

INFACF AND FES ON LINE – WEB BASED SOLUTIONS

AUTHOR: Carol Calvert, Further Education Statistics Branch, Scottish Further Education Funding Council

1. SUMMARY

This paper outlines the design, development and implementation of two related projects. Both projects relate to the same large data collection; one deals with the collection of the data and the other with its publication. Both projects use SAS IntrNet, web browsers, and JAVA.

The collection facility allows secure access to our data providers for data submission and checking. Data is run against SAS programs and error reports are produced immediately for the provider. The data provider uses the error report to correct their data and resubmits data files until there are no errors. At this stage data is loaded into the master data file for the collection.

The publication facility allows the user to access a simple analysis tool and use this to define tables and charts of information that they require. The tables are not predefined but are generated in response to the specification by the user. Access is open to anyone with a web browser.

Both projects were developed with edina software who are an Edinburgh software firm.

2. BACKGROUND

Further Education in Scotland

Further Education colleges in Scotland are major providers of post-compulsory education in Scotland. They provide both academic and vocational education as well as specific skills training. The colleges provide education opportunities for many older learners - over half of the students are aged over 25 years. They are a crucial element in the lifelong learning policy agenda.

Courses are delivered in a range of ways including full time, evening classes and day release. In some cases colleges have formal agreements with Universities for students to take the initial part of a degree at the college and the final part at the University. The traditional strength of FE colleges is in providing technical and vocational skills.

There are currently 47 colleges in Scotland delivering over 25,000 courses with over 400,000 student enrolments annually. Colleges are responsible for

running their own affairs and a college will offer the range of courses that they wish to offer. Public funding for much of their activity comes from the Scottish Further Education Funding Council

The Scottish Further Education Funding Council (SFEFC)

The Scottish FE Funding Council was established in 1999 primarily to administer funding to FE colleges. This was a function previously directly undertaken by a government department and prior to 1993 was the responsibility of local authority government. The Funding Council, as the largest customer for FE college data, was also asked to take over the responsibility for the data collections relating to FE colleges. The Council is now therefore responsible for the national data collections on FE and the data we collect is provided by us to the government for use in national statistics.

The Scottish Further Education Funding Council shares accommodation, staffing and IT resources with its older sister organisation the Scottish Higher Education Funding Council. The Councils jointly employ around 115 staff, of whom about 10 are SAS users. There is an IT support unit but, because of the low proportion of SAS users, this cannot offer any SAS support. We therefore rely on our knowledge and call on consultants for SAS expertise to supplement this where necessary.

The FE Statistics Branch

The Branch deals with all data collections from FE Colleges. The major collection, Further Education Statistics (FES), consists of 430,000 student records and 26,000 course records each year. To deal with the volume of data SAS was chosen as our main analysis tool. In addition to providing analysis and advice to the Funding Council, the Branch has all the responsibilities connected with data collection. This includes the obligation to make the data we collect available, as the national data collection, to government and other organisations, researchers and to the general public.

The data collection / cleaning procedures we inherited were basically manual and paper systems. Raw data was sent electronically but then it was run against checking programmes and paper output was produced. The paper error reports were sent out to colleges and colleges then wrote corrections on the sheets. The corrected paper error reports were returned to SFEFC and the changes manually keyed into the main system. The new information was then run against the checking programmes and remaining / new errors identified and paper reports generated. The cycle continued until all the errors were cleared.

The Branch was aware of the high demands made by the FES error checking process on colleges. The main issue was that all the clearing of errors had to occur in a short period of time between September and December. In September colleges were busy with new enrolments and, by December, the colleges were truly into the next academic year. Clearing up data errors which related to the previous year, in December, was both difficult and of low priority to the colleges.

The dissemination policy, by the government department previously responsible for the data collections, had been through tables in published government statistical bulletins and, typically for researchers, by providing a SAS dataset. Other requests for information were dealt with by writing ad hoc SAS programs and results were frequently then transferred into EXCEL or WORD for the enquirer.

