the selective editing model, e.g. based on a deterministic rule
 - The use of different constraints (e.g. quality or workload) to determine the selection
 - The quality of selective editing relies on the quality of the predictor. This could be an area for future work. However the quality of the predictor is less important for the most extreme errors.
 - It can be useful to take a step back and look at the potential issues with survey questions if there are significant editing issues
 - Different approaches to the treatment of outliers. The decision to treat them automatically or to re-contact the respondent can depend on the score.
 - Selective editing is most effective when true values can be established. In other cases there is a risk of introducing bias.
 - A comparison of software tools would be useful. This was done in the past, but needs updating. Such a comparison could identify the best features of each current tool, which would help to design a new standard tool.
 - The use of open-source software, including R packages. SELEKT, from Sweden, is a collection of SAS macros, which are freely available, but licenses are needed for the underlying software. SeleMix, from Italy, is based on R, and can be used to apply different methods.
 - It is better to think in terms of choosing the best methods rather than a particular tool

II. New & emerging methods

Session Organizers/Discussants: Yves Thibeau (USA, for Ms. Maria Garcia) and Jeroen Pannekoek (Netherlands)

Documentation: Papers by Italy, Hungary, Germany, United States Census Bureau, United States National Agriculture Statistics Service (NASS), Netherlands (2), and France.

22. This topic covered new and emerging methods for improving and/or optimizing the process of data editing and imputation. In this topic, presentations reported on the development of editing and imputation for new kinds of data sets, improving imputation by selecting better methods, detection of outlying, suspect or erroneous values, and evaluation methods/measures for editing and imputation.

23. Italy shared their experience with Structural Business Statistics (SBS) for small and medium enterprises. In recent years, the increasing availability of information from administrative sources has made it possible to use administrative data to improve the quality of the statistics. Administrative sources such as Financial Statement, Sector Studies Survey and Tax Return data have been used to build a microdata file composed of the main economic variables. The choice of producing a microdata file follows from the difficulty of providing coherent estimates at different level of aggregation.

24. The Hungarian Central Statistical Office (HCSO) shared their experience on presentation and development of outlier treatment. The presentation introduced the current practice, the methods used for outlier detection and the process of the workflow in the HCSO and discussed as well some methodological issues of the methods that are used now.

25. Germany presented a comparative study on two multiple imputation methods. They described the setup for the study, which was based on the complete data with simulated missing values. Criteria for evaluation of results were developed and different approaches for potential improvement were compared. The results of the investigation were presented and analyzed.

26. The U.S. Census Bureau described two procedures for imputing monthly earnings, a model-based procedure that makes use of sequential regression multiple imputation (SRMI), and a randomized hot-deck. They presented results of a simulation study designed for comparing the model-based approach to the randomized hot-deck, which showed that SRMI is a feasible alternative to hot deck and it has the potential to improve estimates.

27. U.S. NASS gave a brief outline of the two imputation methodologies used in the Agricultural Resource Management Survey (ARMS), and a comparison of the survey estimates using the two methodologies for two years of data. This survey was administered in three phases and is a multi-mode, dual frame survey. The survey provides an annual snapshot of the financial health of the farm sector and farm household finances, and it is the only source of information available for objective evaluation of many critical policy issues related to agriculture and the rural economy.

28. The Netherlands presented a paper on a generalisation of the Fellegi-Holt paradigm which is proposed to incorporate various complex editing operations in a natural way. In the Fellegi-Holt approach to automatic editing, the amendment of individual values plays a central role. On the other hand, human editors frequently perform more complex editing operations that involve simultaneous changes in several values. It was explained that generalisation may be used to increase the suitability of automatic editing in practice, and hence to improve the efficiency of data editing processes.

29. INSEE presented the project named Scanner Data Project, which is at the stage of methodological studies and experimentation. One of the goals of the project was to improve the accuracy of disaggregated indices. At the same time, INSEE must replace information gathered by the price collectors in the shops by an automatic process. The presentation dealt with some working tracks on data editing.

30. The second presentation of Netherlands discussed, by using applications to real data sets, the implementation of an overall automatic editing process. This involves the configuration of methods and tools, including the specification of the rule-sets, necessary to apply the editing functions. The application of indicators that measure the effects of each editing function and graphical displays of these indicators that allow for a concise review of the progress of the process were also discussed. This monitoring could provide the feedback necessary to further optimise the automatic data editing process.

