

**PDC Center for High Performance Computing**

# PDC Newsletter

**BioExcel Assists Release of Workflow Manager for Large Genetic Studies**

- Rossen Apostolov, PDC, page 3

**Large-Scale Computation Made Possible for Sensistive Human Data**

- Ann-Charlotte Sonnhammer, SNIC, Peter Ankerstål, UPPMAX, and Sverker Lundin, SciLifeLab, page 4

**PDC Pub 2017** - Cristian Cira, PDC, page 6

**iRODS Technology for Research Data Storage** - Dejan Vitlacil, PDC, page 7

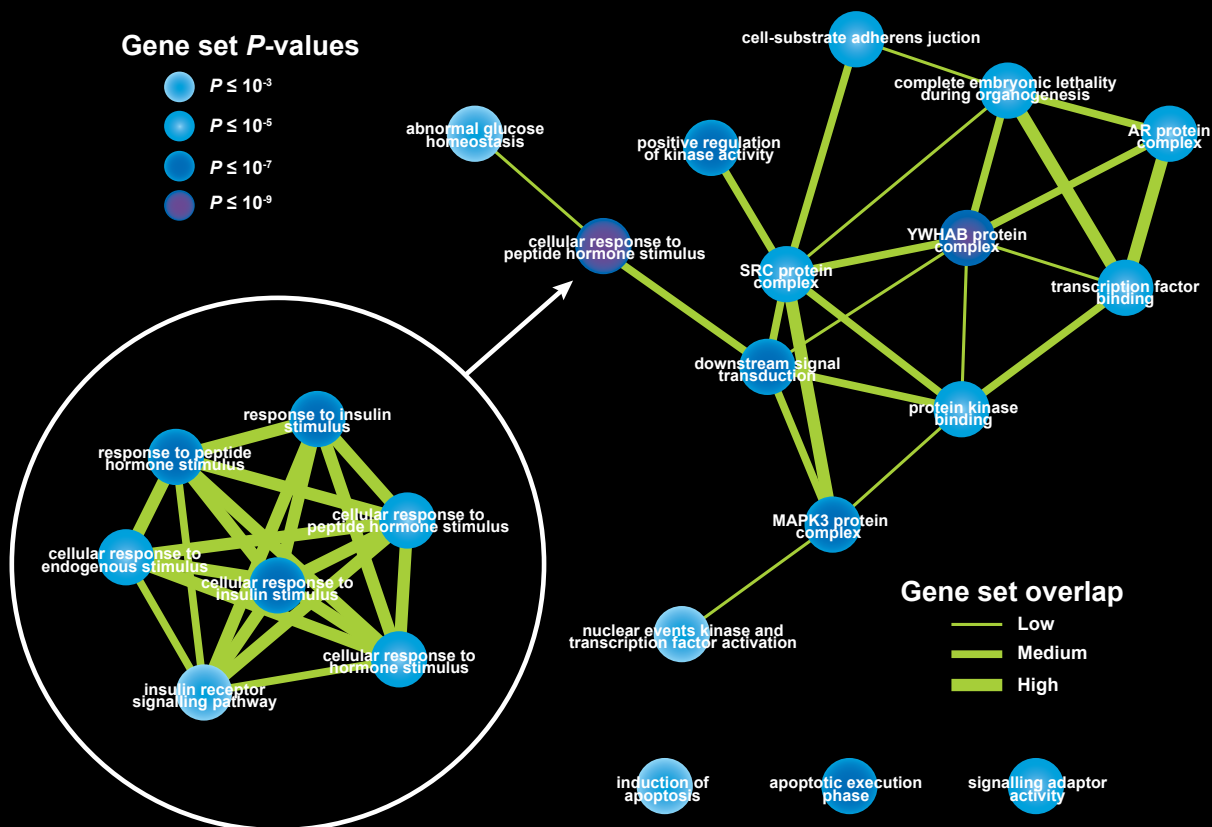
**HPCE3 Grant Approved by European Commission** - Lilit Axner, PDC, page 8

**Introduction to PDC Systems Course** - Cristian Cira, PDC, page 9

**CodeRefinery Workshops** - Thor Wikfeldt, PDC, page 10

**PRACE/BioExcel Spring School 2017** - Michaela Barth, PDC, page 11

**Second SNIC All Hands Meeting** - Gert Svensson, PDC, page 15





**Erwin Laure**  
Director PDC, and CST

### The PDC Newsletter is published by PDC at CSC, KTH.

PDC operates leading-edge, high performance computers as easily-accessible national resources. These resources are primarily available for Swedish academic research and education. PDC, which is hosted by CSC, KTH, is one of the six centres in the Swedish National Infrastructure for Computing (SNIC).

**Editor:** Erwin Laure

**Editorial Staff:** Genet Edmondson

**Layout:** Genet Edmondson

**Graphic Design:** Maria Malmqvist

**E-mail:** [support@pdc.kth.se](mailto:support@pdc.kth.se)

ISSN 1401-9671

### Cover

COMPSs is used to manage workflows for studies of genetic data such as that stored in the [SNIC SENS project](#). The results on the cover are from a genetic study about diabetes using data from seventy thousand individuals.

### How to Contact PDC

**Visiting address:**

Teknikringen 14, "Plan 4",  
KTH, Stockholm

**Mail:** PDC, KTH,  
SE-10044 Stockholm, Sweden

**E-mail:** [support@pdc.kth.se](mailto:support@pdc.kth.se)

**www:** <http://www.pdc.kth.se>

**Phone:** +46 8 790 7800

**Fax:** +46 8 247 784

## Editorial

The Life Sciences are becoming an increasingly important part of the Swedish high performance computing (HPC) ecosystem. Not only is their demand for computing simulations and data analysis ever increasing, but they also have specific requirements when it comes to storing their data – particularly when it comes to sensitive data, like data including personal information. In collaboration with the Science for Life Laboratory (SciLifeLab) and the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX), PDC is helping to provide the technological and legal framework for storing such data. The infrastructure will mainly be used to store sequencing data from the National Genomics Infrastructure (NGI). The underlying technology, iRODS, is also being used for other storage projects, such as a national storage infrastructure for Swedish National Infrastructure for Computing (SNIC) users that is planned to be made available during the summer. Get in touch with our support personnel ([support@pdc.kth.se](mailto:support@pdc.kth.se)) to discuss your storage needs!

