

NHSCR Governance Board Creating A Backup To NHSCR Archive

Background

1. A fundamental review of the work of the NHSCR was carried out in the latter half of 2003, under the leadership of the SEHD Deputy Chief Medical Officer, on behalf of the NHS. One of the Review Group's recommendations was that the NHSCR should examine the cost effectiveness of creating a backup to its paper resources, proportionate to the risk of destruction of vital archive material and irreplaceable information.

The paper resources

2. Almost all of the NHSCR's data is computerised, and there are adequate standard arrangements for that data to be backed-up, and a copy kept remotely, to guard against the risk of destruction by fire, computer malfunction or the like. But the original books recording the 1939 enumeration of the population, which was the NHSCR's original source, exist only in paper form. In that respect, they are no different from the enumeration records of the 1931 or 1951 census. Like these census records, they will be of great value to genealogists. But, unlike these records, they are also a current source for administrative work, rather than being stored securely against the day when they are released, 100 years after the enumeration. So the risk of damage, and the disruption which would be caused, is greater. The loss of these unique historic records would result in a 15 to 20% reduction in our ability to confirm the identity of individuals 'lost to follow up' by Health Boards and medical researchers. The registers also afford us the ability to identify family groups. So there is a clear case for creating a back-up.

Creating a back-up

3. The obvious way of creating a back-up is by digitising the records, as we have done for the Victorian censuses. That is cheaper than transcribing them, and allows the original document to be reproduced and shared (eg on the internet) without transcription errors.

4. We have a great deal of experience in digitising records and, when the Review Group made its recommendation, we used that experience to draw up an estimate for the digitisation of the NHSCR paper records. This is annexed. It shows that, while the project is feasible, it would cost around £800,000.

5. We have recently reviewed that estimate and are satisfied that the figures are still reasonable. It is possible that the costs for the imaging and indexing would come in under the estimate. For instance, a cost of £1.40 per page for imaging may be achievable and this would reduce the overall cost by around £72,660. Similarly, it is possible for the indexing to cost as little as 21p per thousand characters which would save some £26,600. However it may be that the lower cost could involve more in-house quality review work. It is a balancing act which can only be judged when the tenders have been received.

Meeting the cost

6. The cost of digitising the Victorian censuses was met by a special funding allocation from the Scottish Executive, predicated on the expected revenue from the

sale of internet access for genealogists. But, unless we break the rule that this kind of information is not released for 100 years, there will be no similar income stream from the NHSCR records. The only obvious source of funding is from the NHS. It seems to us a disproportionate diversion of resources from patient care.

Other approaches

7. Theoretically, we could simply photocopy the records. In practice, however, that is the same option as we have already explored: modern photocopiers simply create digital images.

8. The real alternative to creating a digital back-up is to continue to keep the paper records safe. We have successfully done so since 1939 and there is no reason to expect that the task will become more difficult. The building at Dumfries in which they are currently kept, and the NHSCR's new permanent home there, are properly constructed and maintained and are not specially vulnerable to fire or flood. During the planning of the NHSCR's permanent home, care was taken to position the water pipes in the ceiling cavity so as to minimise the risk of water spillage in the storage room where the registers will be housed.

Recommendation

9. The Board is invited to consider whether the creation of a digital back-up to the NHSCR's paper records would offer good value for NHS funding, or whether there is an alternative feasible approach.

**General Register Office for Scotland
May 2006**

NHSCR Scanning Estimate

Background

For NHSCR there are five elements to consider:

- The imaging of the books
- The creation of a computer index
- The retrieval software
- The hardware required for the images and index.
- Staff required to quality review and load the image and index data.

Recent exercises have shown that there can be substantial differences in the estimates received from suppliers. This is particularly so with imaging where there have been variances in estimates of up to 100%. Most index creation is now done off-shore and we have found that by dealing with off-shore suppliers direct can result in considerable savings.

As the NHSCR books are in daily use off-shore scanning is not an option and it would be preferable if it could be done on-site. For an exercise of this size it could be made a condition of the contract that quotes for on-site and off-site be submitted. A recent tendering exercise found that the difference in cost did not preclude on-site as a viable option.

The software required is a relatively standard search and retrieval program and a facility to update and amend the index data.

Staff will be required to check the completeness and the quality of the image and index data received from the supplier(s).

