

Guidelines for code development and code documentation.

Version 3 (2 February 2009)

Nanoquanta Integration team 9

Integration of theory and code developments.

C.-O. Almbladh, A. Cucca, X. Gonze, M. Marques, and Y. Pouillon

Context of this document

The software development effort in the NANOQUANTA network and the European Theoretical Spectroscopy Facility is a long-term effort. The pieces of code that are and will be developed in the network and the facility, will likely be examined and/or reused by other developers than the one who originally wrote it. This might happen because cases not originally taken into account have to be developed, because the codes are ported to new platforms, because the initial testing was not careful enough and left some bugs, etc... It is well known that the time spend to modify a complex software (often referred to as the “maintenance” of a software) can be larger than the initial time needed to develop it.

By proposing guidelines for code development inside the NANOQUANTA network, we aim at facilitating the future maintenance of a piece of software, with a minimal increase of the initial development time, mostly through the application of “good practice” rules. We have chosen a set of guidelines that are quite universal (e.g. they are independent of the programming language), widely accepted, and independent of the platform. We also propose some advices to ease the initial development period.

Guidelines

1. Documentation

The time needed to understand a piece of code can vary dramatically with the documentation available for this piece of code. Moreover, it is a large fraction of the time needed to modify a code. Thus, the initial developer should try to ease this process of understanding. He should :

- Choose carefully the names of his variables, subroutines, etc ... so that they express as adequately as possible the underlying quantities, concepts or operations. The question of choosing good, self-explanatory names is important primarily for global entities (functions/subroutines, modules, datatypes ...)
- Write comments (in good English), related to each logical section of his code, typically, one line of comment to explain 10-20 lines of codes (more is even better)
- Describe the purpose, options and arguments of the routines and functions, and refer, when possible, to existing published literature, e.g. cite a paper and the (exact) number of the equation that is coded. Supply similar comments for members of modules, structures ...

- Keep notes of the formulas that are implemented, if they are not (yet) published), and ensure others might get it with the code
- Not try to produce “clever” coding : unless it is very well documented, “clever” coding might be much harder to understand than “dull” coding. If it is not critical for speed, the best code is simple, and easy-to-read.

2. Standards

For optimal portability of code (thus, sparing the time of somebody else), the developer should stick to standards. He/she should NOT use particular features of a compiler only available on some specific platform. As examples, new Fortran code should adhere to the Fortran 90 or Fortran 95 standard unless there are strong reasons for doing otherwise. C code should be ANSI C with every function properly prototyped. New C++ should adhere to the C++ standard. If there is strong need for some system/platform dependent part, it should be isolated into separate, replaceable units.

3. Code reuse and libraries

The best way to develop rapidly a code that works is to reuse a piece of code that has already been written and tested! However, there is often a psychological barrier to reuse existing code: one likes to have the complete understanding of the code, and this cannot be obtained in this case. However, this is often a large waste of time : the own estimation of the time needed to develop a piece of code is very often restricted to the writing part, while there exist a large additional time, usually overlooked, needed to test and debug it. Developing always takes more time than expected, especially if high-quality coding is aimed at.

Libraries are (large) collections of reusable, carefully tested code. We strongly advice the developer to use the following libraries :

- BLAS and LAPACK for basic linear algebra operations (including vector-vector, matrix-vector, matrix-matrix operations, solution of linear systems of equations, eigenvalue problems ...)
- NETCDF (or HDF) for IO operations (allowing portable files, more flexible formatting, etc)

Try to use libraries which are open-source as much as possible, in order to avoid copyright problems for ETSF.

Note that Numerical Recipes routines are copyrighted.

4. Modularity

Aim for a modular structure with as few cross-dependencies as possible. Let the different parts of your code know little as possible about the rest of the program.

Try to dissociate tasks that fulfill different goals, e.g. write separate subroutines for numerical treatments of data, and for their printing. Try that each routine or function perform a logical task of a well-defined character : data handling, or numerical treatment, or driving different routines of equivalent “level”. Try to think in terms of re-usability : will this subroutine easily reused by somebody else ?

5. Use a versioning tool

This is not a “good practice” rule, but proves very useful to developers. There are excellent free software to manage the different versions that always appear in the development process : e.g. Bzr or SVN. Use such tools ! Their use involves a very small overhead, but they can save your day. This is nearly mandatory for group development, but is also quite interesting in case of an isolated person.

The point is that such a software keep track of all the changes that your software is undergoing. This helps considerably the debugging process. As an example, when you develop a new feature, it often happens that there are “ripple” effects, that cause new problems with some functionality of our code, not immediately linked with the functionality your are implementing. Such problems might go unnoticed during some time. When the faulty behavior is seen, the debugging might be very difficult. With the ability to come back step-by-step to the point where the code was working, it is usually easier to isolate the bug.

Additional references :

“The mythical man-month” F.P. Brooks, Addison Wesley, London (1995) ISBN 0201835959

“Software maintenance. Concept and practice” P. Grubb and A.A. Takang, World Scientific, London, 2nd edition (2003) ISBN 981-238-426-X