With the establishment of a new Funding Council we anticipated a high internal demand for published information in connection with the Council's functions relating to funding, planning and financial accounting. We were aware that as a new organisation we did not have the necessarily skills amongst all staff for a high level of ad hoc SAS programming. Nor did we have the staffing level to support any volume of ad hoc activity. However, on the positive side, there was no 'history' of SFEFC produced statistical bulletins that we had to maintain. We also had high quality new IT equipment and an organisation 'mind set' that was willing to consider and embrace innovation and change. Most importantly the new organisation was one committed to openness and transparency, especially in relation to its funding decisions.

3. PROJECT REQUIREMENTS

Data collection / cleaning – FES on Line

There are several different Management Information Systems (MIS) currently in use in FE colleges in Scotland. The basic MIS system has then, in many cases, been tailored to an individual college requirement and a college chooses and maintains its own MIS system. Potentially then, with 47 separate colleges there are 47 different systems. The FES data collection required a flat file, to a specified format, to be written by the college MIS system and sent electronically to SFEFC.

In previous years the 'checks' on the data which were used by SFEFC have been supplied to the college software houses for incorporation in the college MIS systems. However it has proved impossible to achieve this satisfactorily with the result that colleges 'cleaned' their data, using their college software, only to find that on sending the data to us that there were a large number of other errors of which they had been unaware. The fundamental requirement was therefore to improve this position by making our 'checks' available to all colleges.

A second major requirement was that the system adopted had to start from the current position both in colleges in SFEFC. In effect this meant that the ability of the MIS colleges systems to write out flat files, to a defined specification, could be assumed but little else. At SFEFC we required the final 'master' files to be SAS datasets to enable us to continue to use our existing programs. Any time requirement on the colleges and SFEFC to operate the new system had to be minimal.

The existing processing took a large volume of central staff time and time in colleges. Between September and January, SFEFC spent 270 days on FES. That is an average of 5-6 days per college and I believe that colleges spent at least a corresponding amount of time trying to sort out their data errors. We required the FES on LINE project to deliver substantial savings in staff time at both SFEFC and in the colleges.

The old system was very inflexible in terms of timing. All checking / cleaning / submission of data was basically taking place at the end of one academic year and at the start of the next. The flexibility within FE colleges means that many courses start and end at various times in the year. We required the new system to be sufficiently flexible that colleges could submit their data on a 'phased' basis if they wished and add in new details at any stage of the year.

The final major requirement related to security. The students records to be sent to SFEFC, and the error prints being returned, identify individual students and contain details such as name, date of birth, gender, ethnic origin etc. The system had to be secure.

Although this overall requirement was clearly outside the Branch SAS expertise and our IT resources to develop, we did require that the Branch could maintain it. As a new organisation we needed to be careful about introducing change that might have teething difficulties. So we added two further requirements to our list- no 'rocket science' and we needed to use the software we already had as far as possible. We took this requirement to tender.

At the time we ran SAS 6.12 and MICROSOFT OFFICE 97 , we had an NT SAS internal server and another SAS server with SAS IntrNet already on it.

Publication – IN.FACT

The Branch had to be able to deliver accurate responses to a high volume of questions about activity in FE colleges. Many of these questions would be from internal colleagues but there was an additional responsibility, because we were holding the national data collection, to respond to all reasonable requests from the public within 15 days. It was evident that the Branch would be unable to deliver on this minimum requirement given staffing levels and our commitments to many other functions of the Council.

In addition, the organisation's own commitment to openness and transparency would not be met by simply continuing with a policy of statistical bulletins. Inevitably only part of the collected information could ever be made available through a predetermined paper dissemination mechanism. The Council was encouraging colleges to engage with issues about strategic planning to meet the needs of the communities they serve, and it was also encouraging colleges to equip students and themselves with the IT skills necessary for effective working. This clearly pointed us down the route of looking for an IT solution.

Our requirement then effectively came down to:

All data – within data protection limitations- on student and course activity should be made easily and quickly available to anyone in a readily understandable way and in a useful format.

