

# Collaborative Infrastructure for Scientific Software Development

## Project directive

---

This project is a part of NeIC Generic Area Activity for 2015, “Activity GEN-7 Nordic e-Infrastructure for scientific software” and it is in line with Action 10 “Nordic eInfrastructure for scientific software” from the Nordic eScience Action Plan.

# 1. Background

One of the main goals of NeIC is the provision of support to the e-Science research communities. For that it is required to facilitate collaborative work among different groups, provide of state-of-the-art e-Infrastructure services along with the necessary know-how for their utilization.

Software development is at the core of e-Science research. Advances in hardware technology and computational algorithms put substantial pressure on software developers. Thus fast, efficient and reliable development cycles directly affect the speed and quality of e-Science work.

Yet our experience shows that a large number of academic research groups develop in-house software using very inefficient and outdated methodologies. The reason often being that many packages have started as one-person pet projects aimed at solving a specific problem at hand. But over the years the projects would grow with many new developers joining the team, and the codebase expanding considerably. The new functionality would make the software popular, widely used and important for whole research community. However, maintenance of such packages e.g. >10 year old ones, eventually creates substantial friction, which can be overcome with the adoption of proper software development methodologies.

Main reasons for the lack of adoption are unfamiliarity with existing solutions, and fear of potential steep learning curve, which will affect productivity. In fact, the opposite is true - sticking to old practices hurts very much the productivity in the long-term.

One initial step in direction of promotion of such best practices was the first academic course in software development tools at KTH in December 2014 (<http://sese.nu/scientific-software-development-toolbox>). It was organized by Radovan Bast (DALTON developer) and Rossen Apostolov (author of the current proposal and GROMACS developer) from Sweden, and Jonas Juselius (SYMBIOSES developer) from Norway. The teachers are developers of major software packages who have adopted and experienced the benefits of modern software development practices. The course was a very big success (the only negative feedback we had was that the students wished the course was longer!). It clearly showed the lack of expertise and pressing need for proper e-Infrastructure and tools.

SNIC has established a nation-wide network of application experts who acknowledge the need for provision of such infrastructure. During a provider forum meeting all national providers, i.e. SIGMA2, SNIC, CSC, DEIC, and RHNET, have also agreed on its benefits. Finally, the establishment of this infrastructure is in line with Action 10 "Nordic e-Infrastructure for scientific software" from the Nordic e-Science Action Plan.

Several software packages developed by Nordic research groups have large userbases worldwide and a big impact in their corresponding communities. Examples of such are ARC(<http://www.nordugrid.org/arc>), Gromacs ([www.gromacs.org](http://www.gromacs.org)), Symbioses (<http://symbioses.no/doku.php>), NordicESM, Dalton ([daltonprogram.org](http://daltonprogram.org)), Chippster(<http://chipster.csc.fi>). Some of them (Gromacs, Dalton and Symbioses) are already using advanced workflows for development.

## 2. Project idea

This project proposes the establishment of software development e-Infrastructure, which is coupled with necessary technical expertise to address the growing needs of the computational communities.

The project includes three aspects.

First: Deployment of infrastructure. It will provide the necessary systems and tools, which could include:

- Distributed version control (DVC) repositories for scientific codes  
=> allow for extremely easy collaborative code development
- Issue tracking systems  
=> a necessity for any software project
- Code review systems  
=> major improvement in the quality of the code, collaborative work, and skills of the programmers
- Code analyzers, debuggers and profilers framework  
=> indispensable for catching errors or inefficiencies in the code
- Build systems, unit testing frameworks and report boards  
=> automated testing before sharing of code
- Code Integration and benchmarking  
=> automated software validation and performance monitoring

Depending on the available resources, the software e-Infrastructure can provide some or all of the above services. It could be extended with additional features, as needed.

Second: Extensive training activities. In order to quickly bring researchers up-to-speed, we will organize series of training events. The events will be organized as close to users as possible, e.g. at research institutions with large communities where we can achieve biggest impact.

Third: Nordic-wide network of application experts. Building on the experience of e.g. the SNIC application experts' initiative, the project can create the nucleus of a wider communication environment for exchange of expertise. This will be achieved naturally through jointly developing and the infrastructure and running the training activities.

### 3. Expected benefit

Adoption of modern tools and best practices for software development will bring considerably benefits to researchers by increasing:

1. productivity of the researchers,
2. quality of their codes,
3. skills of the developers.

National providers will benefit from 1) the increased competence among their users, which will lead to 2) improved utilization of their resources through development of better software, as well as 3) the long-term maintainability of locally developed software.

The Nordic region as a whole will benefit from the nucleation of a community and increased interactions between application experts from different countries, as well as with the corresponding user groups.

### 4. Basis

- Establishing a software development e-Infrastructure
- Coupling it with necessary technical expertise
- Extensive training and on-boarding activities

### 5. Contact persons

This is the list of persons that have knowledge on subject and have helped in the preparation phase.

Name	Email	Role
Rossen Apostolov	rossen@kth.se	<b>Author of the proposal</b> and SNIC Application expert in Molecular Simulations at PDC. He is a developer of one of the major Nordic codes, Gromacs, which has a sophisticated development infrastructure.

Dejan Vitlacil	vitlacil@kth.se	NeIC Generic Area Coordinator and <b>main contact</b> with NeIC
Jonas Juselius	jonas.juselius@uit.no	Jonas (University of Tromso) has expert knowledge in the domain of the proposal and is leading activities in line with the suggested activities. We have discussed the project and he is very interested in participating.
Radovan Bast	radovan.bast@gmail.com	Radovan (PDC) was the main organizer and leader of the “Scientific Software Toolbox”. He is a SNIC Application Expert in Electronic Structure, core developer of Dalton and Dirac, and has expert knowledge in the domain.
Andreas Jaunsen	andreas.jaunsen@uninet.t.no	Technical Provider Forum member (NO)
Jura Tarus	jura.tarus@csc.fi	Technical Provider Forum member (FI)
Ann-Charlotte Sonnhammer	ann-charlotte.sonnhammer@it.uu.se	Technical Provider Forum member (SE)
Torben Rasmussen	torbenr@nsc.liu.se	SNIC Application Experts Coordinator
Jon Kerr Nilsen	j.k.nilsen@usit.uio.no	Senior Engineer and ARC lead developer