The Life Sciences, however, encompass much more than storing data – in fact, they are one of the research communities that drive the current exascale developments. PDC is supporting these efforts through an EC H2020 funded Centre of Excellence, BioExcel, which is working to improve the efficiency, scalability, and usability of key Life Science codes. In this edition of the newsletter we report on the successes of our colleagues from the Barcelona Supercomputing Center (BSC) in parallelizing large-scale genetic studies with workflow systems. And this year BioExcel, together with PRACE, organized the first “HPC for Life Sciences” spring school, which was hosted by PDC.

If you develop code yourself, you know how difficult it is to write correct code, and writing good code is even harder! With the support of the Nordic e-Infrastructure Collaboration (NeIC), we have started a Nordic-wide effort to train people in software engineering best practices: the CodeRefinery project.

In May we also held our legendary Pub event, where PDC users and staff can get together in an informal setting. The short scientific highlights presented by some PDC users were particularly interesting for PDC staff – it is always good to know what problems researchers actually solve using our resources!

Finally, the process of determining the future funding of SNIC has reached an important milestone. At the end of April, Uppsala University (as the SNIC host university) sent a document about the future organization and funding structure of SNIC to the Swedish Research Council (VR). This document, which foresees extending the financial

contributors (currently VR and the SNIC centre host universities) to include all the major Swedish universities, was supported by all the affected universities and we expect VR to make a decision about the future funding of SNIC in the autumn. With this milestone reached, it should be possible to have a stable financial situation for SNIC in the coming years. The future of SNIC was also a major topic at the Second SNIC All Hands meeting, held in Gothenburg at the end of April.

With all this positive news, I wish you a relaxing summer time and I'm looking forward to all the great research that will be done on the PDC systems in the autumn.

*Erwin Laure, Director PDC, and CST*

## BioExcel Assists Release of Workflow Manager for Large Genetic Studies

Rossen Apostolov, PDC

The BioExcel Center of Excellence has assisted in the recent release of a new version of the **COMPSs** programming model which powers **GUIDANCE**, a pipeline for large-scale genetic studies, by parallelizing it at the task level and enabling it to run on distributed computing platforms.

To understand the significance of this, one needs to be aware that the computational requirements associated with large genetic studies keep growing alarmingly, both in capacity and in complexity. For example, the full analysis of the genotypes of thousands of individuals involves thousands of different types of tasks, each of which has specific computational requirements. In order to address these needs, the BioExcel partner team at the Barcelona Supercomputing Center (BSC) developed **GUIDANCE**, a modular compilation of programs for performing complete genetic association analyses. **GUIDANCE** makes it possible for researchers to perform all the steps involved in a large-scale genome- and phenome-wide association analysis in a single execution – it also enables users to perform the steps in a modular way with optional user intervention.

The **GUIDANCE** implementation is based on **COMPSs**, which is a task-based programming framework that facilitates the development and execution of parallel applications and workflows in distributed infrastructures, such as high performance computing (HPC) clusters, grids and clouds, making this application integrable into multiple parallel platforms. **COMPSs** is able to parallelize (at task level) sequential applications written in Python, Java and C/C++. At execution time, **COMPSs** schedules, balances and organizes all the necessary subtasks to ensure efficient usage of the computing resources. It also takes care of the data transfers between tasks, when those are distributed between remote nodes.

## In This Issue

### Editorial

*Erwin Laure* ..... 2

### Staff Focus

*Jaime Axel Rosal*..... 4

*Ilari Korhonen*..... 5

*Niclas Jansson* ..... 9

### BioExcel Assists Release of Workflow Manager for Large Genetic Studies

*Rossen Apostolov*..... 3

### Large-Scale Computation Made Possible for Sensitive Human Data

*Ann-Charlotte Sonnhammer, Peter Ankerstål and Sverker Lundin* ..... 4

### PDC Pub 2017

*Cristian Cira*..... 6

### iRODS Technology for Research Data Storage

*Dejan Vitlacil*..... 7

### HPCE3 Grant Approved by European Commission

*Lilit Axner*..... 8

### Introduction to PDC Systems Course

*Cristian Cira*..... 9

### CodeRefinery Workshops

*Thor Wikfeldt* ..... 10

### PRACE/BioExcel Spring School 2017:

#### "HPC for Life Sciences"

*Michaela Barth*..... 11

### Second SNIC All Hands Meeting

*Gert Svensson*..... 15

**PDC-Related Events** ..... 16

**HPC Sources**..... 16

## Staff Focus



Jaime Axel Rosal

Jaime's main research area is in high-performance and low-complexity algorithms for quantum chemistry. He did his Ph.D. in the Theoretical Chemistry and Biology department at KTH. Before that Jaime worked as a mathematics and science teacher, writing movie critiques and doing technical support. After his Ph.D., Jaime did a one-year postdoc in Linköping before moving back to Stockholm.

Jaime is currently working part-time as an Application Expert at PDC, providing support for researchers in Sweden and the rest of the EU, and working as a developer with the QM/CMM project in Dalton, parallelizing Dalton and merging it with the CPP (complex polarization propagator) code. Jaime is also the creator and main developer of the FIESTA quantum chemistry package and works on the upcoming GATOR program. Jaime helps PDC users with (hybrid) parallelization, code optimization for high performance architectures and quantum chemistry software.

