DSC 2001: Cooperation between programs (and programmers).

Thomas Lumley

The second international conference on Distributed Statistical Computing was held in March 2001 in Wien, Austria. Hosted by the Center for Computational Intelligence of the Vienna University of Technology (Technische Universität Wien), it was substantially larger than the first Vienna DSC conference held two years earlier. There were about 30 presentations over the three days of the conference, with speakers from universities, industrial research groups, and software companies in twelve countries. Despite the programmed sessions' running about 30% over time there was lively discussion over coffee and lunch and in during the conference's social program.

DSC 2001 emphasized the value of high quality open source software as a statistical research platform but also for the ultimate goal of delivering modern statistical methods to those with little knowledge of (or interest in) statistical computing.

One clear theme of the conference was the leverage to be gained by joining together specialized tools. Statistical computing environments (most speakers preferring dialects of the S language) need to handle large data sets: databases can do this. They should provide interactive dynamic graphics for displaying complex, high-dimensional data: programs such as GGobi and Orca do this. The must supply familiar interfaces, from menus and dialog boxes to web pages and spreadsheets, and again there are both free and proprietary programs that do this.

Now that modern programming languages, and applications from Excel to Oracle, make much of their functionality available to external callers it is unnecessary for statistical environments to duplicate these features. The traditional retort to unnecessarily complicated feature requests, based on Dennis Ritchie's "If you want PL/I you know where to find it.", becomes much more helpful when you can link "it" in to what you currently use. As always, this works better in theory than in practice, but there is now real hope that the problems can be solved.

Another theme of the conference was the value of object-based programming, partly in inter-process communication but also in static and dynamic graphics, documentation, and statistical modelling. This is hardly new in the computer world, but statisticians have not traditionally thought in OOP. Some of these uses, such as a more flexible notation for statistical models, will require more understanding of the benefits of an object-based approach by the wider statistical community. Other uses will be invisible to users, apart from the increased convenience they will produce.

Literate Programming, and its new cousin, Literate Data Analysis, made an appearance, and it may be that tools will soon (finally) be available to make these more than intellectually interesting. They are among the many areas that may benefit from the use of XML and XSL to store and manipulate text, code, and data.

These proceedings cannot capture the full content of DSC 2001, but we hope they will be a useful resource for statistical computing users and researchers.