Working Paper No. 13
ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint UNECE-EUROSTAT Work Session on Registers and Administrative Records for Social and Demographic Statistics
(Geneva, 9-11 December 2002)

Invited paper – SESSION 2

SYSTEMS BEING DEVELOPED TO COLLECT, PROCESS AND DISSEMINATE NEIGHBOURHOOD STATISTICS, INCLUDING THE PROVISION OF TOOLKITS FOR USERS

Submitted by the Office for National Statistics, United Kingdom[1]

[1] Paper presented by Liz Tadd, Programme Manager for the Neighbourhood Statistics System, ONS, UK

I. INTRODUCTION

1. At the June 2001 plenary session of the Conference of European Statisticians, a paper was presented[2] which described in some detail the need for small area statistics and in particular the demand from the UK Government for information to enable social exclusion in deprived neighbourhoods to be tackled. The paper also described the UK's initial plans in response to this need for better information for small geographic areas: the development of Neighbourhood Statistics. This system is making available a range of data for local areas which can be analysed or downloaded, free of charge, by anyone with access to the internet.

2. A first version of Neighbourhood Statistics went live in February 2001. This system met the immediate need to make available, in one easily accessible place, the data that were currently available from different sources for small areas.

3. However it was recognised that this was only a small proportion of the data which could be collected together and disseminated, and that systems needed to be put in place that would enable the collection, processing and dissemination of large volumes of data for small geographic areas. The systems also needed to provide the capability to view data through a Geographic Information System, and to allow analysis of data over time against a consistent geographic base.

4. This paper describes the systems, including software tools, now being developed to meet these needs and our plans for future developments and enhancements. The development of Neighbourhood Statistics is an ambitious long-term programme which is being delivered in stages. Section IV of this paper describes these stages: a first website in February 2001; the addition of detailed mapping in May 2002; the release of 2001 Census data for very small areas in Spring 2003; improved data handling capabilities in August 2003; and further enhancements over the period August 2003 to March 2004 which will improve the automation of data collection and processing as well as analytical functionality. Finally, Section V of this paper describes the software tools which will be developed to enable data owners to supply validated, geo-referenced, aggregated, non-disclosive data for small areas to Neighbourhood Statistics.

II. THE POLICY BACKGROUND

5. In April 2000, the Government had published a report[3] as part of the National Strategy for Neighbourhood Renewal. The report identified the need to

> "produce a way of measuring the 'gap' between the poorest neighbourhoods and the rest, so the Government can see whether policy nationally is succeeding in narrowing this gap and act if it is not".

6. There was recognition that many of the statistics that were then available were providing relevant information eg crime statistics, health, receipt of social security benefits, but that the geographic areas at which these data were available were too large for identifying deprived neighbourhoods. Neighbourhoods are by definition small, and so the focus for the development of Neighbourhood Statistics has been at the level of UK electoral wards (NUTS[4] level 5) and lower. Wards vary considerably in size (the smallest 10 per cent have populations below 2,000; the largest 10 per cent have populations exceeding 12,000, with the largest populations being around 30,000), but this level of disaggregation of data should enable many deprived neighbourhoods, located in otherwise affluent regions, to be identified and their needs addressed.

---

[2] Paper reference CES/2001/19: Strategies and Approaches for Small Area Statistics by Dev Virdee, ONS, UK.

[3] National Strategy for Neighbourhood Renewal – Report of Policy Action Team 18: Better Information. HMSO, April 2000.

[4] Nomenclature of Units for Territorial Statistics.

III. THE VISION FOR NEIGHBOURHOOD STATISTICS

7. The vision for Neighbourhood Statistics is:

> "to deliver easier and wider access to geographically referenced, quality assured information that is consistent and coherent. Confidentiality will be the only limitation to enabling general access."

8. The vision for the current programme of work is, that by March 2004, there will be a web-based statistical system for small geographic areas, disseminating information on population, access to services, community well-being, crime, economic deprivation, education skills and training, health, housing, physical environment and work deprivation. The core of this information resource will be the results from the Census conducted in April 2001, amplified by other data held across the public sector. Wherever possible, statistics will be provided for areas smaller than wards to allow deprived neighbourhoods to be identified more precisely. The information will be freely and easily accessible to a wide range of users from government departments, through to local authorities and to the citizen.

