Opening the treasury of R language to a wider scientific community: *GCDkit*, a package for interpretation of geochemical data

V. JANOUŠEK[1], C. M. FARROW[2], V. ERBAN[3]

[1] *Institut für Mineralogie, Universität Salzburg, Hellbrunnerstraße 34, A-5020 Salzburg, Austria, vojtech.janousek@sbg.ac.at*

[2] *Computing Service, University of Glasgow, Glasgow G12 8QQ, Scotland, c.farrow@compserv.gla.ac.uk*

[3] *Czech Geological Survey, Klárov 3, 118 21 Prague 1, Czech Republic, erban@cgu.cz*

Similarly to many branches of science, the flood of numerical data on the one hand, and the dearth of potent and flexible software for their recalculation and plotting on the other, is often a limiting factor to a creative work in igneous geochemistry. The Geochemical Data Toolkit for Windows *(GCDkit)* is our answer to the challenge, taking advantage of the tools available in R and exploiting its functions facilitating Windows-like interaction. The current version does:

1. Offer an alternative, more user-friendly interface to powerful statistical and graphical functions built in R. The R is provided with a command line interface, which on the one hand allows excellent control for experienced users but on the other tends to discourage many scientists, accustomed to software with graphical user interface. All functions of *GCDkit* are accessible via pull-down menus, as well as in an interactive regime.
2. Provide core routines for effortless data management, i.e. loading and saving of free form text files, copying data from/to clipboard, data editing, searching and generation of subsets using regular expressions and Boolean logic.
3. Allow the analyses to be organised into unique groups, which are subsequently utilised by the statistical and plotting functions. This can be done on the basis of various attributes (locality, rock type,…), ranges of a numerical variable, by cluster analysis or using a selected classification diagram.
4. Contain flexible high-level graphical functions. Most of the diagrams are defined as templates for *Figaro*, a set of graphical utilities for R. *Figaro* provides a means to create figure objects, which contain both the data and methods to make subsequent changes to the plot (zooming and scaling, adding comments or legend, identifying data points, altering the size or colour of the plotting symbols…). The templates can be used also as a basis for classification; the general algorithm looks for the name of the polygon within the diagram, into which the analysis falls according to its x–y coordinates.
5. Offer brand new calculation, modelling and plotting tools designed specifically for geochemists.
6. Permit expansions by the end users. Text files containing R code defining new calculation or plotting options when placed in the sub-directory *Plugin* are automatically loaded at the system start-up. These can be made accessible via newly appended menu items.
7. Avoid any licensing problems. The *GCDkit* is distributed as freeware via the WWW; the current version can be downloaded from *http://www.gla.ac.uk/gcdkit.*

The whole *GCDkit* system, which is modular and straightforward to modify, provides potentially a platform for DIY additions written by R literate geochemists for their less fortunate colleagues. Our mission is to amass eventually a platform-independent version of *GCDkit*, using Tcl/Tk-based interface on Unix/Linux and Macintosh (System X). Moreover we intend to extract a core of the system (without geochemical bias), creating a simple GUI front-end to R, intended specifically for use in natural sciences.

This project was supported by the grant FWF 15133–GEO.