Jaime's hobbies include playing go (weiqi) at the 1 dan amateur level, programming video games, playing the piano and composing music.

As an example, GUIDANCE was used for a recent genetic study identifying the locations of genes associated with type 2 diabetes. The study was based on the reanalysis of seventy thousand publicly available genetic samples, and led to the identification of seven new gene locations associated with the illness, which included variants of low and rare frequency in the population that could have only been found using GUIDANCE methodology (see [cover image](#)).

Both COMPSs and GUIDANCE are open source and can be used free by biomolecular research communities. If you would like more information about using COMPSs and GUIDANCE, please [contact us](#) or visit our workflow support forums at [Ask.BioExcel.eu](http://Ask.BioExcel.eu).

---

## Large-Scale Computation Made Possible for Sensitive Human Data

**Ann-Charlotte Sonnhhammer, SNIC, Peter Ankerstål, UPPMAX, and Sverker Lundin, SciLifeLab**

There have been dramatic technological advancements within genomics in recent years. These advancements are transforming the way researchers can investigate the genetic basis of disease. This development is also impacting healthcare by bringing forth more precise methods for diagnosis and treatment.

Swedish researchers are in a unique position to contribute to the understanding of disease due to the combination of a tradition of technology development, the existence of Swedish bio-banks with large collections of tissue and cell samples, along with well-organized and comprehensive Swedish population and disease registries, and longitudinal population-based epidemiologic studies in different geographical regions of Sweden.

Now that human whole genome sequencing is possible on a large scale, the amount of data is also dramatically increasing, in turn demanding large resources for data-intensive computations. In addition to this, human genetic data is, by its very nature, personal and sensitive and the Swedish National Infrastructure for Computing (SNIC) lacked a suitable computer system for dealing with this type of data. To address this, SNIC was granted funding to set up and maintain computing and storage resources to handle sensitive personal data generated by large-scale molecular experiments. The funding comes from the Swedish Research Council, and the Knut and Alice Wallenberg Foundation, with co-funding from Uppsala University and the Science for Life Laboratory (SciLifeLab). A new SNIC project, called SNIC SENS, was initiated to set up and maintain the resources, with the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) and PDC as the participating SNIC centres.

*Below: Close-up of one of the systems at UPPMAX (Photo: Peter Ankerstål)  
Note that, for security reasons, the SNIC SENS resources are all in physically locked racks (which are not shown in this photo).*



The aim of the SNIC SENS project is to set up and maintain a production resource for the Swedish National Genomics Infrastructure (NGI), and a national SNIC resource for handling sensitive personal data that originates from large-scale molecular experiments, such as next-generation sequencing. Since the data is sensitive personal data, the information and IT security work, plus the legal considerations, have been essential parts of the project. The Security and Safety Division at Uppsala University has been of great help, as well as the Legal Affairs Division of Uppsala University and the Legal Department at the KTH Royal Institute of Technology. The information security work started before the procurement process, and provided important information regarding the system requirements, and has continued since then.

The production system for NGI is called Irma and consists of 250 compute nodes with 2 x 8 cores each and 256 GB RAM per node. The storage system associated with Irma is called Lupus and is a 1 PB Lustre file system with a peak write performance of 25 GB/s. Irma is a “regular” high performance computing (HPC) cluster with fast storage and infiniband interconnect for MPI and file traffic. Irma was put into production in March 2016.

The new national computing resource within SNIC is called Bianca. It consists of 200 compute nodes with 2 x 8 cores, with 128 GB RAM per node and 4 TB of node local storage. The storage system associated with Bianca is called Castor, and it is a 6 PB Gluster file system with a peak write performance of 25 GB/s. Bianca was put into production on the 7th of April 2017, and inaugurated on the 24th of April 2017. The first users in the pilot phase started in mid-December 2016.

For both parts of the project, the secure backup resource at PDC is used, and plays an important part in ensuring that the information security guidelines are fulfilled. The project has given SNIC and the participating SNIC centres new knowledge, which can be utilized in new projects.

One of the first groups of users on Bianca was the SweGen project. It is one of the projects in the [Swedish Genomes Program](#) of the

## Staff Focus



Ilari Korhonen

Ilari Korhonen started his academic studies in computer engineering at Tampere University of Technology, Finland. His studies continued in physics at the University of Jyväskylä, Finland, and he eventually graduated with an M.Sc. degree in Mathematics from the University of Jyväskylä. Ilari's studies have always revolved around the field of scientific computing and while he was studying he also worked as a research assistant at the University of Jyväskylä, as well as in private companies, doing development and Unix and Linux systems administration. Before joining KTH, Ilari was working at his alma mater building a research data management and storage system.

Ilari is now working at PDC mostly with the SNIC iRODS project, the goal of which is to build a nationally accessible storage system for research data in Sweden.

When not working, Ilari likes to spend time with motorcycles to get away from computers and is also trying to dedicate some time to playing the guitar and photography.

National Projects initiative at SciLifeLab funded by the Knut and Alice Wallenberg Foundation. The aim of the SweGen project is to construct a reference dataset for the genetics research community and clinical genetics laboratories. A high-quality genetic variant database for the Swedish population is being established from the genomes of one thousand individuals selected to reflect the genetic structure and geographical distribution of the Swedish population. The variant frequencies have been made available at <https://swefreq.nbis.se>.

To have researchers from several universities and research institutes analysing sequencing data using a national resource has required strengthening the IT security, and the development of appropriate routines and procedures. Due to the security concerns that arise when working with sensitive personal data, the UPPMAX team has worked hard to maintain full separation between projects. Moreover, data transfers are logged in such a way that it should be impossible to extract any data from the system without having a complete audit trail. In order to achieve this, Bianca provides a fully virtualized and compartmentalized environment based on OpenStack, where every project gets its own virtualized version of an UPPMAX cluster, so there is no such thing as sharing a login or compute node.

