Introducing RSF, a computational platform for geophysical data processing and reproducible numerical experiments

Sergey Fomel*, University of Texas at Austin, Paul Sava, Colorado School of Mines, and Felix Herrmann, University of British Columbia

This talk is the first public announcement of RSF (from *regularly sampled format*), an open-source software package developed in collaboration by a group of geophysicists, petroleum engineers, and computational scientists. It is also an open invitation for other scientists to join the project as users or developers.

The four main features of RSF are:

1. RSF is a *new* package. It started in 2003 and was developed entirely from scratch. Being a new package, it follows modern software engineering practices such as module encapsulation and test-driven development. A rapid development of a project of this scope (more than 300 main programs and more than 3000 tests) would not be possible without standing on the shoulders of giants and learning from the 30 years of previous experience in open packages such as SEPlib and Seismic Unix. We have borrowed and reimplemented functionality and ideas from these packages.

2. RSF is a *test-driven* package. Test-driven development is not only an agile software programming practice but also a way of bringing scientific foundation to geophysical research that involves numerical experiments. Bringing reproducibility and peer review, the backbone of any real science, to the field of computational geophysics is the main motivation for RSF development. The package consists of two levels: low-level main programs (typically developed in the C programming language and working as data filters) and high-level processing flows (described with the help of the Python programming language) that combine main programs and completely document data processing histories for testing and reproducibility. Experience shows that high-level programming is easily mastered even by beginning students that have no previous programming experience.

3. RSF is an *open-source* package. It is distributed under the standard GPL open-source license, which places no restriction on the usage and modification of the code. Access to modifying the source repository is not controlled by one organization but shared equally among different developers. This enables an open collaboration among different groups spread all over the world, in the true spirit of the open source movement.

4. RSF uses a *simple, flexible, and universal* data format that can handle very large datasets but is not tied specifically to seismic data or data of any other particular kind. This "regularly sampled" format is borrowed from the traditional SEPlib and is also related to the DDS format developed by Amoco and BP. A universal data format allows us to share general-purpose data processing tools with scientists from other disciplines such as petroleum engineers working on large-scale reservoir simulations.