

ARCHIVES

Researchers launch hunt for endangered data

Global effort will catalogue information languishing in drawers and basements.

BY LINDA NORDLING

Around the world, key scientific data are at risk of being lost, either because they are held on fragile or obsolete media or because they may be destroyed by researchers who are unaware of their value. Now a team of scientists is planning to scour museums and research institutes to draw up a global inventory of threatened data. Launched on 29 October, shortly after the biennial conference of the Committee on Data for Science and Technology in Stellenbosch, South Africa, the project aims to publish the inventory online in 2012.

The effort should enable future rescue programmes to target the most valuable and threatened material, and it should also alert researchers to resources of which they were unaware. Climate-change studies, for example, require data series on temperature and rainfall reaching back further than digital records. Some scientists are having to leaf through old ships' logs for clues to past weather patterns.

It is not only data in old books or photographs that are at risk. Digital information collected between 1950 and 1980 is also threatened, because it is stored on outdated media often subject to deterioration, such as magnetic tape and floppy disks, making it increasingly difficult to access and retrieve. Developing countries host some of the most valuable data — from land use to disease statistics — and the most threatened, for example by local conflict and inadequate or patchy storage. The task group hopes to catalogue such data by simply asking scientists, by e-mail and telephone calls, what records are lying in their vaults.

"I think the CODATA initiative is a very worthy one," says Sam Pepler, curation manager at the British Atmospheric Data Centre in Chilton. "A lot of our data heritage is locked up in an analogue form. Long-term,

accurate records are not an invention of the digital age, they have been kept for hundreds of years and the people collecting the data expected us to use them."

The inventory is the brainchild of Elizabeth Griffin, an astronomer at the Herzberg Institute of Astrophysics in Victoria, Canada, and William Anderson, an information specialist at the University of Texas at Austin and an associate editor of the *Data Science Journal*. Both will serve on the task group with ten other scientists.

"As things stand, researchers are only vaguely

aware of the existence of non-digital repositories of data in their own disciplines," says Griffin, who has relied on non-digital data in her own recent work. From old glass plates recording astronomical spectra, she has been able to study historic ozone levels in Earth's atmosphere. "I had to go from observatory to observatory with a suitcase, rummaging around these plates and books," she says. She estimates that some 3 million glass plates storing information about the night sky are gathering dust in observatories worldwide. Each is an

irreproducible window on the past, she says.

Some disciplines already have projects to recover old data. The Global Oceanographic Data Archaeology and Rescue Project, for example, has collected more than 3.5 million sea-temperature profiles — some dating back more than 100 years — since the effort began in 1993. Most fields are not so well organized, however, and it is unlikely that all the at-risk data that the task group identifies can be saved for posterity. As data sets start to be measured in petabytes (10^{15} bytes), scientific institutes are struggling to manage the vast streams of new data generated by projects, without adding to their burden with legacy data.

But without knowing where the old data are, says Anderson, it is impossible to make rational choices about what to keep. "I think everybody knows that stuff is going to be lost," he says. "But it would be nice to be on top of it." ■



Millions of astronomical plates are just waiting to be rescued.

R. ASTRON. SOC./SPL

➔ **NATURE.COM**
For more on data
curation see:
go.nature.com/vnklyt