When working with the design and implementation of the resources, the UPPMAX team has made several internal and external security assessments of the system. The efforts to maintain and improve security will continue throughout the entire lifetime of the resources.



Above: PDC Pub and Open House, 24 May 2017

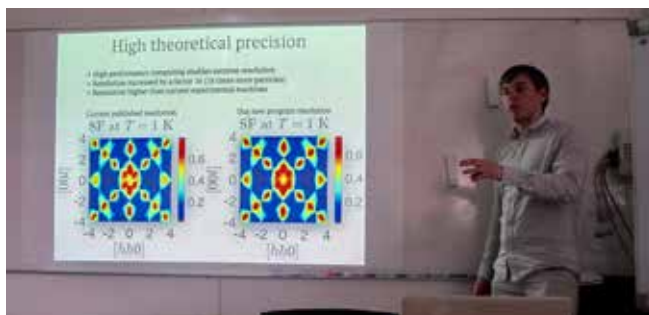
## PDC Pub 2017

Cristian Cira, PDC

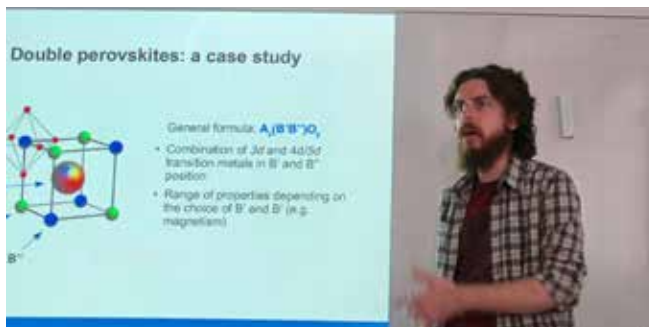
About fifty people attended this year's PDC Pub and Open House on the 24th of May. Last year the mini-lectures about research being done on PDC's systems proved so popular that we needed to move to a larger room this year!



Above: Arne Elofsson talking about computational challenges when predicting protein structure on a genomic scale



Above: Mikael Twengström discusses spin ice and whether magnetic monopoles exist

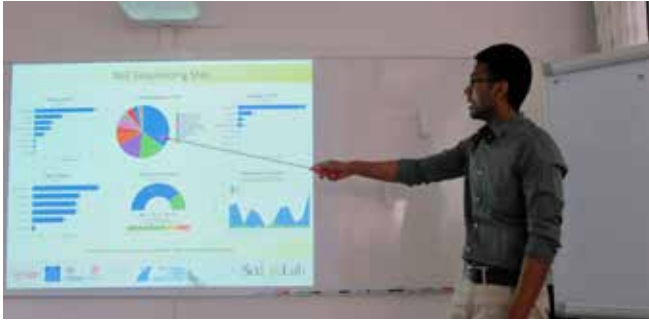


Above: Michele Visciarelli on high throughput formation energy calculations for screening novel double perovskites



Above: Patrik Norman elaborates on probing molecular interactions with DNA by means of circularly polarized light

*Below: Senthilkumar Panneerselvam talking about long-term storage at PDC for NGS data*



*Below: Ricardo Vinuesa on high fidelity simulations of the flow around wings*



## iRODS Technology for Research Data Storage

Dejan Vitlacil, PDC

In collaboration with other partners from the Swedish National Infrastructure for Computing (SNIC), PDC is developing a new service for research data storage. The new service will make it easier to manage research data and will be a significant step towards providing open access data. Researchers who make use of this storage service for live project data will be able to start generating metadata right from the very beginning of their project. This will make it a lot easier to package data and archive it when their projects come to an end. The expectation is that archiving services will also benefit from these efforts immediately, as metadata is an important part of making data searchable and is also useful when it comes to publishing data.

The Integrated Rule-Oriented Data System (iRODS) is open-source software that provides a comprehensive set of tools to support data management tasks from the initial collection of data through to archiving and reusing the data. This

is particularly important given the implications of the worldwide movement towards Open Science and Open Data Access. iRODS is supported and maintained by the iRODS Consortium, which is based at the Renaissance Computing Institute (RENCI), a research institute of the University of North Carolina at Chapel Hill (UNC). The membership-based consortium receives funding from UNC and the other twelve consortium members, which include hardware vendors such as DDN, HGST and IBM, as well as universities such as Utrecht University and University College London. The consortium reports that iRODS is used by research organizations and government agencies worldwide.

Over the last few years the iRODS software has been subjected to very significant refactoring and reorganization – the result of which is that iRODS is now being released as a production-level software distribution with commercial support, as well as a strong user community.

Our goal at PDC during this year is to implement both a new iRODS-based storage service for SNIC, as well as a separate service that will be available at PDC for Swedish researchers whose research data storage requirements are not addressed by the available SNIC services.

The new iRODS-based SNIC service is being developed to expand the service portfolio of Swestore, the Swedish National Research Data Storage Infrastructure operated by SNIC. The new service will have the advantage that the data will be stored in such a way that it will be interoperable with the services for European research data provided through the EUDAT Common Data Infrastructure (CDI). This will make it easier for Swedish researchers to collaborate and share data with other European researchers, as well as to have transparent access to European e-Infrastructures (such as PRACE, EUDAT and EGI), which is in line with the aims of the European Open Science Cloud initiative. While concrete terms of usage for this service are still being worked out, we invite potentially interested pilot users to contact us at [support@pdc.kth.se](mailto:support@pdc.kth.se).

# HPC Europa3

## HPCE3 Grant Approved by European Commission

Lilit Axner, PDC

The European Commission has recently granted approval for a proposal for a new European project called HPC-Europa3 or HPCE3.