We also had to make minimum demands of the user – in terms of IT equipment, IT skills, knowledge of what data were actually available and analysis skills.

Considerations of our clients for the data helped us further refine the requirement. One priority client was Government who required the data to compile national statistics and with whom we have a Service Level Agreement. This client was a SAS user and employed qualified Government Statisticians. Therefore we could continue to adequately meet their requirements through provision of a raw SAS dataset and they were not considered a client for the new product – although in practice this has turned out rather differently.

Another major client group was the internal Council user. These users would always have the expertise of the FE Statistics Branch to call upon but we needed a way to deal with their more straightforward enquiries. As a new organisation, with many staff new to working with the FE sector, it was likely there would be a lot of internal 'background' relatively straightforward requests. Effectively we needed to filter, and answer, those enquiries where detailed statistical knowledge / expertise were not required.

The next major client group were the data suppliers i.e. the FE colleges themselves. They had, under previous arrangements, received paper feedback showing their college information in relation to national average figures. Additional information would be supplied on request to colleges by running an ad hoc enquiry. It seemed, to us, that the colleges had need for information at two quite distinct levels – the broad picture and at a much more detailed course level or student level.

The final major client group was the general enquirer who wanted some information, usually for background, about the sector activity as a whole or on a particular area of activity. We anticipated that with new Scottish Parliament, the increasing interest in social inclusion and widening participation and the importance of the lifelong learning agenda we would receive a high volume of such general enquiries.

Consideration of these groups led us to define our client group as interested in FE activity but not necessarily informed about what information was available. We therefore envisaged the new project as providing the 'first' stop for information and we put another data source, an ACCESS database, in place for the more detailed enquirer.

Thus we envisage a product that did not have to 'be attention seeking' but which did need to be readily and easily available for anyone interested in FE in Scotland to access. User requests for information were likely to be very

variable and perhaps, initially, ill defined due to lack of knowledge of what was available. We decided that, to meet the organisations commitment to openness and transparency, we could not go down the route of pre defined tables. We also felt we needed to give the user the ability to 'refine' their request as they went along. Users were likely to want to use the information in their own reports so we needed to integrate our project output with other spreadsheet and word processing packages. Finally the user might be more comfortable with a simple chart rather than a table of information so we needed some rudimentary graphing facility.

As with the data collection / cleaning package we required no 'rocket science' and we needed to use the software we already had as far as possible. At the time we ran SAS 6.12 and MICROSOFT OFFICE 97 , we had an NT SAS server and were about to purchase a new web server to replace an ageing APPLE MACINTOSH.

4. SOLUTION

Data collection / data cleaning

The solution was to have a bespoke package – FES on Line- which only requires a college to have a web browser and which uses SAS , on an existing web server, to check the data and generate errors where necessary. The communication is handled via SAS IntraNet and uses Java servlets for username and password authentication to provide security. All FES on Line traffic is also encrypted via Secure Sockets Layer (SSL).

Colleges log into a SFEFC server and through a series of screens can send, validate and submit their data. On validation an error report is produced which colleges can view on screen or save and download to their own PC. A college can leave files that contain errors on the system or delete them as they wish. Files that have not passed the validation checks cannot be submitted by colleges to the master files.

In addition, SFEFC have a range of administrator rights which include changing college passwords, resetting files, and the right to 'force' files through which have not passed all the validation checks.

Publication

The solution we have is a bespoke package called IN.FACT. The analysis of the data is supported via a WWW browser query screen. The Query screen runs a SAS program that extracts, summarises and presents to the enquiring user in the form of a frequency table or graph. The user can then view, print the results, or cut and paste the data across to other software packages for further analysis. The performance objective is that analysis should generally complete within 10 seconds and can be run while the requestor waits for the result. Users can access IN.FACT either directly or through our general web server.

The classification variables and values and comparison operators are specified using lists and / or pull down menus. Users can request either or both row and column totals and results are delivered as a table or a table and

a chart. The information presented in each cell (or element of the graph) is either an absolute number (count, sum or mean as appropriate) or a percentage (relative to the row or column total).