Conclusion

I have attached information on the various options including costs and the reasoning for choosing a particular option.

I consider the following to represent a reasonable solution:

Imaging	363,300 images at £1.60 per page	£581,280
Indexing	Contracted off-shore	£160,000
Software	Contracted off-shore	£3,000
Hardware	To support good quality working images	£12,000
QR Staff	B1 and 5 A1 for 3 months	£41,650
TOTAL		£797,930

Ron Bell
September 2004

Cost Calculations

The preferred options as included in the conclusion on page 1 are in dark red/bold font

Cost of Imaging

Number of books	2,076		
Average number of pages	175	Maximum number of pages	200
Average Total pages	363,300	Maximum Total pages	415,200
Cost at £1.235 per page	£448,676		£512,772

Note: The NHSCR entries cover the 'open' two pages of the book and would have to be captured as a single image. The £1.235 was for single page capture I would therefore expect to pay more for a double page capture. However this is a much larger exercise for which I would expect a more favourable rate to be charged. Judging from past experience I would anticipate costs to be in the range £1.40 to £1.60 per page

Cost at £1.60 per page	£581,280		£664,320
------------------------	-----------------	--	----------

Cost of Indexing

Total pages (average)	363,300	Total pages maximum	415,200
Entries per page	46		
Characters per entry	38	(see below)	38
Cost at 25p per 1,000 char	£160,000	(off shore)	£182,000
Cost at 96p per 1,000 char	£610,000	(UK)	£700,000

Index Format:

<i>Field</i>	<i>No of Characters</i>
Forename	11
Surname	7
Sex	1
Date of Birth	10
Volume Number	4
Page Number	3
Keying Characters	2
TOTAL	38

Notes:

The average for the name fields were calculated from a sample of the DIGROS index.

Some companies charge for control characters e.g. end of line

Cost of Software

In our recent exercise the cost for a retrieval system ranged from £1,000 to £30,000 with other bids in the £15,000 region. This had a much smaller base of data and I would therefore anticipate that the lower end cost would rise to the region of **£3,000** but there should be no difference in the other totals.

Cost of Hardware

Number of images	420,000		
Size - Good quality working	1Mb	Disc space required	600Gb
Size - Medium quality working	600kb	Disc space required	350Gb
Size – Archive quality	4Mb	Disc space required	2 Tera Bytes

Note: The disc space required allows additional capacity for contingency and recovery.

For the first two options a good quality but standard server would suffice. To include back up and power failure features this could be obtained for around **£12,000**. For archive quality images we enter a whole new ball game and would be looking at spending at least £50,000 to £80,000.

The advantage of the medium quality image is that the image will load quicker however given the original size of the page, that the data is handwritten and as the images are for internal use by a limited number of users the better quality image is the more suitable option.

Quality Review Staff

To review some 400,000 images, remove duplicates, identify missing pages and load the images onto the server. At an average of 300 images per hour a 100% check would take approx 200 man days. As this is a one off exercise and the images will be used as the source for creating the index I would recommend a 100% check.

There are two parts to reviewing the index data; checking that the supplier has correctly completed the work as per the contract and then completing the items which could not be transcribed by the contractor. There will be some entries which due to poor writing, poor quality image or various other reasons cannot be transcribed by the contractor. As the standard reviewing of the contractors work will have to be completed within an agreed period following completion of the transcription (usually 4 to 6 weeks) this latter work is sometimes undertaken as a separate task.

There will be approx 17,000,000 index entries to be reviewed. Given that the entries will have already had a review by the contractor (I would recommend that as a minimum double keying of entries is undertaken) we do not have to undertake a full review of the index. However the review and completion of the missing data is a considerable piece of work. The best guide I have to this is the work required to check and complete the 1841 to 1871 Census indexes. Given that the majority of the time is in completing the unreadable entries and that by capturing NHSCR data from

the original document will provide a good quality images my estimate is that a team of 6 (1 B1 and 5 A1/3 or casuals) for 3 months will be required.

Estimated Staff Costs:

Imaging:	B1 + 5 casuals for 2 months
Indexing:	B1 + 5 casuals for 3 months
Total:	B1 + 5 casuals for 5 months
	= £11,650 (5 x £2,330) + £30,000 (5 x 5 x£1,200)
	= £41,650