31. Points raised in the general discussion included:

- How to deal with the results of multiple imputation, and whether to provide these to users? Volume could be an issue, and there may be little value to research users in having multiple values for the same observation.
- The extent to which outlier detection can be considered an integral part of statistical data editing
- When multiple sources are available, quality assessment may be easier, but if statistical sources are phased out, new approaches will be needed for quality assessment of other sources.
- In-depth knowledge of new sources is necessary to be able to understand changes, and their impact on editing
- Outliers may be errors, but may also be genuine, but unrepresentative of the population. However removing them when calculating estimates, or as donors for imputation, may increase biases.
- Emerging trends in statistical software, such as the use of SAS or R may have an impact on the ability to implement emerging methodologies
- How to identify (and count) operations in the context of optimal approaches to the Fellegi-Holt Paradigm
- Some of the selective editing issues relating to the use of administrative sources are also likely to be relevant in the context of Big Data, including the French example of scanner data.

IV. Editing of Census and social data

Session Organizers/Discussants: Mr. Daniel Kilchmann and Ms. Felipa Zabala (not present, New Zealand)

Documentation: Papers by Canada, US Census Bureau, University of Southampton (UK) / Universidade Federal do Rio Grande do Norte (Brazil), and Mexico

32. This topic focused on the methodological advances of editing and imputation (E&I) techniques applied to census or social data. The presentations discussed the results of applications of E&I for census or social data.

33. Canada presented an overview of the CANCEIS methodology and recent improvements that have been implemented to the software. The highlights of the E&I processing of the 2011 census were presented. Options being considered for the 2016 census E&I processing were described, which include combining more processes to improve coherence between variables and restricting the number of donors used for imputation.

34. Mr. Li-Chun Zhang presented a joint paper of the University of Southampton (UK) and the Universidade Federal do Rio Grande do Norte (Brazil), on estimation of the variance due to imputation in the 2011 UK Census. The paper presented a simple method for assessing the imputation variance that uses information retrieved from the CANCEIS output files. In this way the approach may be useful for other applications of the CANCEIS platform. The methodology has been implemented on the UK 2011 Census data. The feedback suggests that uncertainty in the Census-based population estimates due to imputation variance is perhaps negligible. This should serve to increase confidence for the users of the data.

35. Mexico shared experiences of automatic data editing in the Mexican Census, where efficiency gains led to more timely results. The methodology was described with particular emphasis on reviewing the editing strategy of the local offices of the Institute, whose maximum level of disaggregation was the municipality. The use of theoretical vectors methodology to automate data editing helped ensure the timely release of census results.

36. The U.S. Census Bureau presented their paper, which describes the characteristics of persons with missing race and Hispanic origin data, and then assesses the coverage of administrative records data for respondents who do not answer race and Hispanic origin questions in Census data. The presentation also discussed the distributional impact of using administrative records race and Hispanic origin data to complete missing responses in a decennial census or survey context.

37. Points raised in the general discussion included:

- The selection of donors for imputation in CANCEIS – this is based on the smallest distance score for the set of selected variables. It may be useful to treat outliers first, to reduce the risk of bias.
- When imputing with census data there are many potential donors, but this may not be the case for survey data, so different approaches or imputation methods may be needed
- The difference between CANCEIS and BANFF – BANFF is designed for business and financial data with linear edits, whereas CANCEIS is better suited to social data sets which tend to be larger and include categorical data as well as a very large number of within-person and between-person edits
- Whether mode effects resulting from the increase of web-based data collection should be taken into account when selecting donor records for imputation. More research is needed in this area.
- There should be a clear link between editing metrics and questionnaire design for future survey waves. Improved questionnaire design can reduce the amount of editing required.
- There are many challenges integrating data from different sources without common identifiers

Modernization Committee on Production and Methods under the High-level Group

38. Mr. Claude Poirier (Canada) and Mr. Steven Vale (UNECE) presented the work of the High-Level Group for the Modernisation of Statistical Production and Services, and the new Modernisation Committee on Production and Methods, and how these groups are encouraging greater interaction between specialisms such as methodology and information technology (IT). The meeting was asked for ideas on how to organize future events to maximize their value to participating organizations. Suggestions and comments included the following:

- Communication between specialisms could be improved, this may be best achieved through small multi-disciplinary groups working on specific issues
- Statistical organisations need more people with both methodology and IT skills
- Some more IT input to data editing meetings could be useful, but the IT perspective should not dominate
- It would be useful to have input on the possibilities offered by new IT tools
- Data editing is just one aspect of methodology. Any joint methodology and IT events should have a wider focus
- The current format of work sessions on statistical data editing is unique and remains relevant, so should be maintained
- The series of papers of the workshop give a unique overview of the field and are used all the time NSI's as a reference

V. International collaboration and software & tools

Session Organizers/Discussants: Mr. Claude Poirier (Canada), Gergely Horváth (Hungary) and Pedro Revilla (Spain).