Selection clauses are available to the user using WHERE and IN statements. The results of a query are delivered as an HTML table and include a description of the results, the selection criterion used and the date on which the query was executed. Users can build queries with multiple selection criteria in stages. They first specify a single selection clause, run the query, add another clause to restrict data further, run that query etc. This strategy allows a user to drill-down into the detail in any way they choose.

Users can also request a proximity analysis. This allows students within a specified distance of a specified centre to be selected. This analysis allows a college to generate comparative statistics for themselves and adjacent institutions.

There are a number of safeguards in the package; some deal with data protection issues, some help prevent an inexperienced user misinterpreting the information and some protect the system from a user crashing the system accidentally or deliberately.

Detail

A user links to our web site using their web browser. The software was designed to run with Internet Explorer 4.0 and Netscape 4.5 running under Microsoft Windows.

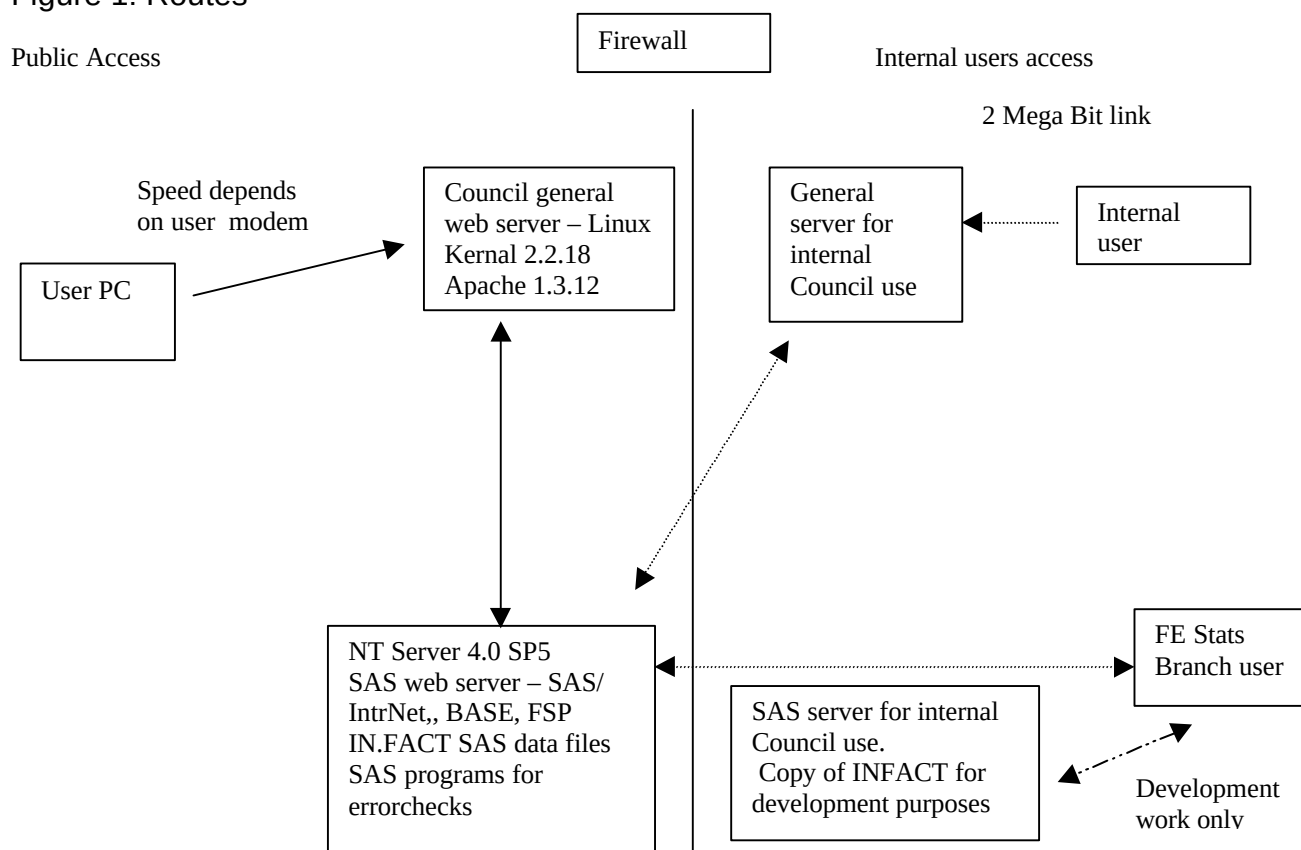
The Funding Council provides a dedicated SAS web server which is a NT server with Base SAS, SAS FSP and SAS/ IntrNet. At the time both these projects were developed we operated under SAS 6.12 although we are now on SAS 8.1. The NT server runs Microsoft Internet Information Server (IIS)