HPC-Europa3 aims to ensure that a high quality service providing transnational access to the most advanced high performance computing (HPC) infrastructures is persistently available in Europe for the European research community. European researchers have been able to rely on the existence of such transnational access activity for almost two decades, and the main aim of this new innovative third edition of HPC-Europa series of projects is to fill the gap that has been left in the four years since the end of the last HPC-Europa initiative.



Above: HPCE3 kick-off meeting in Bologna, 1 May 2017

HPC-Europa3 is also a powerful force for enlarging the HPC PRACE user base by supporting researchers early on in their careers in HPC and numerical simulation, and by dedicating attention to two European regions where there is still potential to improve the HPC environment for research: the Baltic and the Western Balkans regions.

The objectives of HPCE3 are to:

- provide researchers with access to eight European HPC centres (with a target of 1,220 visits), via a single application and an international peer-review process, free of charge and with minimal administrative overhead,
- provide mentoring in the usage of the most advanced European HPC facilities,
- facilitate new scientific collaborations to be formed within an extremely wide network of scientific host labs in all the computational sciences domains,
- increase awareness of the benefits arising from the use of HPC amongst small and medium-sized enterprises or SMEs,
- increase synergy and collaboration with other HPC initiatives, and
- identify a long-term sustainability roadmap to facilitate future access to HPC resources.

The HPCE3 project is based on a program of visits for undertaking approved HPC research projects. These visits will be in the form of traditional transnational access with researchers visiting HPC centres and/or scientific hosts where the researchers will be mentored scientifically and technically to enable them to take full advantage of the HPC resources in their research. The visiting researchers will be funded for travel, accommodation and subsistence, and provided with an amount of computing time suitable for their approved research project.

The HPCE3 project started on the 1st of May 2017 with the kick-off meeting. If you have questions about the HPC-Europa3 project, feel free to contact Dr. Lilit Axner ([lilit@kth.se](mailto:lilit@kth.se)).



# Introduction to PDC Systems Course

Cristian Cira, PDC

The first Introduction to PDC course of 2017 took place on the 14th of February with ten postdoctoral researchers and Ph.D. students attending. They learned about the organization of PDC and its role in the Swedish National Infrastructure for Computing (SNIC) research community, as well as finding out about the overall mission of PDC and the staff that are available to help researchers using PDC's high performance computing (HPC) resources. The course featured a tour of the computer room, which was followed by being given an overview of the architecture of PDC's systems and the software available at PDC. Other topics that were covered included necessary basics such as instructions on getting individual accounts on PDC's computer systems, applying for a time allocation on PDC's computing resources (for doing a particular research project), submitting job allocation requests on specific computer systems at PDC (for running simulations and other jobs as part of a project's total time allocation), and how to actually run specific simulations or applications.

For those of you who are not familiar with the SNIC terminology, it is useful to know that a time allocation is the total amount of run time (in terms of core hours – which mean one hour of run time on a CPU core) that research groups are granted when they apply to SNIC for computing resources to carry out a research project. (Note that a project that is undertaken with support from SNIC can have multiple time allocations at different high performance computing centres, as well as allocations for storage.) Once a research group is granted a time allocation for PDC (which will be for use on Beskow or Tegner), they can request a job allocation (that is, to be able to use a specified number of computing nodes for a period of time on the relevant computing resources) in order to run their code. The request for resources is handled by PDC's job scheduler, which is called SLURM. Once SLURM grants the job allocation, the researchers can execute their code (which is generally referred to as running a job).



Above: Tour of PDC computer room, Introduction to PDC Systems course, 14 February 2017

## Staff Focus



Niclas Jansson

Niclas Jansson received an M.Sc. in computer science in 2008 and a Ph.D. in numerical analysis in 2013, both from the KTH Royal Institute of Technology. His research focused on all aspects of high performance adaptive finite elements methods, such as efficient linear solvers, mesh refinement and dynamic load balancing.

Between 2013 and 2016 Niclas was a postdoctoral researcher at the RIKEN Advanced Institute for Computational Science in Japan, where he was part of the application development team of the Japanese exascale program, Flagship 2020, focusing on developing extreme scale multiphysics solvers for the K computer.

Currently Niclas is the project manager of the EU-funded Horizon 2020 project ExaFLOW that addresses key algorithmic challenges in computational fluid dynamics in order to facilitate simulations at exascale.



Above: Henric Zazzi, PDC Pub 2017

# CodeRefinery Workshops

Thor Wikfeldt, PDC

A key aspect of the CodeRefinery project (<http://coderefinery.org>), which was launched last autumn, is to train Nordic research groups to take full advantage of state-of-the-art tools and practices for modern collaborative scientific software development. The CodeRefinery training takes place in the form of interactive three-day workshops involving demonstrations, live coding exercises and type-along types of presentations.

The very first CodeRefinery workshop was held in Espoo, Finland, on the 14th-16th of December 2016 in the headquarters of the CSC IT Center for Science, and the second workshop was held at the KTH Royal Institute of Technology in Stockholm on the 20th-22nd of February 2017. The most recent workshop was held in Copenhagen on the 9th-11th of May 2017. Most of the participants in the workshops were Ph.D. students or postdoctoral researchers who came from various scientific disciplines, ranging from mathematics and computer science to the physical and biological sciences, engineering and psychology. Their programming background was similarly diverse: some had extensive experience of software development in C, C++ or Fortran, others were Python or R aficionados, and yet others had less prior coding experience. However, they all shared a keen interest in improving their coding practices and learning to use modern software development tools.



*Above: CodeRefinery workshop, 14-16 December 2016, Espoo, Finland*

*Below: Coffee break at CodeRefinery workshop in Espoo*



This is exactly the aim of the CodeRefinery project: helping researchers to write modular, reusable, maintainable, sustainable, reproducible and robust software, regardless of their academic discipline or preferred programming language. CodeRefinery workshops cover a diverse range of topics including collaborative distributed version control, automated testing, documentation, Jupyter Notebooks, CMake, integrated development environments and how to manage code complexity. Overall, feedback from participants in these first two workshops has been highly positive - the core topics covered by the CodeRefinery workshops are clearly in high demand by researchers who develop scientific software in their daily work but who may not have received any formal training about using modern software development tools.