Documentation: Papers by the Netherlands, Slovenia, Italy, Netherlands, Austria/Vienna University of Technology, Finland and New Zealand

39. This topic focused on portable solutions, collaboration initiatives targeting software development or reuse of tools, governing data editing open source tools, competing goals of local programs against international initiatives, modernization using Enterprise Architecture, international collaboration to solve practical problems and to share knowledge and solutions, identifying national needs, schedules and expectations, and moving national successes into an international perspective.

40. The Netherlands briefly described the ESSnet project Memobust (Methodology of Modern Business Statistics). It updated and expanded a previous handbook on methodology for business statistics, making use of many existing sources of information, such as the EDIMBUS Edit and Imputation Manual. The Memobust handbook and the description of the editing and imputation topics were discussed.

41. The Statistical Office of the Republic of Slovenia launched in 2011 a project aiming to upgrade existing solutions and build one global solution which would cover all parts of data processing, and which could easily be used for most statistical surveys. They presented the results achieved so far and described the plans for further developments. A part of the presentation was devoted to the consideration of what the introduction of such a general tool means for the strategic changes in the design of statistical production.

42. Italy described experiences related to the implementation of the CSPA (Common Statistical Production Architecture) specifications for an "Error Localization" service, realized in a Proof-of-Concept carried out within an international collaboration project. The service wraps functions that are offered by the R package "editrules" developed at Statistics Netherlands. The service has been deployed on the CORE platform used by Istat for process industrialization. They also showed how such a CSPA service could be combined with an error correction service in order to perform a full editing and imputation process.

43. The Netherlands presented a paper explaining how a wide range of data validation checks, including in-record, cross-record, and cross dataset checks can be analysed in a generalized way. Each check can be

assigned a boolean result, a severity measure (measuring the amount of discrepancy between a desired and the current score) and an impact measure (measuring how much data needs to change in order to be valid under a [set of] rules). All three measures can be derived directly from a general formulation of a data validation function.

44. The presentation of Austria described the application of the methods available in the package VIM and demonstrates the usage of the graphical user interface in VIMGUI. Special attention was also given to the new imputation functionality and the VIM integration of survey objects that are output from R's survey package.

45. Finland presented the SAS EG project that should implement selective editing methods for all suitable statistics in Statistics Finland. This project for editing is on-going, and there is no demo version available yet. The development of imputation processes is under way.

46. New Zealand presented the challenges of building a micro-economic platform (MEP) that includes a non-response imputation package and a non-sampling variance package. MEP uses both the Banff software for outlier detection and imputation, and the SEVANI (system for estimation of variance due to non-response and imputation) software to estimate sampling errors, taking into account the non-response in the data.

47. The following points were raised during the discussion:

- The introduction of a standard metadata-driven solution for all data processing can be seen as short-term pain for long-term gain. A balance is needed between simplicity and functionality, and subject matter staff will require training.
- The need to provide standardised metadata about business processes can help to improve understanding of those processes amongst staff
- Ensuring consistency of editing rules can be a challenge in metadata-driven systems. BANFF has a function to help with this.
- The Memobust handbook would benefit from on-going maintenance, and inputs from the wider data editing community. A wiki-based approach was proposed. Although the initial focus was business statistics, it would be relatively easy to extend it to cover other domains, such as social statistics.
- The use of the k-Nearest Neighbour algorithm proposed in the Austrian presentation might be processor-heavy for very large data sets with many potential donor records
- The need for tools to allow monitoring of the impact of editing and imputation during the different steps of the statistical production process
- There are several different standards for data validation syntaxes. Convergence on a single standard would facilitate software development.
- The extent to which subject-matter areas can take responsibility for configuring and setting parameters for data processing tools and systems. Some knowledge of programming is increasingly required.
- Both SAS and R now allow distributed or parallel processing to help speed up tasks that are very processor-heavy

* * * * *