Response time for FES on Line varies between colleges depending on the size of their data sets. However, even for the largest colleges it is below 5 minutes from sending data to an error report being available. Response time for IN.FACT internal users, over our 2Mb link, is 2-3 seconds and for external users is dependent on the speed of their modem. Response time appears to be satisfactory for users.

In addition, to allow us a platform for development work, we hold a complete copy of FES and IN.FACT on our internal SAS server. This can only be accessed by members of FE Statistics Branch.

Log are generated for both FES on LINE and IN.FACT. The FES on Line log provides a record of all actions taken by the system "administrator" and of any messages sent to us by colleges. For IN.FACT the log includes for each query, the date and time, the query options selected, the network address from which the query originated, the approximate amount of CPU used executing the query and the elapsed time taken to execute the query. This enables us to monitor the use of IN.FACT.

Figure 1: Routes



5. PROCESS OF IMPLEMENTATION

Time table

IN.FACT was the first project and was conceived of as an idea in August 1999, the requirement, the funding and the contractors were agreed by December 1999. Work commenced in January 2000 and IN.FACT went live in July 2000. FES on Line was developed as an idea during the summer of 2000 and contractors agreed after a tendering exercise in November 2000. Work started in January 2001 and by June 2001 we had an operational system that colleges were using.

Non technical issues

Senior staff support for the first project- IN.FACT - was relatively easy to secure because the package clearly met a major need. Funding was relative easy to secure because the potential for saving staff time was so large. The timing of the project helped, as did the commitment of the organisation to using IT. Support for the second, and larger project FES on LINE, was also forthcoming not least because IN.FACT had proved so successful.

In both cases the same contractor was used. For IN.FACT the contractor was chosen on the basis of previous work with SFEFC. For FES on LINE, after a tendering process, the same contractor was appointed.

After securing funding, and a contractor, the next most important issue for us was to ensure that data providers were content. For IN. FACT this meant colleges needed to be happy with the with the level of information being made publicly available. All the data providers were consulted and many saw a demonstration of INFACT given by FE Statistics Branch staff - linking remotely to the development copy on our internal server. For FES on Line we needed to reassure the colleges that the package was actually giving them more control and flexibility rather than increasing their workload. As with IN.FACT demonstration's were arranged and many colleges were visited for individual demonstrations.

In parallel we sought legal advice to ensure that our proposals were going to meet the requirements of the current data protection act.

Technical issues

Collection / cleaning

The biggest single issue was resolving the security issues. The major issue was to maintain security and ensure both college and SFEFC systems could adequately 'talk' over the secure links.

As colleges started to use the system it became apparent that a number of minor modifications would be helpful. This included altering how often the college was required to change their passwords; removing the limit of the size of the error report; and linking the FES on LINE 'master' files to the 'master' files produced by the more traditional route.

Publication

We decided to protect individual student data by entering n/a if the number of students was 1,2,3 or 4. To discourage identification of individuals by subtraction we also opted to place n/a in any row or column totals that contained cells with n/a entries.

Variable names, which we already used in our SAS datasets, were in some cases so long that the tables were badly sized on the screen – this particularly effected the bar charts. We therefore had to produce some shorter, abbreviated field names.