9. To achieve this vision requires:

- the development of a high quality, high performance IT infrastructure:
  - the 'front end' of the IT infrastructure will provide a web-based Geographic Information System; (GIS) interface allowing users flexibility of analysis at different levels, by socio-demographic characteristics and over time;
  - the 'back end' of the IT infrastructure will support the capture, cleaning, storage and dissemination of small area statistics;
- the development of a common geographic referencing framework for all official statistics at national and local level;
- the collection on a continuous basis of a range of statistics, supplemented by estimates from statistical modelling;
- the provision of a package of tools, training and support to enable data owners to supply validated, geo-referenced, aggregated, non-disclosive data to the Neighbourhood Statistics system.

10. It also requires a design based around user needs – where users includes those accessing the data and those supplying data to Neighbourhood Statistics. In order to maintain and support this new business operation a modernised service delivery model will be developed based on a one-stop "customer care service" concept accessible to both data owners (suppliers) and users in the widest sense.

11. The vision beyond March 2004 is for a Neighbourhood Statistics system in which data capture and processing, including the application of disclosure control, are automated, and the system is disseminating a much wider range of data down to neighbourhood level. To enable this to happen the software tools for geo-referencing the data, for aggregation and for disclosure control, will be rolled out to a wider range of data owners: these to include local government and the voluntary and private sectors. Estimates for small areas based on modelling techniques will be added to Neighbourhood Statistics, complementing the data available from administrative sources. Increased functionality of the website will allow users to 'pick and mix' the data and analyse these data across a range of geographic areas selected by the user.

III.1 The challenges

12. A project of this scale also brings with it many challenges. Some of these challenges are technical, for example having the correct information management infrastructure in place to handle the volumes of data; or methodological, for example ensuring that the risk of disclosure of information about an individual is acceptably low, or ensuring that all data are referenced to the same geographical boundaries. Further challenges include making the data truly accessible to all – the website needs to be user friendly to the

inexperienced user but allow the expert to quickly and easily analyse multiple datasets. Finally there are external constraints on delivery of the ideal system. These include legal constraints on the sharing of data supplied for administrative purposes, financial constraints, and practical constraints such as having sufficient people with the necessary skills and expertise to manage and execute the developments. Some of these challenges are also addressed in this paper.

III.2 User research

13. Potential users have been consulted about their needs and views on what the systems should deliver. Research has looked at what the users of the website require, and at what tools data owners require to geo-reference and supply data.

14. The range of potential users of the website is wide and there is considerable variability in their needs and likely usage of the site. Research has looked at the requirements of those in local and central government and health organisations, charity and other voluntary bodies, and of citizens. The research has indicated that many dedicated analysts are interested in accessing and downloading data for loading into their own systems. Their interest in analytical functions on the site is limited.

15. However less advanced users want a range of analytical capability such as

- aggregation of supplied areas to form user defined areas;
- estimation and aggregation of data within user defined areas where these cross the boundaries of the supplied areas;
- graphing and mapping;
- comparative analysis of small areas within the larger local authority;
- ability to find ‘like’ areas or ‘extreme’ areas on a number of criteria.

16. Users also want

- profiles of areas to see at a glance the characteristics of areas
- good searching mechanisms both by geography and by subject index
- to be able to save enquiries so that they can return at a later time to the same selection of variables or areas
- rapid email and telephone support

Research into the way people use the existing site has also clearly demonstrated the need for a simple to use, easily to navigate site.

