on seed chemical composition that we could access from USDA[10]. Mining data from floras, faunas and virtually any other retrievable source can be fun and often leads to simple, yet important patterns[11]. Finally, as my colleague Ganeshaiah demonstrated, you can get cracking by mining data from far and wide, from cricket to stock markets, without ever having played cricket or engaging the bulls (the stock brokers)[12,13].

Gone are the days of coal hunters and coal mining or for that matter gold hunters and gold mining. Let us unashamedly and without loss of much more time enter the age of data hunters and data miners. And why not, considering that data hunting and mining is environmentally safe?!

1. Balaram, P., *Curr. Sci.*, 2000, **79**, 1511–1512.
2. Singh, K. S., *Curr. Sci.*, 1993, **64**, 5–10.
3. Singh, K. S., *People of India: An introduction*, Seagull Books, Kolkata, 1992.
4. Conventionally, the term 'data mining' has been used to refer to the process of treating large sets of data to large-scale data analysis and discovering patterns in them. However, in this article I use the term to also include the process of collating widely dispersed and seemingly disparate data onto a common platform and then examine them for certain underlying patterns.
5. Ganeshaiah, K. N. and Uma Shaanker, R., *Curr. Sci.*, 1998, **75**, 292–298.
6. Ganeshaiah, K. N. and Uma Shaanker, R., *GIS@Development*, 1999, **III-V**, 67–69.
7. Sen, N., *Curr. Sci.*, 2000, **79**, 1046.
8. There is an inherent bias among students that their thesis should contain new (original) data. In fact my own graduate student was extremely uncomfortable with the fact that three of the four chapters in his thesis were based on data mining, while all his friends reported original data. He seriously doubted if his thesis would be approved by the external examiner. The fears of the student might not be completely unfounded. External examiners and reviewers of manuscripts tend to regard papers that do not contain significant amount of primary data as soft work and thus such papers run the risk of a high rejection rate.
9. Sarma, J. A. P., Nangia, A., Desiraju, G. R., Zass, E. and Dunitz, J. D., *Nature*, 1996, **384**, 320.
10. Lokesha, R., Hegde, S. G., Uma Shaanker, R. and Ganeshaiah, K. N., *Am. Nat.*, 1992, **140,** 520–525.
11. Uma Shaanker, R., Ganeshaiah, K. N. and Ravishankar, K. V., *Curr. Sci.*, 1997, **73**, 646–647.
12. Ganeshaiah, K. N., *Curr. Sci.*, 1992, **63**, 345–347.
13. Ganeshaiah, K. N., *Deccan Herald*, Science and Technology Supplement, 18 May 1999.

R. UMA SHAANKER

Department of Crop Physiology,
University of Agricultural Sciences,
GKVK,
Bangalore 560 065, India
and Jawaharlal Nehru Centre for
Advanced Scientific Research, Jakkur,
Bangalore 560 065, India
e-mail: rus@vsnl.com

Towards a biodiversity information network for India

Convergence of research in modern biology and informatics is a reality now that genomics has become an accepted branch of scientific research[1]. The recent announcement of the completion of the sequencing of the *Arabidopsis* genome, while a significant milestone in modern biology research by itself, is also notable for the rapidity with which results have been made available to the community of researchers via the World Wide Web[2].

The transformation of outputs of modern biological research into easily dispersible and transferable digital data is thus now fully functional at the molecular level. At the levels of species and ecosystems, such a convergence is still incomplete. However, significant initiatives are under way in applying informatics at the species level. The Global Biodiversity Information facility (www.gbif.org) and the project Species2000 (www.sp2000.org) are notable efforts to build global gateways for taxonomic data in digital format held in collections located in different countries and continents. The challenge facing them is one of overcoming the limitations posed by the existence of widely different and occasionally incompatible digital formats used by various database system[3].

These initiatives are spearheaded by organizations located in countries that are members of the organization for Economics Cooperation and Development (OECD). At the international level, the World Conservation Monitoring Centre (www.unep-wcmc.org) and the clearing-house mechanism of the Secretariat of the Convection on Biological Diversity (www.biodiv.org/chm) provide detailed information on conservation and policy issues, but their emphasis is not on providing in-depth data on biological species.

In India, large collections of herbarium data exists. However, they are not readily available in a digital format. The Botanical Survey of India (BSI) has long pursued a project for the computerization of its accessions, but the digitized data are not yet available to public. With the passing of the Biodiversity Conservation Bill, a number of individuals and organizations connected with administering the national sovereignty over biological resources will be required to access species level information frequently. The Plant Variety Protection and Farmers' Rights Bill, when passed, will also create a need for access to varietal information.

There is thus a need to establish a digital network for species level information in India, which is the easiest and surest way to enable access by individuals and institutions connected with implementing the provisions of the above-mentioned Bills. Centralized databases such as the one under construction by the BSI are slow in coming. When such databases are constructed, the problem of incompatible formats cannot be ruled out, as seen with the OECD initiatives mentioned above. There is a clear need to avoid such bottle-necks.