As a further step towards assisting research groups to migrate from ad hoc in-house software development solutions to state of the art collaborative infrastructures, CodeRefinery has recently deployed a web-based source code repository platform for Nordic research projects which offers free unlimited hosting of private repositories. The platform features components for issue tracking and integrated code review and will, in the near future, be connected to a continuous integration platform to automatically build software and test code changes. To sign up for this service, visit <https://source.coderefinery.org>.

The CodeRefinery project will continue to deliver three-day workshops during the coming years: the next is the “midnight sun workshop” in Tromsø on the 19th-21st of June. (For details of

*Below: CodeRefinery workshop, 9-11 May 2017, Copenhagen*



the additional workshops in 2017 and 2018, please see the [project website](#).) In tandem with delivering these workshops, the project organizes half-day or one-day events focusing on either training in one particular topic or on general seminars and discussions. An example of the latter is the get-together event that was held in Stockholm on the 19th of May (<http://coderefinery.org/workshops/2017-05-19-stockholm>). By bringing together researchers, software developers, CodeRefinery members and SNIC application experts, this meeting was designed to contribute to building a Nordic community of users and developers of research software, as well as providing an environment for those people to communicate and exchange their expertise.

Anyone interested in attending any of the CodeRefinery events can get the latest news on the project's website, <http://coderefinery.org>, or follow [@coderefine](#) on Twitter. Furthermore, CodeRefinery recently launched a web forum where researchers from Nordic universities and research centres can ask questions and take part in discussions on software development topics. To join this forum, visit <https://groups.google.com/group/coderefinery>.



*Above: CodeRefinery get together, 19 May 2017, Stockholm*



## **PRACE/BioExcel Spring School 2017: "HPC for Life Sciences"**

**Michaela Barth, PDC**

Together PRACE and the BioExcel Center of Excellence were organizing a joint Spring School on the topic of "HPC for Life Sciences". The school was to be held from the 10th-13th of April at the KTH Royal Institute of Technology main campus in Stockholm and was also being supported by the Swedish National Infrastructure for Computing (SNIC). The start of the school was unexpectedly dramatic though!

It was Friday afternoon before the Easter vacation weekend and all the preparations for the school that we had been planning for over a year had finally been completed, and I had just sent an email wishing "safe travelling" to our speakers and students when – suddenly – news of an apparent terrorist attack in central Stockholm hit the PDC office.

It felt as though the world stopped for a moment there. There was no longer any point in fiddling around with the layout of my welcome presentation. The whole city of Stockholm had entered crisis mode – text messages started arriving from the embassy of my home country





PRACE has run 32 seasonal schools since 2008 – they are top-quality face-to-face training events that are organized and held all around Europe. The themes of the different schools have ranged from programming techniques to more specialised topics, and have been aimed at different levels from generic intermediate to advanced levels. The 2017 Spring School was a great opportunity to offer a high-quality event together with the locally-based BioExcel Center of Excellence.



*Above: Lectures and coffee break at PRACE Spring School 2017*

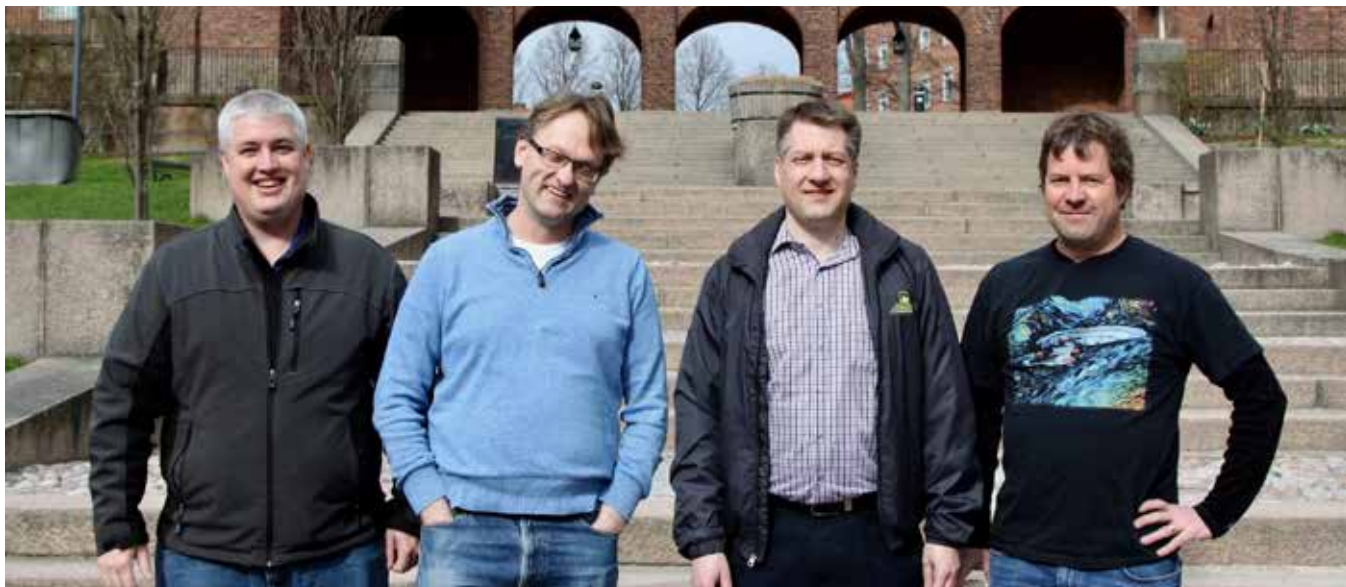
and from our children's day care centre. Obviously my embassy was not the only one to react. While watching the unfathomable news later on TV, I saw the first emails coming in from the students from Greece and Italy who had been going to attend the school saying that their embassies strongly advised against travelling to Stockholm and that they would therefore not be able to attend the school.