17. The research into the needs of data owners supplying data to Neighbourhood Statistics also uncovered a range of issues about the expectations of potential users of distributed tools. The following points of interest emerged:

- The task is complex - data are held in a number of ways within the data owners’ organisations;
- Many of the data were still under development and therefore technical details were not yet available to inform the design of the distributed tools;
- Mostly, but not exclusively, data are held in a variety of databases (rather than spreadsheets, or flat file);
- Metadata presents a real problem, with many data owners not collating it at all. If metadata is held it is in a variety of formats and these are not particularly structured;
- Guidance/ leadership from ONS was expected in most cases;
- There was widespread interest in the Distributed Tools suggested by ONS;
- At the time of the study there was still widespread concern about disclosure and confidentiality.

IV. BUILDING THE NEIGHBOURHOOD STATISTICS SYSTEM

IV.1 Where we are today

18. The first Neighbourhood Statistics website went live in February 2001. Information from 15 datasets was released for wards, including Indices of Deprivation 2000, benefits data from the Department for Work and Pensions and vital statistics from the Office for National Statistics (ONS). In June another year's worth of data on benefit claimants was added at ward level, together with parliamentary electorate figures. Information on hospital episode statistics and drug mis-users presenting for treatment were added for larger areas. Users were able to download these data in file formats.

19. The Neighbourhood Statistics home page received an average of nearly 34,000 hits per month in the period March 2001 to March 2002, this was about one in eight of the number of visits to the home page of the National Statistics website.

20. The version of the website put up in February 2001 had understandable limitations. It allowed users to find statistics for wards (on boundaries effective in 1998) either by typing the postcode of an address or by drilling down from a local authority name. Both of these methods had limitations for finding wards, and when the area was found there was no accompanying mapping to show its location. Also, the original method of selection of data became harder to use as the number of datasets increased.

21. In order to resolve these issues it was decided to insert an intermediate stage in the programme plan and develop a website to tackle them. This stage would also allow development of skills necessary to build the site required for the release of small area census data in the second quarter of 2003. This site was made live in May 2002.

22. This website employed a mapping system to allow users to see the areas which wards covered. It also allowed users to select data tables using tabulation software. Users were now able to:

- recognise in which ward a property or street is located by viewing a map of the ward;
- find areas of interest by place name;
- select areas and generate tables cutting across national and local authority boundaries;
- select datasets more easily;
- arrange table dimensions flexibly in the way that best suits their needs.

23. These improvements in functionality increased the popularity and use of the site. In the period May to August 2002 the average number of monthly visits was just under 44,000.

IV.2 Releasing the 2001 Census results

24. The next version of the system is due for release in April 2003. It is designed to disseminate the small area Census data at a geographic level of ‘Output Area’ as these data become available over the subsequent months. Output Areas are dissemination units, rather than collection units. Households are grouped together into homogeneous units according to some Census characteristics. Typically, Output Areas contain about 100 to 125 households, and 250 people (lowest size 40 households and 100 people). There will be approximately 200,000 Output Areas across England. These Output Areas will fit within the administrative ward boundaries on the 2003 boundary definitions. This will be the first release on Neighbourhood Statistics of data down to this very low level.

25. Output Areas are the smallest geographic level at which data will be made available on Neighbourhood Statistics. However it will not be possible to disseminate all data (particularly from non-Census sources) down to this level because of the low frequencies within many datasets and the consequent risk of disclosure of information relating to an identifiable individual. The Output Areas are therefore also

seen as 'building blocks' such that by combining them, data can be collected and disseminated at the lowest level of disaggregation that is non-disclosive.

26. A research project, not reported on here, is underway to define 'Super Output Areas'. These are areas larger than Output Areas (but built up from them) which are more closely aligned with a user view of neighbourhoods, and therefore more suitable than wards for policy analysis. Consultations are taking place on the definitions of these areas, and in the longer term they will be made available on the Neighbourhood Statistics system for analysts to use as an alternative to the standard administrative areas.

IV.3 What will the April 2003 system offer?

27. This system will offer all the functionality present in the earlier systems and in addition will offer more data and at lower geographic levels. It will also provide some online statistical analysis tools and additional GIS functionality eg vector maps.

