

A lesson in sharing

Earth scientists need better incentives, rewards and mechanisms to achieve free and open data exchange, says **David Carlson**.

When the polar-research community planned the International Polar Year (IPY) of 2007–08, it embraced a revolutionary goal: to establish free, open and ready access to all data. After decades of reports with ‘data’ and combinations of ‘integrated’, ‘interoperable’ and ‘distributed’ in their titles, the IPY presented an ideal test case — interdisciplinary but limited in duration and regional in focus. Yet the community found inadequate services, almost no international support and few solutions.

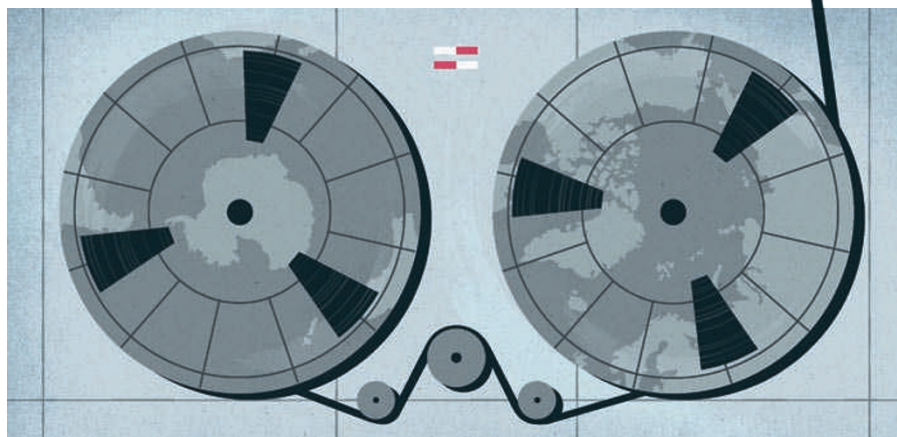
We have come out of the IPY with a rich burst of data, but the information uses the jargon and units of specialities from anthropology to astronomy, referenced to everything from Cartesian coordinates to postal codes. And despite the best efforts of the IPY Data and Information Service (www.ipydis.org), we cannot say how users might discover or access IPY data five years hence. Indeed, it emerged just last week that an upcoming report from the US National Academy of Sciences in Washington DC identifies the lack of data sharing as a barrier to understanding rapid changes in polar ecosystems (see *Nature* **469**, 145; 2011).

What caused these failures? Technical impediments exist relating to formats, permissions, bandwidth and so on, but the real problem is behaviour. The Earth sciences, like the science community as a whole, lack incentives for widespread data exchange.

Long before the film *Avatar* popularized it, I learned from engineering colleagues the whimsical but useful term ‘unobtainium’ — used to describe something perfect but elusive. A perfect data-sharing system is science’s unobtainium. We must respond creatively to the challenge. Steps begun as part of the IPY by the Earth-science community include establishing a polar information commons and instigating a journal for data publication and citation.

A LEGACY OF CONFUSION

The challenges of preserving and sharing IPY data have come up repeatedly. Meteorological data from the first IPY in 1882–83 emerged in digital, accessible form only during the planning of the latest IPY. Data from the 1932–33 IPY were scattered: some were rediscovered only in recent years at the Danish Meteorological Institute in Copenhagen. By the time of the 1957–58 International Geophysical Year, the International Council for Science (ICSU) was forming World Data Centres (WDCs)



JESSE LEFKOWITZ

to help solve the problem. There are now more than 50 WDCs, all of which pledged to support the latest IPY. Most struggled. Few received increased funding to respond to new or bigger IPY data streams, and the system had no mechanisms for handling the ecological or social threads of the IPY programme. The current WDCs, which have been supplemented by national and speciality data centres, cannot meet the needs of modern international interdisciplinary science.

The ICSU is establishing a World Data System to reform and reinvigorate the WDCs, and the World Meteorological Organization is upgrading its global information system. I endorse these efforts. But without fundamental changes to the incentives for data sharing, scientists will only perpetuate bad habits.

Data centres depend on willingness to share. All IPY projects opted in to an explicit free and open data-sharing policy (go.nature.com/byf9b4). But many researchers do not recognize, much less comply with, this policy, and few national funding organizations have the motivation or means to enforce it. Many researchers worry about others ‘stealing’ or misusing their data, and so hoard them.

To circumvent these attitudes, a team including myself developed the concept of a ‘Polar Information Commons’ (PIC; www.polarcommons.org) data label in 2009. PIC data can be freely accessed and used according to voluntary rules on attribution, citation and recognition, version control and notification, and appropriate use. The idea has been favourably received in the scientific community, and a pioneering group of polar data

centres in Australia, Canada, Japan, Norway, Britain and the United States have indicated their support. However, when it comes to the nitty-gritty of making data fully available, the PIC often stalls in institutional or national legal departments. The collection of PIC-labelled data is growing — but slowly.

Another effort is the *Earth System Science Data (ESSD)* journal, which I started with

“A perfect data-sharing system is science’s ‘unobtainium’.”

Hans Pfeiffer-berger, head of IT infrastructure at the Alfred Wegener Institute for Polar and Marine

Research in Bremerhaven, Germany. It publishes complex and comprehensive data sets, giving data providers credit just as for a traditional publication. The journal uses a Creative Commons copyright policy to encourage free use of articles as long as the original authors and citation details are identified, and insists that authors deposit their data in a well-known open-access repository of their choice. It is a small effort so far — ESSD has published 28 data sets since its first issue in 2009, and remains unique in Earth sciences. We hope it will inspire similar projects.

The IPY crystallized our view of the unobtainable ideal, but hinted at solutions. The grand vision of the IPY ran aground on practicality and pragmatism, so we must take practical steps to change behaviours. ■

David Carlson served as director of the IPY International Programme Office. He is now education and outreach director for the non-profit geodesy consortium UNAVCO in Boulder, Colorado.
e-mail: ipy.djc@gmail.com

► NATURE.COM
For more on data-sharing mechanisms and practice, visit: go.nature.com/kpfn1z