The PRACE Spring School had been fully booked. That was quite understandable with such a prestigious event which was offering – for the first time – the unique opportunity to learn directly from the main developers of four of the major and most popular codes for molecular modelling and simulations: VMD, GROMACS, NAMD and AMBER. When applying to hold such a seasonal school, we had promised PRACE to organize an event for 60 students and had set our own limit to an absolute maximum of 65 students. (Our limit was actually determined by the ventilation capacity in the computer lab rooms where the students, speakers and tutors would be.) We had extended the registration deadline by one week after the decision about awards of travel grants from BioExcel so that other people could apply and take the places of the students who unfortunately could not attend the school as they had not received one of the travel grants. Even though the event, as a sponsored PRACE seasonal school, was free of charge, the students still needed to pay for their own travel and accommodation. The popularity of the school was such that the students who applied were based at universities and organizations in 19 different countries – including China, the Czech Republic, India, the Russian Federation, Slovakia and Saudi Arabia – which meant significant (and sometimes prohibitive) travel costs for some of the attendees. BioExcel offered three travel grants that were awarded to students from Spain, Ukraine and Poland. Many of the “Swedish” students (that is, those with their home university or organization in Sweden) were originally from other countries as



*Above: (from left to right) Travel grant winners, Vladyslav Kravchuk, Laura Llorach Pares and Ewelina Rutkowska, with Vera Matser, the BioExcel Training Coordinator*

*Below: Main speakers at the PRACE Spring School (from left to right): John Stone (VMD), Prof. Erik Lindahl (GROMACS), Dr. Jim Phillips (NAMD) and Prof. Tom Cheatham (AMBER)*



well, which also contributed to the international flavour of the Spring School.

My BioExcel counterparts – BioExcel Project Manager, Rossen Apostolov, and the BioExcel Training Coordinator, Vera Matser – had put together an excellent programme focusing on extensive hands-on sessions. The main speakers that were expected at the school were Prof. Thomas Cheatham (University Information Technology, The University of Utah) – the main developer of AMBER, Dr. Jim Phillips (University of Illinois at Urbana-Champaign) – the main developer of NAMD, Prof. Erik Lindahl (Stockholm University) – the main developer of GROMACS (and also the chair of the PRACE Scientific Steering Committee), and John Stone (University of Illinois at Urbana-Champaign) – the main developer of VMD.

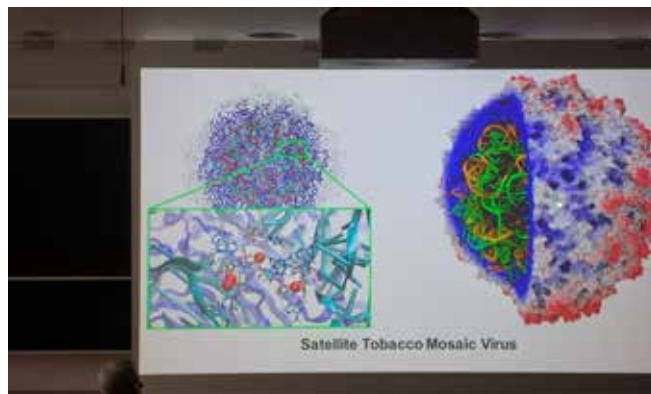
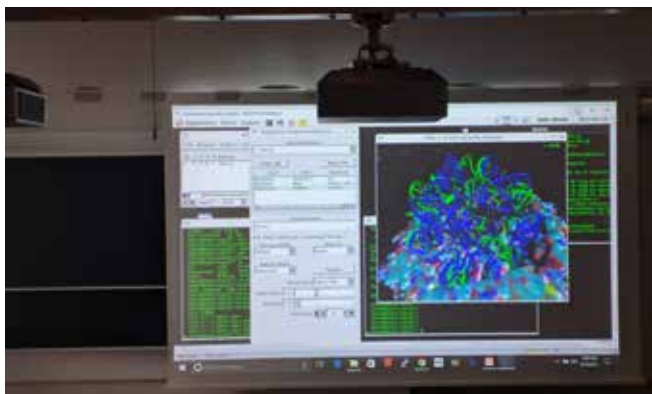
So, in a fraught weekend before the school was due to start, contingency plans and web streaming alternatives were discussed, but it became clear relatively quickly that all the speakers would come despite the situation. The “capital of Scandinavia”, as Stockholm likes to style itself, chose to approach the recent events quite matter-of-factly. Problems were solved as they appeared, people helped each other, a memorial vigil called the “Love manifestation” was organized during the weekend following the attack, and the subway and commuter trains were rolling again soon.

Finally Monday, the first day of the school, arrived and thankfully everything seemed to work smoothly: 55 students had arrived to attend what was the last of six seasonal schools within PRACE-4IP, but then, in the afternoon during the hands-on session, we unfortunately encountered some unforeseen login problems for those relying on locally provided workstations, which prevented a smooth start to the school. Fortunately, these problems were fixed fairly quickly and the school continued smoothly with the students submitting jobs to our PRACE Tier-1 system, Beskow, which is still the largest Tier-1 system within the Nordic countries.



*Above: Welcome dinner, “Syster O Bror” restaurant*

*Below: John Stone's lecture on VMD*



According to the later feedback reports, the students felt they had received helpful support and that PDC had acted professionally in solving the problem. Despite the unforeseen hassles, overall more than 75% of the students gave the spring school a good or excellent rating in the feedback form. The balance of theoretical and practical content of the classes was also given a very good rating, the overall workshop organization was rated favourably, and especially the food (including the welcome dinner at the “Syster O Bror” restaurant and the visit to the Vasa Museum and its restaurant) received many,

many positive remarks and a 60% excellent rating in the feedback form.

