

A bit of a problem



Examples from CASPAR project <http://www.casparpreserves.eu>

The valuable digital assets of many cultural and scientific institutions across Europe represents the cumulated knowledge of generations. However they are increasingly vulnerable and at risk of being no longer accessible, given the fast evolution of computer technology.

We need to address the issue of how digitally encoded information can still be understood and used in the future when the software, systems and even everyday knowledge will have changed.

Things we take for granted now would otherwise become incomprehensible, something to be guessed at, even if we preserve the bits and bytes. In addition to benefiting future generations, immediate advantages result from doing the preservation right - it enables people to find and make use of unfamiliar data generated today, thereby

extracting more value from the investment.

Many digital objects are mainly “simply” rendered, in other words displayed on a screen or played through loudspeakers. Even this can be a challenge after a few years – just think about documents created by your favourite word processor a decade ago, or your spreadsheet, or your email.

In currently planned and future experiments, several orders of magnitude more scientific data will be generated than has been collected in the whole of human history. This deluge of data is matched by the blizzard of digital footprints which track each and every one of us through our lives and beyond, such as our health records or bank

transactions. These need far more than just rendering. We need to be able to process them, but even if we can extract numbers from the bits that make up a file, how can we be sure what they mean – for example was the measurement in feet or metres or miles? – what does it mean?

Some data sets are relatively cheap to collect, others such as those collected from space, are very expensive. Much can never be repeated and if they become inaccessible or unintelligible, how can we be sure, for example, how much the climate might have changed or how far our society had changed over the years.

Solution?

One simple but impractical solution is to supply endless amounts of money to pay talented people to care for the data generation after generation.

Clearly we need to keep the “bits”; the cost of storage (£ per gigabyte) is declining but volumes are growing, generations of hardware and software technologies pass increasingly quickly, and no funding is guaranteed. The chain of preservation is only as strong as its weakest link.

How can the costs be controlled?

Perhaps the most practical is to arrange to share the costs wherever possible, and to plan from the start for the long-term.

How can we know if we are doing the right things? The truth is that there are no guarantees but efforts are underway to produce a standard and a process to give us some confidence by certifying archives as trustworthy.

There are very many projects tackling aspects of digital preservation being funded by the Research Councils, JISC, the European Community, the USA and many other countries around the World – CASPAR is one of these. What we need to identify is the common support infrastructure which will help to share costs of preservation of the deluge and then be prepared to support that infrastructure over the long term.

For more information please contact David Giarretta (Director of the CASPAR project) via email at d.l.giarretta@rl.ac.uk or telephone 07770326304.