IV.3.1 Data:

28. The key deliverable will be access to the Census data at Output Area level and above as they become available. The higher geographic areas will be wards and local authorities. These data will include analyses of the population by age, sex, marital status, educational qualifications and many other variables. These data will all be presented consistent with current, 2003, ward boundaries. In addition there will be a selection of pre-created thematic maps.

29. The volumes of data are immense: approximately 7,000 variables for 200,000 areas. Census metadata associated with the data will also be added to the system.

30. Some other new datasets will also be made available referenced to the 2003 boundaries. A number of priority datasets have been identified and work is ongoing to enable these to be delivered to the system by April 2003. Delivery of data involves quality assurance of the data as well as ensuring that the dataset is non-disclosive. Manual disclosure routines will be applied to datasets for this version of the system. In addition, it is hoped to make available some small area income estimates, the result of a modelling exercise using sample survey and other data.

31. The existing data, which are currently held on the system at 1998 ward boundaries, will be transferred to the new system. Initially these data will continue to be held consistent with the 1998 ward boundaries: they will not be precisely recalculated to the new 2003 ward boundaries because of the risk of disclosure of information which could result from differencing the old and the new areas. A project is underway to identify the best methodology for re-casting existing data to new geographic boundaries. This should result in the provision of existing data on the new boundaries by summer 2003 which will then enable these data to be readily analysed with the Census data and with other data made available at 2003 boundaries. The re-casting of all data to 2003 boundaries will also provide longer consistent time series data with which to assess changes over time. The provision of time series on a consistent geography is important to many analysts using Neighbourhood Statistics.

32. The availability on Neighbourhood Statistics of population counts from the Census will also provide denominators for datasets across the whole system. These will in turn enable the calculation of rates and proportions – considerably more useful to analysts wanting to compare areas of different sizes than the simple counts of occurrences of, eg benefit recipients.

IV.3.2 Functionality:

33. The system will offer some online analysis functionality. This will include accessing data through either statistical topics (domains) or through the geographical area (ie selecting data for a particular administrative area, place name or postcode). When using geography, users will be able to enter via a map or by typing in the postcode or name of area. It will also be possible to select multiple areas to display

together for comparison purposes. Data will be presented as tables and some manipulation of data will be possible, graphing will be available and further functionality may be available including dynamic map output However this functionality may have to wait until the August 2003 enhancement (see below).

34. However, the software being used to hold the data, Beyond 20/20, imposes some constraints on manipulation of the data. The software's primary use is as a tabulation package; it does not allow the flexible interrogation of the data that would be possible with a database solution. This issue will be addressed in future versions of the system but, for this version, strategies are being put in place to mitigate the effects of the constraints.

35. To facilitate the analytical capability to compare data from different sources, a pre-selection of key variables is being made and these will be held such that any of them can be analysed together for the chosen geographic area(s). All rates and proportions will need to be pre-calculated and they will be held in the same way as the original variables. There may still be limits on the cross region analysis and so complex analysis may require a number of runs of the analysis. Again this limitation is being addressed in future developments.

36. Finally, text-based profiles and pre-created maps of key statistics (including Census outputs) will be available for administrative wards and Parliamentary constituencies. The provision of key statistics for other geographies such as National Parks is under discussion.

IV.3.3 User support:

37. Much of the user support will be an extension of the support already available to users of Neighbourhood Statistics. This includes a help desk providing support and advice both by telephone and email. Contact details for advice on specific datasets are published as part of the metadata on the website.

38. In addition, there are plans for an online ordering service, a facility to download digital boundaries via the main National Statistics website, and a facility for users to save their selections so that analyses can be repeated at a future visit. It is also planned to improve the 'Help' texts and to make available more communications materials such as press releases.

IV.4 Further enhancements: August 2003 to March 2004

IV.4.1 August 2003

39. The next key delivery date is August 2003 when new systems for the collection, validation and storage processes will be introduced. These will give much improved data handling capabilities. Whilst there are some functionality improvements in this release, the immediate beneficiaries of this development will be those supplying and processing the data rather than users of the Neighbourhood Statistics website.