One strategy would be to allow a variety of individuals and organizations to create and share or digitize and share species level data using the World Wide Web and through use of the technique of peer-to-peer file sharing. Creation of a minimum common format is essential for this purpose. It can be designed on the basis of consensus among the major national organizations such as the Botanical and Zoological Surveys of India, the National Bureau of Plant Genetic Resources and the allied bureaus, and the National Bioresources Board. Data can be either input

afresh into the common format or transformed from an existing digital format. This process need not be centrally directed and can be carried out at the level of each participating individual or organization. The technique of peer-to-peer file sharing brought to larger public use by Napster (www.napster.com), can be used for searches and downloads and is useful in setting different privileges of access. This will help overcome the technological limitations that developing countries are facing in applying informatics to biodiversity[4].

The emphasis here is on multiplicity of institutions and individuals holding and sharing data in a common minimum format, rather than centralized databases that have a tendency to develop mutual incompatibility. The size of the data holding or the institution does not matter in this approach and that is likely to contribute to its success. India is in a position to promote and use peer-to-peer data sharing for biodiversity databases. If implemented successfully, this can become an international model, similar to GBIF or Species2000.

1. Pennisi, Elizabeth, *Science*, 2000, **290**, 2220–2221.
2. Rhee, Seung Yon, *Plant Physiol.*, 2000, **124**, 1460–1464.
3. Edward, James, L., Lane, Meredith, A. and Nielsen, Ebbe, S., *Science*, 2000, **289**, 2312–2314.
4. Zhang, Y., Zhang, L. and Zu, D., *Science*, 2000, **290**, 2073.

ACKNOWLEDGEMENT. We thank Dr Francesco di Castri and Dr James Aronson for comments on the subject.

V. BALAJI

Centre d'Ecologie Fonctionelle et Evolutive (CEFE), CNRS, Montpellier BP 9056, France
e-mail: balaji@cefe.cnrs-mop.fr

Significance of impact factor with regard to mathematics journals

We read with interest the note of N. C. Jain (*Curr. Sci.*, 2000, **79**, 1513–1514) on the findings of Science Citation Index (SCI). We would like to congratulate the Editor and the Editorial Board on the high impact factor of *Current Science*, though it is hardly a surprise as *Current Science* has been popular with the scientific community over the last several years.

For a journal like *Current Science* with its coverage of a wide range of scientific disciplines, the impact factor (IF) of SCI, no doubt, serves as a good indicator of its success in realizing its objectives. However, we would like to draw the attention of the readers to the fact that this IF may not have the same significance for other journals with a sharper focus, devoted to specific disciplines; this is particularly so with regard to mathematics.

The IF of SCI for a given year is the average number of times 'recent' articles from the journal were cited during that year; 'recent' here means that the article has appeared within the two preceding years. In fields where rapid obsolescence is a major factor to reckon with – and this is often no doubt the case, especially with fields of applied science – it may be reasonable to measure impact by such recent citation. But in mathematics, where 'classic' works of a Gauss, a Riemann or a Ramanujan continue to inspire current research, it is not a good gauge for quality. Titles in bibliographies in most mathematics papers, especially from prestigious journals like the *Annals of Mathematics*, *Inventiones Mathematicae*, etc. often date back to 20 or 30 years (and occasionally even more); recent papers (dating back to not more than 2 years) make up barely 5 to 10 per cent of the references.

In many disciplines, a few dominant themes become the focus of attention of a large number of researchers, for relatively short periods; however this does not happen in mathematics. Different branches of mathematics go through a slow and steady development and these diverse streams meet time and again and spread out further. In such a context it is hardly surprising that most mathematics journals have a low SCI IF. Often the papers which get cited within a short span are those which are incomplete in one respect or the other, so that either the author, or someone within the narrow area around, is able to say something more about the topic soon after. Thus the two-year reference window is somewhat arbitrary, and especially in the specific context of mathematics, may well exaggerate the importance of work of indifferent quality. This makes it clear that the IF does not serve to make meaningful comparisons of quality between mathematics journals, nor do changes in the IF of a mathematics journal over a period of time have a serious meaning.

For the Mathematical Science Proceedings of the Indian Academy of Sciences, for example, the IFs for the five years from 1995 to 1999 were computed to be 0.154, 0.143, 0.184, 0.149 and 0.048, respectively (only the last one has been quoted in the note by Jain cited above). During the six years from 1993 to 1998, the journal published 25, 53, 45, 31, 37 and 26 papers, respectively. The values of the IF as above therefore mean that during 1995 to 1999 the papers from the preceding two years were cited 12, 14, 14, 10 and 3 times, respectively. Seen in raw figures the information means much less than what the variation in the IF would suggest; in such a small sample, simple variations due to the specific nature of some papers (including good papers) can change the averages drastically.

Thus the IF of SCI cannot be taken seriously for a calibration of the quality of mathematics journals: we have a little doubt that professional judgments of quality will be at considerable variance with what the IF may suggest. It is not our case that, judged by other more appropriate criteria, Indian mathematics journals will come out with flying colours. Many of our journals are far from being serious competitors internationally. However, the situation is significantly better than what the IF would indicate.

M. S. RAGHUNATHAN
V. SRINIVAS

School of Mathematics, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005, India