

Establishing an environmental-extension system at the township level could also help to prevent overuse of fertilizers and pesticides.

Peng Gong, Lu Liang and Qiang Zhang *Institute for Global Change Studies, Tsinghua University, China.*

penggong@tsinghua.edu.cn

Data archiving is a good investment

We have found that ongoing financial investment in data-archiving infrastructure yields an impressive scientific return, and believe that it should be wholeheartedly supported by research funding agencies (see, for example, go.nature.com/nzftf3).

We used Dryad (see <http://datadryad.org>), an international, open, cost-effective data repository for the biological sciences, to estimate the cost of archiving data from more than 10,000 publications. We found that these could be curated and the data preserved at an annual cost of about US\$400,000.

As an example of how much research is typically published per grant dollar, core grants in population and community ecology from the US National Science Foundation averaged 3–4 publications per \$100,000 of grant between 2000 and 2005 (S. Reyes, A. Tessier and S. Mazer, unpublished results). That is, \$400,000 invested in original research resulted in about 16 papers.

Dryad cannot yet tell us how effective data archives are in facilitating primary research publications, but the Gene Expression Omnibus (GEO) database at the US National Center for Biotechnology Information offers some insight. To estimate data reuse, we searched the full text of articles in PubMed Central for mention of any of the 2,711 data sets deposited in GEO in 2007. We excluded articles whose authors' names overlapped with those depositing the data set. Extrapolating the 338 hits in PubMed Central to all of PubMed, we estimate that the

GEO 2007 data sets made third-party contributions to more than 1,150 published articles by the end of 2010, and reuse continues to accumulate rapidly (H. A. Piwowar, T. J. Vision and M. C. Whitlock *Dryad Digital Repository* doi:10.5061/dryad.j1fd7; 2011).

Assuming that Dryad has a comparable rate of reuse and collects at least 2,500 data sets annually, an investment of \$400,000 in one year should contribute to more than 1,000 papers in the next four years — far more than the accepted value for a research dollar.

Heather A. Piwowar *Dryad, and the National Evolutionary Synthesis Center, Durham, North Carolina, USA.*

hpiwowar@nescent.org

Todd J. Vision *Dryad, and the University of North Carolina, Chapel Hill, North Carolina, USA.*

Michael C. Whitlock *Dryad, and the University of British Columbia, Vancouver, Canada.*

Noisy oil exploration disrupts marine life

Fossil-fuel operations in the Arctic will inevitably compromise habitat — regardless of spills (*Nature* **472**, 163; 2011).

The seismic airgun surveys used for hydrocarbon exploration and for monitoring deposit conditions can disrupt foraging behaviour of bowhead whales at long distances. They also seriously diminish fisheries catches of haddock and other Arctic species, and have halted the migration of fin whales (a non-Arctic species) at a range of more than 175 kilometres.

Bowhead and beluga whales avoid oil-derrick operations. Many other noises associated with fossil-fuel exploration and production — such as construction, shipping, transport helicopters and underwater acoustic telemetry — have a deleterious impact on the marine acoustic environment.

We do not yet know about the impacts of noise from thruster-stabilized exploration

platforms and from sea-floor processing equipment such as wellhead chokes, separators and re-injectors, which operate out of sight and under extreme pressures.

In collaboration with the World Wildlife Fund and the Natural Resources Defense Council, we are developing a peer-reviewed website that can be understood by a lay audience in order to explore some of these issues (see go.nature.com/5vuebe).

Michael Stocker *Ocean Conservation Research, Lagunitas, California, USA.*
mstocker@msa-design.com

Address education inequality in India

Narrowing the educational achievement gap between different social groups in India remains a major challenge, despite 60 years of affirmative-action policy (*Nature* **472**, 24–26; 2011). Using publicly available data from the country's top medical school, the All India Institute of Medical Sciences (AIIMS), we found that performance was poor among students admitted under a government scheme for socially disadvantaged groups.

All government and government-aided institutions in India allocate a fixed percentage of places on educational courses to socially and economically disadvantaged students. But in 1995–2005, out of more than 600 indigenous tribes with access to such positions, one small group from northern India accounted for 36% of students admitted to the AIIMS.

Between 1998 and 2006, socially deprived students accepted into the AIIMS scored 13.6% less in the entrance exam than students from non-disadvantaged social classes ($P < 0.001$). In 1989–98, such students also had double the dropout rate of non-disadvantaged students (6% versus 3%; $P > 0.05$). In the ten years for which data are available (1995–2005), 61.4% of students admitted to

government-reserved positions had to resit examinations in at least one subject, compared with 15.2% of non-disadvantaged students ($P < 0.001$).

To address such inequality, India should adopt measures that have proved successful in other countries. These include wider access to quality primary education; standardized assessment of students; and academic support for students who are lagging behind. More research to assess this inequality is also needed to inform education policy.

Manas Kaushik, Subha Ramani *Boston University, USA.*
mkaushik@post.harvard.edu

Fund experiments on atmospheric hazards

The radioactivity released from Japan's damaged Fukushima Daiichi nuclear power plant has increased the urgency to fund tracer experiments that will improve models of atmospheric dispersion and reinforce confidence in emergency procedures.

The last major tracer experiments were conducted in the mid-1990s. So the predictive capabilities of current atmospheric-dispersion models have not been properly tested, hindering their evaluation and development.

To generate more observational data, multiple-scale atmospheric tracer experiments should use non-hazardous, climate-neutral substances and a realistic release term with varying source strengths. Modellers could estimate emissions in real time using a limited set of observations without knowing the actual release rates, and later improve their models and data-reconstruction methods on the basis of the real source terms and measurements.

Stefano Galmarini *European Commission Joint Research Centre, Italy.* **Andreas Stohl** *Norwegian Institute for Air Research, Norway.* **Gerhard Wotawa** *Central Institute for Meteorology and Geodynamics, Austria.*
gerhard.wotawa@zamg.ac.at