40. The intention is to introduce Oracle as the strategic relational database management system. This will involve the creation of a new hardware and software infrastructure and introduce new applications development tools. The development will enable data to be loaded from a variety of sources and validation procedures applied. Validation will include checks on completeness of the data and that the correct geo-references have been used. It is also intended to increase the automation of the disclosure control routines and ensure that the data are sufficiently protected even when analysed alongside data from other sources. Also included within the system will be the provision of management information for controlling, monitoring and reporting on the flow of datasets.

41. The August 2003 system release will also provide some increased analytical functionality. Plans include the provision of dynamic mapping, improved data manipulation and manipulation of outputs. There will also be a facility for users to define the geographic area for which to output the data. This will be based on aggregation of available building blocks by a best fit methodology. Following on from the development work to recast data between wards defined at different points in time (paragraph 31), methodologists are

looking at estimation procedures to enable users to define geographic areas which cross standard boundaries.

42. The ability to find and compare data for different areas is also a planned enhancement but this requires a different type of web dissemination tool from that currently used. The timing of the full development is dependant on assessments of alternative software including the use of functionality within Oracle. Some interim solutions will be put in place.

43. So, in August 2003, the benefits to website users will be seen in the provision of more data to the website and in increasing analytical functionality. There will also be better metadata provided as the result of standardising our requirements and ensuring that data owners comply.

IV.4.2 August 2003 – March 2004

44. Between August 2003 and March 2004, further improvements will continue to be made to the automation of the data collection and processing (back end) of the system. In particular, the system will be prepared for the delivery of data by means of the software tools to be rolled out to data owners in March 2004. These tools are described in detail in the next section of this paper.

45. It is also planned that the software used for the web dissemination of the data will be replaced with more suitable software which will allow rapid routing through the data (which by March 2004 are likely to comprise the Census data and more than 100 other datasets) and better capability for flexible selection and analysis of data. Some of the interim solutions developed in the previous period will be replaced with more robust and comprehensive analytical capabilities.

IV.5 Non-functional capabilities

46. Research into the current and future level of usage of the website has shown that the number of users is expected to grow substantially as the value of the website increases. Assessments are being carried out to inform the design of the system to ensure that it continues to deliver a reasonable performance as volume increases.

47. However, some factors affecting performance are outside the control of Neighbourhood Statistics such as fluctuations in general internet loading, performance of ISPs, and the modem speeds of the users. To allow for these, a range of facilities are being included to maximise useful performance. These include:

- a rapid route to simple statistics without the need to load and display detailed maps, which are bandwidth hungry and computationally intensive;
- information and control when downloading large files.

The use of outsourcing enables the system to adapt to changes in capacity demands and maintain an acceptable speed of response, as well as providing agreed levels of reliability and availability.

48. Usability centred design is a high priority for the future versions of the web-site. Research has been, and will continue to be, conducted with real users to identify the strengths and weaknesses of the system, and improvements will be made on an on-going basis as the site develops. This will ensure that the system remains easy to use although the complexity and quantity of the data available increases over time.

49. Security is implicit within the design, preventing unauthorised modification of data held within the system, as well as protecting data before it is formally released for public access. The design is also being undertaken with the aim of facilitating maintenance, enhancement and changes to meet future business needs and take advantage of improved technology.

V. COLLECTION OF DATA: THE PROVISION OF DISTRIBUTED SOFTWARE TOOLS TO DATA OWNERS

V.1 The Tools Vision

50. As already identified (paragraph 9) the vision for Neighbourhood Statistics includes the provision of software tools to enable data owners to supply data to Neighbourhood Statistics. The original vision for distributed tools encompassed:

- A package of tools, training and support;
- Provision of tools to enable the adoption of a common geographic referencing framework;
- Tools freely available to data suppliers;
- Disclosure protection procedures.

51. It is central to Neighbourhood Statistics that existing administrative records be processed, at source, to produce geographically precise information. The key to this is seen to be the central development of software and improved geographic referencing datasets which will then be supplied to data owners, free of charge, to enable them to create small area statistics. The provision of centrally developed tools will enable the creation and integration of statistics for small areas. This is not currently possible because of the many different methods used to add a location reference to data. Data owners should also benefit from the provision of these tools – the ability to link geographic codes to administrative records will allow organisations to exploit Geographic Information Systems to access their data in new ways, enhancing the usefulness of the data to them.

