Empowering African scientists: an investigation into a CD-based installer for scubuntu

P VAN ZYL AND T FOGWILL
Meraka Institute, CSIR, PO Box 395, Pretoria, 0001, South Africa
Email: pvzyl@csir.co.za, tfogwill@csir.co.za - www.csir.co.za

CORE IDEA

Ubuntu is a popular distribution of Linux which is an open source operating system. There are scientific tools that run on Linux for almost all scientific disciplines. However, despite its significant advantages, scientific computing on Ubuntu is not widespread; a number of challenges make scientific computing on Ubuntu problematic:

- Few people are aware of Ubuntu as a platform for scientific computing
- Finding and choosing the best scientific tools on Ubuntu are difficult
- Open source software (OSS) scientific tools are not integrated in cohesive suites
- Some OSS scientific tools provide only limited functionality
- OSS tools are not available for all scientific tasks
- Ubuntu is not certified to run all proprietary scientific software
- Software maintenance requires technical proficiency.

To address these problems and to promote scientific computing on Linux, the Meraka Institute of the CSIR is developing scubuntu, an Ubuntu-based Linux distribution for scientists. The project vision is that scubuntu will become the premier choice of desktop operating system for researchers and scientists.

The approach taken by scubuntu addresses some of the problems listed above by:

- Making all work publicly available
- Engaging in formal and informal collaboration
- The creation of user and developer communities
- Integrating scientific OSS into cohesive suites of packages
- Centralising the technical management and maintenance of packages.

Investigative work was undertaken to determine the suitability of a **CD-based installation** program for scubuntu. An experimental prototype of such an installer was developed, based on the standard Ubuntu installer. This prototype installer allows users to **select and install scientific software** based on their field of study or profile.

PROFILES

One of the key activities in scubuntu is the determination of **profiles of scientists**, based on their computing and software requirements. These profiles are coupled with a specification identifying the OSS scientific tools that are most appropriate for that profile. The refinement of these profiles and package collections is an ongoing activity that will be driven by the scubuntu user community.

Figure 1 describes the situation where scubuntu has a top-level profile called *biology* with sub-profiles *biochemistry* and *molecular biology*. Each of these sub-profiles then contain lists of software packages appropriate for biochemists and molecular biologists, respectively

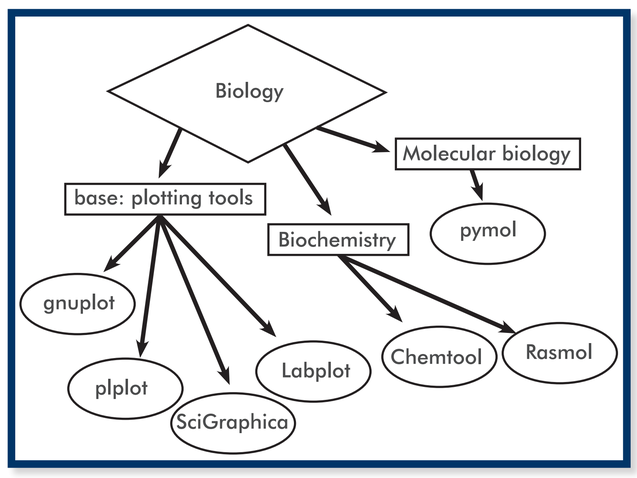

Figure 1: An example of a profile with meta-packages

INSTALLER PROTOTYPE

There are a number of possible approaches to installing customised sets of packages during the Ubuntu installation process. These include:

- The automatic installation of a predefined set of packages (without user input) from the installation CD
- The use of additional screens in the installation program to prompt the user to select the desired packages from a list
- The use of 'add-on' CDs to install selected packages after the installation of the base system
- The use of network-based installers to fetch packages from the internet.

The scubuntu team investigated the use of a CD/DVD-based installer for scubuntu, with all the scientific applications already on the CD. This approach could work well for the African market where the lack of good connectivity is a constraint to the use of network-based installers.