From the talks and the hands-on session, “awesome” John Stone especially entered the hearts of the students with an interactive and engaging demonstration on how to use VMD to do molecular visualizations. Erik Lindahl's presentation on GROMACS was also very favourably remarked on, both scientifically and in terms of presentation, by several of the students. In general, I was very impressed by the dedication, enthusiasm and keen level of interest shown by all of the main speakers who made themselves



*Above: The weather was not totally with us: what was supposed to be a spring school actually felt more like a winter school most of the time. However we still managed to catch a brief glimpse of the sun for the group picture in the KTH courtyard on the 11th of April 2017. High resolution versions of this group picture, and further pictures from the school, can found at <https://events.prace-ri.eu/event/502/page/2>.*

*Below: Prof. Erik Lindahl explaining about GROMACS*



available throughout the whole of the school (including all of the tutorials, not only their “own” sessions), as well as listening to each other’s talks, and who were always ready to assist the students.

We knew from the beginning that there would be quite a spread in the experience level of the students – basically ranging from total beginners to quite professional users for some of the software – which made it a real challenge for the presenters to find the exactly right level, but everyone I talked to said that the school had been really useful and that they had learned a lot.

The intensive time that had begun so dramatically finished, at least for me, with a tear in the eye: it was over too soon and I felt we should do something like this again soon; the well-polished program was certainly worth repeating.

For more information on future PRACE training events, please go to <http://www.training.prace-ri.eu>, and for more information about BioExcel, please visit their website (<http://www.bioexcel.eu>) where you can sign up for their newsletter.



*Above: Vasa Museum restaurant*

## Second SNIC All Hands Meeting

**Gert Svensson, PDC**

The second SNIC All Hands meeting was held on the 25th and 26th of April 2017. SNIC is the acronym for the Swedish National Infrastructure for Computing, the body that organizes the academic use of high performance computing (HPC) in Sweden. The purpose of the All Hands meetings is to share new information about SNIC and the SNIC activities and projects, from both the technical and organizational perspectives, with all the staff at the six HPC centres in Sweden that are funded by SNIC.

This year’s meeting paid special attention to the proposal for a new SNIC organization called SNIC 2.0. Today the Swedish Science Council (VR) funds SNIC, which is formally an organization within Uppsala University. SNIC distributes the funds to the six HPC centres that are members of SNIC. These centres are hosted at various universities around Sweden. Normally the university hosting each centre adds some funding and pays for some in-kind costs at their centre, like rent or power when a new SNIC computer is placed at that centre. This means that Swedish universities without a SNIC HPC centre do not have to contribute anything to the costs of HPC research in Sweden, although their researchers can use SNIC resources on the same terms as researchers based at universities that help to fund SNIC centres. This is being changed in SNIC 2.0 so that all the universities contribute. In addition to making the overall funding situation fairer, this also means that all the contributing universities will have a strong influence on the governance of SNIC.

The SNIC All Hands meeting in April was held at Hotel Arken, which is near the harbour in Gothenburg. Although Hotel Arken is located in an industrial area surrounded by petrochemical and automotive industries, not much of that is visible from inside the facility. The restaurant, for example, has a magnificent view of the entrance to the harbour.

Below: View from the Arken hotel restaurant, Second SNIC All Hands Meeting



The first day of the meeting covered news from SNIC and the different SNIC projects. The SNIC Director, Hans Karlsson, started by presenting “SNIC Today and from 2018”. Dejan Vitlacil from PDC presented the [new SNIC storage project](#) where an iRODS-based solution for storing research data will soon go into production, complementing the existing dCache implementation. Ann-Charlotte Sonnhammer from SNIC presented the [SNIC SENS project](#), which handles data storage for sensitive personal data with all the special considerations that are required for security and preserving anonymity with such data. Magnus Persson from Lund University described some of the potential security threats to computing centres and how to mitigate some of them. Torben Rasmussen from the National Supercomputer Centre, NSC, at Linköping University presented some facts and new ideas about how to best organize the provision of HPC support and in particular application experts in Sweden.

The second day of the meeting was almost totally devoted to presentations and discussions of SNIC 2.0. Sverker Holmgren, who chaired the group that developed the SNIC 2.0 proposal, gave a presentation outlining the motivation for the ideas behind SNIC 2.0. There were numerous comments and questions from the staff of the SNIC centres. Following this presentation, there was a group discussion where the attendees, in smaller groups, worked out advice for SNIC and VR about the future work and implementation of the SNIC 2.0 proposal.

## PDC-Related Events

### PDC Summer School 2017

14-25 August 2016, KTH, Stockholm

<http://www.pdc.kth.se/education/summer-school>

## HPC Sources

We recommend the following sources for other interesting HPC opportunities and events.

### CERN

<http://cerncourier.com/cws/events>  
<http://cds.cern.ch/collection/Forthcoming%20Events?ln=en>

### EGI

<https://www.egi.eu/category/events>

### HPC University

<http://www.hpcuniversity.org/events/current/>

### HPCwire

<http://www.hpcwire.com/events>

### NeIC

<http://neic.nordforsk.org>

### PRACE

<http://www.prace-ri.eu/HPC-access>  
<http://www.training.prace-ri.eu>  
<http://www.prace-ri.eu/events>  
<http://www.prace-ri.eu/news>

### SeSE

<http://sese.nu>

### SNIC

<http://www.snic.se/news-events>  
<http://docs.snic.se/wiki/Training>

### XSEDE

<https://www.xsede.org/conferences-and-events>

## CodeRefinery Workshops

**Would you like to improve your coding skills?**

If so, join us for a CodeRefinery workshop in one of the Nordic countries in 2017 or 2018. Further details are available here:

<http://coderefinery.org/workshops>