SOUNDINGS.......

"Recovery of Digital Data"

As a result of some experiences with the publication of *LIGHTHOUSE*, edition 59, I would like to share with you some thoughts about digital data. The experiences were somewhat painful at the time, but in the long run perhaps they will be of benefit to those in the hydrographic and other communities who collect and store large amounts of digital information. I recognize that my concerns may not shared by all and will perhaps be considered by those in the business of archiving digital data to be superficial and with little substance. However, my recent discussion of this topic with several individuals, convinces me that the topic deserves some serious thought.

The "theme" of this article is digital data or digital information and it's archival and subsequent unaltered retrieval.

As a "field" hydrographer I collected digital data, including soundings, as far back as 1972 (others were involved earlier than I). The digital soundings collected were presented in "traditional" field sheet format and, as well, a copy of the digital data was archived, presumably with the intent that it would and could be retrieved again at a later date. That data was collected over 30 years ago and since then a lot of digital information has been collected, processed and archived in much the same manner. Modern survey systems collect huge amounts of digital information. Can all of that data that has been archived be recovered and presented in its original form? Most believe it can – but can it really?

Rigid procedures and standards have been introduced over the years and considerable effort has been expended to ensure that the digital data being archived is in fact "good" data. As I recall, data was measured as being "good", if the digital file was identical to some printed hardcopy. In the case of "field sheets" the digital files had to be certified as being identical to the printed hardcopy (if there were exceptions, they were noted).

The data printed on the field sheet from the digital file was used as a measure for the quality of the digital file. The main concern was that data had to look good on the printed sheet. Nothing is wrong with that - but there are other considerations which may be problematic.

Let's return to *LIGHTHOUSE* edition 59 and see how these issues are related and why I believe this story needs to be told.

As edition 59 was being finalized, we had a digital file on CD and a hardcopy produced from that file. This was relatively the same situation as a file of digital soundings and a printed hardcopy of that data, which we call a field sheet.

The hardcopy of the digital *LIGHTHOUSE* file was used for final proofreading. As expected, errors were found and corrections were made. We had to ensure the digital file was exactly the same as the corrected hardcopy.

When we were satisfied that the digital file was exactly the same as the "good" hardcopy, we went to the printer. In one hand we had a good looking hardcopy and in the other a digital file from which that hardcopy was produced. We had to be happy - right!

Wrong! We admit to making a few costly rookie errors, but more importantly, our fundamental thinking was flawed. We believed that all necessary quality control was in place as the digital file and the hardcopy were identical. We assumed that any and all subsequent hardcopies would also be "good".

By now you have probably correctly concluded that our copy from the printer was not identical to our hardcopy from the same digital file. Some of you have seen the results and they were not pretty. Hopefully we have learned a good lesson and will not repeat our errors. Life is full of learning experiences and perhaps we can all benefit from this exercise.

With *LIGHTHOUSE* edition 59, our main problem after the initial printing was a font change. The printed text was difficult to read because of little spacing between words. A more significant problem, which could not be explained by the printer, (or anybody else) was that some text was simply omitted. When the same file was subsequently printed with different software, most of the missing information magically appeared. Real scary!!

Different computer programs can and do produce different results. Different versions of programs can result in different output. If I print this Word file with a Mac or a PC, I can get something that may look different.

I write about this issue so those responsible for digital information are aware of possible problems. This is not new - but sometimes we need to be reminded. Don't assume that your data is safe because of whatever reason.

Many of these concerns can perhaps be explained by some of our experts. However, we are not all experts and most of us believe what we see.

Can all offices responsible for the storage of digital information recover that data so that it looks exactly the same way it was when it was filed?

Do I believe that the sounding information archived from the survey I conducted in 1972 can now be routinely recovered and used to produce a hardcopy identical to the one that was used as a quality control document?

With the computer world and that of digital information as well as software and hardware evolving so very quickly, I'm not sure that collectively we have given enough thought to the painless recovery of error free digital information at any point in time down the road.

Your comments are invited.

Earl Brown
Hydrographer (retired)