To aid user interpretation we placed some restrictions on the output from INFACT. For example, if a chart is requested the corresponding table of information is also produced. Percentages cannot be displayed without the numbers on which they are based also being displayed. In a two way table only column percentages can be displayed.

We had to purchase two SAS/IntrNet licences. Running an internal developmental version of INFACT is very helpful but costly because of the

additional licences required. However, for the moment we will continue to do so whilst IN.FACT is still being enhanced.

6. Reality

Use

In the year that IN.FACT was running we had 11,438 requests for tables. We use the request for tables as an indicator of real enquiries and of the table requests 3,637 (32%) were for internal users. Amongst the external users the Scottish Executive- the government department already supplied with the SAS raw data were the major user requesting 913 tables (8% of all requests). Discussion with them has established that they find IN.FACT a very good way to deal with most of their internal, government department requests.

FES on Line has only been available to colleges for the last couple of months. The reception in this time however has been very positive and colleges are already suggesting a lot of additional features that they would appreciate being added. The real test will be when the major data return takes place in September when we hope to see an increase in data quality sent to us. Then throughout the academic year 2001-02 we hope colleges will use the package on an ongoing basis.

Benefits

The major benefit of IN.FACT is that we have been able to meet a large volume of requests for information. Prior to IN.FACT, to deal with an external request for information would easily take two hours; there would probably have been an initial call, followed by writing some simple SAS code, then a call to confirm the extracted information met the need, and finally a letter sending out the details. On average this would have taken at least two hours and hence to process just the 7,000+ external requests would have been well beyond the ability of the Branch to staff ($7,000 \times 2 = 14,000$ hours and at 37 hours per week and 45 weeks a year that is $14,000 / 37 \times 45 = 8+$ members of staff). We have spent perhaps five days in total of Branch activity on specific external data requests using INFACT. In addition we spent a total of five days of Council time in setting up INFACT.

Users have been able to answer their own requests immediately and accurately; no potential for mis interpretation on our part of the users request, no copying errors to and from SAS, and no errors in SAS programming of requests.

We have made a large contribution to our organisations commitment to openness and transparency. This helps us in our wish to encourage colleges in continuing to develop their use of IT. It also results in us being in a strong position in relation to delivering on the government target that 100% of public services which can feasibly be delivered electronically should be available on-line by 2005.

We have had a lot of positive response from all users. As a Statistics Branch we are more used to people voicing concern that data are not available rather than people telling us how much they like our dissemination approach and package. It is good for the Branch to get so much positive comment.

Users seemed to have been impressed by how much is available rather than critical about what is not. It is likely that increasingly we have users specifying a table from INFACT where they would not have entered a formal request with its consequent time delay and perhaps cost.

It is rather early to identify the full benefits of FES on LINE. However, it is already clear from the colleges that have used it that we provided something they find helpful. Colleges are full of praise for the product and enthusiasm at the prospect of being able to use at any stage they wish during the year.

Issues and next steps

The major issue is that time and technology move on. The mechanisms we are currently using will already be outdated. If we were designing now doubtless the operation mechanism would be different although the product look would probably not be.

We have requests for back years more information to be put on INFACT and for more information on the variables to be put in the help files. The reasoning behind these requests is to enable users to do more analysis, especially time series analysis. We have also had requests to make information on activity in Higher education Institutions similarly available.

Colleges are already asking us if we can link FES on Line to an ACCESS data base with built in queries. This would give them a current picture of what was happening and allow them to investigate their data in a more flexible way than they can at the moment using their MIS systems.

Contact details

Scottish funding councils for further and higher education
Donaldson House
97 Haymarket Terrace
Edinburgh
EH12 5HD

INFACT: www.sfefc.ac.uk/infact/

Email : ccalvert@sfc.ac.uk

Tel : (+44) 0131 313 6537

Fax : (+44) 0131 313 6501

Software house

EDINA Software Limited
40 Queen Street
Edinburgh
EH2 3NH

Scotland

UK

Email: info@edina.co.uk

tel : (+44) 0131 624 6624

fax : (+44) 0131 624 6625