

DAO VICTORIA



Rescue old data before it's too late

If we don't act soon to preserve past records, invaluable knowledge will be lost, warns Elizabeth Griffin.

In the late nineteenth century, astronomers began to photograph stars using prisms and gratings. They recorded stellar spectra — the dispersal of starlight into colours — to learn what the stars are made of. Since then, those photographic plates have become useful for another purpose: they let scientists map past concentrations of ozone in Earth's stratosphere, and help to reveal whether some changes to the ozone hole are natural. The hardest part is getting hold of these glass plates. I know, because I spent many weeks going through collections at observatories across the world, from Germany to Australia, to search for them.

What other historic data could be useful? Tales abound. The thousands of logs recorded during ship voyages in past centuries are a bonanza for studying weather patterns today. Photos of glaciers from the past and the present have startled the world, and yielded incontrovertible evidence of climate change. Medical records on dusty punch cards, abandoned in the late 1950s and decoded decades later, have helped to show how varying levels of cholesterol predict later disease.

To model the future, we need to be able to examine the past. But our chances to do so are fading fast, sped by misunderstanding and negligence. Few forms of 'heritage data' — whether stored on glass plates, paper, old tapes or floppy disks — are easily available for today's research, so the information on them is effectively lost.

Scientists used to complain that they could never obtain enough data. Today, we speak of Big Data as if it were an untameable beast. Measurements collected now are increasingly sophisticated, but they tell us only about the present. Measurements recorded long ago can show us how Earth's weather, ecosystems and more are changing, and data taken from individuals in decades past can inform modern medical and policy guidelines. If we want those data, we need to start recovering them now.

Why aren't scientists from all domains scrambling to preserve old records, the better to study long-term trends? Part of the answer is human psychology. At one talk I gave on the need to bring astronomy's near-lost data into lasting, easily shared formats, an audience member challenged the effort. "Modern data are so much better," he said.

He missed the point. Few want to poke around musty archives for heritage data captured using yesterday's technology, but these provide information not available in any other form. Hydrologists in Cape Town, South Africa, have converted 70-year-old, handwritten stream data to deduce how non-native tree species affect water distribution across a landscape. High-resolution, full-colour photographs of extant birds cannot replace images of extinct passenger pigeons and laughing owls.

The time is ripe to rescue heritage data. In many cases, the original scientists are still alive to provide context. Technologies for digitizing

many sorts of records are cheap and convenient.

Digitization will not preserve everything. At least one epidemiologist has tracked the spread of cholera in the Iberian Peninsula by sniffing envelopes. How? For centuries, post offices used vinegar to disinfect outgoing mail from afflicted towns, and the smell has persisted.

So, what can be done? The Data Rescue Interest Group, part of the international Research Data Alliance, offers guidelines (go.nature.com/2pgzkfs) to steer a researcher through the initial stages of rescuing data, determining the equipment needed and deciding how best to tackle the rescue. The most important data capture conditions from before large-scale human changes were felt. Fields such as biodiversity (<http://rebind.bgbm.org>), volcanology and oceanography have made strides in preserving old data, but more needs to be done — soon, and with better coordination.

We will not be able to save all data. Prioritizing means looking for the potential to illuminate questions that could not be answered otherwise. Too often, researchers dismiss heritage materials without considering what uses they might have. Treasure troves of data, and the knowledge they could offer, are left mouldering on shelves.

Everyone can help. The first challenge is to locate records, photographs or other items, or simply to recognize their value. Most have not been used for yonks, and are stored in some almost-forgotten location where damp, spiders and mice are probably doing their best to destroy them.

The second is to ascertain that the necessary metadata (such as date, location and limitations)

are available, so that when data are converted into modern formats, they can be assigned accurately to time and place.

Finding the resources for preservation is often difficult. Funding is sparse and erratic, but enthusiasts have secured grants from agencies ranging from NASA to the US Agency for International Development and the German Research Foundation. It is worth casting a wide net. University archivists can supply expertise. Citizen-science groups have also been mobilized.

One overlooked resource is success stories. When researchers learn of once-neglected data that have been revived and transformed into modern insight, they themselves are more likely to recognize hidden opportunities. The next heroic rescue tale could be your undertaking.

But hurry. Some data are decaying as I write, some will have gone past retrieval by tomorrow, and the ageing memories needed to make them meaningful might not be available much longer. ■ **SEE NEWS P.278**

**TREASURE TROVES
OF DATA, AND THE
KNOWLEDGE
THEY COULD
OFFER, ARE LEFT
MOULDERING
ON SHELVES.**

Elizabeth Griffin is an astronomer at Canada's National Research Council Observatory in Victoria, and co-chair of the RDA Data Rescue Interest Group
e-mail: elizabeth.griffin@nrc-cnrc.gc.ca