52. The creation of aggregated statistics for small areas will remove the need to pass potentially confidential records to ONS. There will, however, remain a need to ensure that these aggregated statistics do not disclose confidential information about any individual or household. A programme of work is underway in ONS to develop disclosure control methods and systems that can be used by data owners supplying data to Neighbourhood Statistics. A policy on data sharing and confidentiality, which will operate across National Statistics within the UK, is also being developed. One principle of supply of data to the Neighbourhood Statistics system is that the data are aggregated to small areas and have had disclosure control routines or methods applied before receipt.

V.2 The development of distributed tools

53. A pre-requisite to the development of tools for data owners is an understanding of the environment in which the data owners operate, in particular the way in which their data are held and the tools already available to them for delivery of small area data. Although some user research had been undertaken (see above, paragraph 17) it had been based on only a small number of interviews with data owners. It had also raised a number of issues which required further investigation.

54. It was therefore decided that a more detailed study of the needs of data owners and the availability, or otherwise, of existing suitable software should be undertaken before any development of tools commenced. A more in-depth Feasibility Study was therefore undertaken over the period July – September 2002.

55. The remit for the Feasibility Study was to establish the business need, the feasible options and the associated cost levels for distribution and on-going support for the software tools. The objectives of the study were:

- to research the data owners' (including potential data owners) technical requirements of the tools;
- to consider the scope of the software component of the distributed tools in terms of the functionality required and potential software platforms;
- to identify a range of technical options for the distribution of the tools (both software and reference data);
- to identify a range of technical options for the ongoing future support of the tools.

56. As a result of the Feasibility Study, the original set of tools envisaged by the programme was broken down into a set of more detailed processes. The processes identified were:

Dataset Extract: extract the raw data from the data owner's database in readiness for processing

Data Record Validation: validate and clean the addressing information to ensure that the geographic location of the data record is correct, and that it uses a referencing standard that is usable in the rest of the process

Add Standard Geo-referencing Codes: add the geo-reference codes to each data record; these to include the grid reference and also standard geo-codes for Neighbourhood Statistics including Output Area and Ward

Aggregation: aggregate data to the standard geographic areas required by Neighbourhood Statistics, eg to Output Areas. Also aggregate data to any standard classifications applicable to the dataset, for example age range, sex, marital status

Apply Disclosure Control: apply the appropriate disclosure control methods to the aggregated data to ensure the desired levels of confidentiality. This may involve further aggregation of the data to remove the smaller areas if a risk of disclosure of information about an individual or household remains

Add Metadata and Format: add the appropriate metadata, eg information on quality of address matching (which will have been collected as part of the process). Format the output dataset for transfer to Neighbourhood Statistics

Dataset Transfer: transfer the dataset to ONS (or any intermediate agency) by means of physical media, for example CD, or via the Web

Visualisation: the ability to visualise the data against a mapped background. Visualisation aids the checking of data quality as well as being a useful tool for analysing local data for the determination of local policy

57. The Feasibility Study report made two proposals which differed from the tools 'Vision'. These were:

- that the tools be available individually in the form of a toolkit as opposed to the vision of integrated tools; and
- that the value of a visualisation tool was questionable because most data providers already have this facility.

58. The study report also advised that it was not possible to develop a set of distributed tools in time for the August 2003 system. It therefore proposed a two tier approach of an interim solution for August 2003 and a final solution for March 2004. For the interim solution, a combination of accredited software, in house software development and development of templates was recommended. The recommendations in the report have been broadly accepted (except that it was decided that there was insufficient time to accredit commercially available software) and the solutions proposed are now being turned into concrete plans for development. Adoption of the interim solution will ensure that the flow of data into the Neighbourhood Statistics system will continue, but the processes will be more manual (and hence more resource intensive) than the automated approach of the final tools.