Scubuntu is based on the Ubuntu 'live CD' and a modified version of Ubiquity, the standard Ubuntu installation program. Ubiquity is a simple graphical installation program that runs from a live CD and is designed to integrate well with Debian and Ubuntu-based systems. It is written largely in Python and uses the Debian Installer (d-i) as a back-end for many of its functions.

The scubuntu installer provides an additional screen where users can select scientific packages from a tree-based package list. This tree of available packages is created dynamically and grouped by scientific profile. **Figure 2** shows what this selection screen looks like.

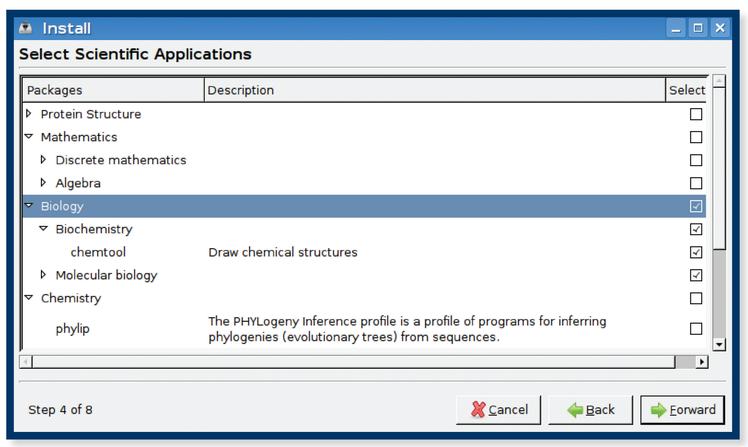

Figure 2: Scubuntu's scientific package select screen

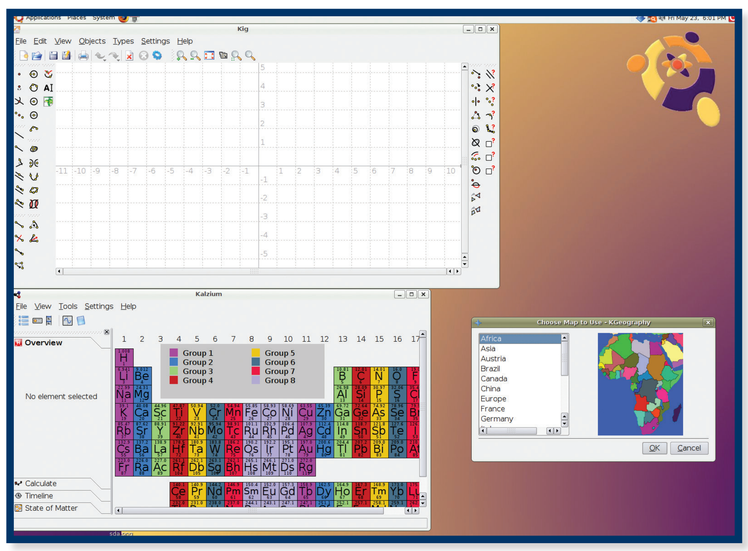

Figure 3: The scubuntu desktop

CONCLUSIONS AND FUTURE WORK

The work on the installation program presented in this paper addresses some of the difficulties associated with scientific computing on Linux by:

- Pre-selecting the best-of-breed scientific open source software and grouping these packages into profiles
- Predefining profiles for supported scientific domains
- Providing a user-friendly way for scientists to select these profiles and install the associated package sets.

The scubuntu prototype presented here will be tested by scientists in a number of fields. Future work will include the development of new profiles to cover additional scientific fields, the refinement of existing profiles, and the investigation of additional options for the installer program. In particular, it is thought that the current design may not be optimal, as it is severely restricted by the amount of data that can be placed on a CD/DVD. There may be significant benefit to taking a hybrid approach, combining elements of the current design with the use of the 'add-on' CD approach. This could result in scubuntu comprising a number of CD/DVDs:

1. A modified Ubuntu live CD that includes (in addition to normal packages) a set of packages commonly useful to all scientists, regardless of their field of study.
2. Additional 'add-on' CDs for each supported scientific field.

As an Ubuntu-based Linux distribution, Scubuntu will provide scientists and researchers in Africa with a quick and easy installer for selecting and installing scientific applications based on specific fields.