59. The costs presented in the study report are very high. This is largely the result of very high licence costs for the software to be accredited and the allowance made for training data owners in the use of the tools. The costs are being examined to see where savings can be made.

V.3 Interim/Final Approach

60. The interim solution will provide sufficient functionality to data providers to allow their data to be presented to Neighbourhood Statistics in time for inclusion in the August version of the system.

Discussions will be held individually with each data owner, due to supply data before the final solution is in place in March 2004, to determine their requirements for disclosure control, geography and data acquisition etc. The solution will then be built around the data owners' needs.

61. A summary of the interim solution for meeting data owners' needs is presented in the table:

Tool	**Interim solution**
Geo-Referencing (incorporates 'Data record validation' and 'Add standard geo-referencing codes')	ONS to provide a bureau service to data owners for address matching and adding geo-referencing codes, subject to feasibility
Aggregation	Use data owners existing aggregation tool or ONS to advise on developing a spreadsheet or alternative simple tool
Disclosure Control	ONS to provide Disclosure Control Standards and Guidance document together with a specification of routines for random rounding
Meta Data & Format	ONS to provide a template for the data owner to complete.
Dataset transfer	Method (probably CD or internet) to be agreed with data owner
Visualisation	Do not provide this tool

62. The recommended **final solution** for the delivery of tools is still under consideration. The recommendations have been broadly accepted, but issues remain around the accreditation of commercially available software as against in-house development or in-house adaptation of software. This particularly applies to the 'Data record validation' and 'Add standard geo-referencing codes' tools.

63. For the 'Aggregation' and 'Disclosure Control' tools it is hoped to specify and directly commission further development of an existing disclosure control tool specifically to meet the needs of Neighbourhood Statistics. For the 'Metadata' and 'Format' tools, the recommendation was for ONS to complete work on developing a software tool. This is being taken forward. The 'Visualisation' tool will be removed from the set of tools because of the associated high costs but low benefit to Neighbourhood Statistics. The research showed that many data owners already have a visualisation tool for their own purposes.

64. The decisions on the way forward for developing these final tools will be taken over the next couple of months. Work will then need to proceed quickly to ensure that the final tools are delivered to data owners by March 2004.

VI. SUMMARY

65. Neighbourhood Statistics is an innovative, long-term development which is already bringing benefits to users of small area statistics. Key users, including analysts within central and local government, are able to access more data more easily and to use this to inform those who are responsible for developing policies to reduce deprivation and social exclusion. Comparisons of deprived (or other) areas across the whole country can be made on a consistent basis.

66. As more data are included on Neighbourhood Statistics, the benefits of a system which holds data from many sources on a comparable basis, in particular all geographic areas defined using the same reference data, will increase. The methods being used to geo-reference data and aggregate them to small 'building block' areas will also reduce the impact of boundary changes and enable analysis over time – important for monitoring the effects of policy changes – to be conducted with confidence.

67. Data owners will also benefit from the provision of software tools that will allow them to geo-reference and aggregate their data and then to analyse them for their own purposes.

68. A number of enhancements to the system are planned and many have been described in the earlier sections. There will be increasing pressure to continue these enhancements:

- to increase the functionality offered to website users;
- to include more data from a wider range of sources including estimates for small areas derived from statistical models;
- to improve the automation of data capture and processing.

69. But challenges remain:

- With increasing numbers of datasets from a wider group of data owners, how do we ensure that the quality of the data remains high?
- How do we ensure a standard 'look and feel' for data from many sources?
- How do we ensure that the navigation around the website continues to provide quick access to the required data?
- How fast do we roll out to data owners the software tools for capturing data? How do we prioritise?
- How do we manage the expectations of users, and meet the needs of different types of users?
- What systems do we need to put in place to support the potentially large numbers of users and their enquiries?

REFERENCES

National Strategy for Neighbourhood Renewal – Report of Policy Action Team 18: Better Information. HMSO, April 2000.

Neighbourhood Statistics website: www.statistics.gov.uk/neighbourhood.

- - - - -