PLAN-E Workshop eSkills
Dublin, May 9 2016

Introductions

Four invited contributions embedded the following discussion:

Open Science Conference (Patrick Aerts)

The report on the Open Science Conference in relation to the eSkills discussion was mainly to stress the importance of a general change of mind if it comes to sharing data and software and creating awareness about FAIR and Open Science principles. This does not happen without initiatives and should become part of the mind-set education of young students.

eSkills for Science (Lene Krøl Andersen)

Lene Krøl Andersen, DeIC eScience Center, Denmark: Besides informing the audience about the escience developments in Denmark and the particular role of DEIC in that- sketched eSkills as one of the more urgent challenges faced. New generations can be lectured in due course on how to approach Big Data (analysis, data set interaction, data base technology, etc.), but today's PhD students, post-docs, and other staff need to be educated as well, in order to get access to the wealth of information hidden behind Big Data. Education is also required to deploy efficiently and expediently all elements of modern ICT-infrastructures, not just its single components. Hence, both tracks (curriculum education and courses for more advanced scientists) need full attention.

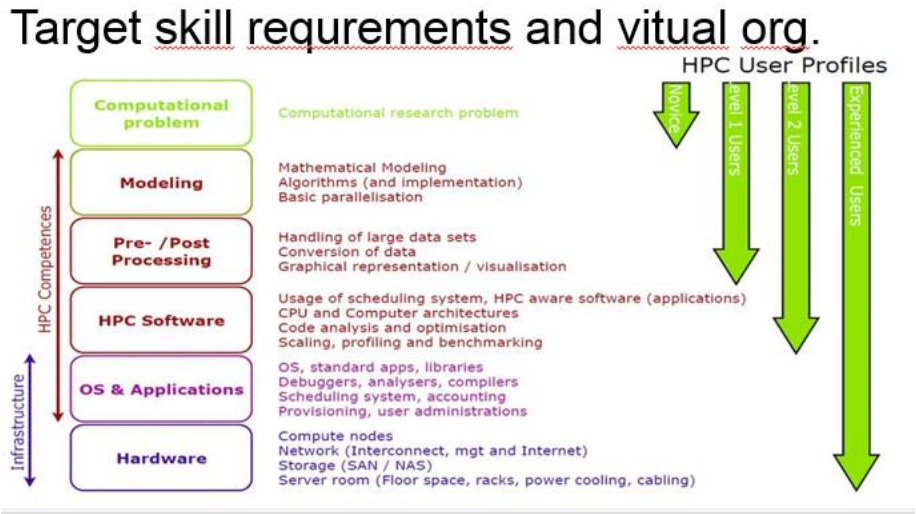

Figure 1: Overview of eSkills for supercomputing (Danish eInfrastructure Corporation)

DeIC has established a national eScience Portal: https://vidensportal.deic.dk/?language=en which informs and identifies eScience resources around the universities, archives, libraries etc. and provides user information on; how to get access to a national supercomputer, data management, legal issues, eScience contact points through the country, and other related eScience topics. DeIC eScience Center is furthermore working on a national eScience training platform to be incorporated into the DeIC eScience Portal – and the ambition is to be able to launch this national eScience gateway to eScience courses hosted by various eScience-parties, during autumn 2016.

The Nordic eInfrastructure Collaboration (NeIC) has furthermore put high emphasis on eScience and eInfrastructure skills among scientists and e-infrastructure personnel in the Nordic region, for the purpose of raising the quality of the scientific output overall.

Within the broad concept of training there is a distinction between courses which give academic credits, and other types of training. This would also include best practice guides and online training material. NeIC and DeIC takes an interest in all types of training that can further the interests of the end users and science and focuses on working with the e-infrastructure providers.

The NeIC training calendar (https://neic.nordforsk.org/training/) shows upcoming training events throughout the 5 Nordic countries. The training events/courses/seminars are either arranged by NeIC or partners, or simply of interest to NeIC and its stakeholders. NeIC has set up a world-readable calendar where NeIC and stakeholders can share information on upcoming training events. Training coordinators from stakeholder organisations are invited to add information on their upcoming training events to this calendar. The information is shared in this calendar for the purposes of coordination of efforts, and for advertising the training events to potential interested participants.

Nordic training coordination
There are several use cases where Nordic coordination of training can help contribute to the goals and purposes outlined above.
Potential uses for training coordination:

- Allocating seats reserved for Nordic participants at current training events
- Broadening the training offering and audience
- NeIC acting as a facilitator for established courses being in another country
- Technical support to make training online and accessible
- Mobility enhancement amongst trainers and trainees

PLAN-E must ensure to interact with respectively the *NeIC Pool Comptencies* and *NeIC CodeRefinery* steering groups which are in the process of establishing a Nordic eScience resource platform to enhance the Nordic eSkills. The eSkill related NeIC projects will run until 2018. National surveys in the Nordics are furthermore in process and interlinked through NeIC. These will feed into identification of specific needs for user-support, tools, competencies and other resources. Denmark expect to have results by the end of 2016 on a national survey regarding *usage of supercomputing within Danish research*.

eSkills in education and post graduates (Sverker Holmgren)

Sverker Holmgren sketches the progress made by the Swedish national graduate school in scientific computing (NGSSO, http://ngssc.se/ (in Swedish)). NGSSO went through two major phases. One from 1996-2006 and one from 2007-2012.

In phase 1 the goal was: “to increase the competitiveness of research and development in academia and industry in Sweden by increased and improved use of advanced computational techniques”. In phase 2 the goals was set to: “Produce Ph.D.’s that combine skills in important applied disciplines with a broad knowledge of advanced numerical and computational techniques and to encourage interdisciplinary

work and spread the use of computational techniques into new areas of research and development. This was achieved through the following approach:

- Disciplinary training at home institutions
- Intensive courses to provide the computational and mathematical core
- Courses developed and given by leading experts on different locations within Sweden.

To set up and fill the courses with students the following measures were taken:

- The courses were normally organized by two universities, uses local expertise at the university
- NGSSC funds development and giving of the courses
- NGSSC students receive funding for travel and stay to participate in the courses
- The courses are open to other students
- Normally given at the end/beginning of a semester.

There were wo types of courses: basic and advanced.

Basic:

- Numerical Methods in Scientific Computing
- Programming in Science and Technology
- HPC I: Parallel Computers and Programming
- HPC II: Algorithms and Applications
- Data Management and tools.

Advanced:

- Specialized subject
- Each course given once or possibly repeated every second year
- Given in the NGSSC format or as a summer school
- May be given in collaboration with research groups or other research schools

Topics:

- Finite Element Modelling and Simulation
- Computational Methods in Statistics with Applications
- Computational Biophysics/Pharmacology
- Data analysis in Genetics
- Other

Experiences gained from the project are:

- Local support from supervisor important
- Teachers must be able to handle a vast difference in scientific background

- Intensive format makes it possible to fit in the courses with other activities
- Interdisciplinary interaction mainly between the students.

Skills for eScience (Joshua Ryan-Saha)

Joshua Ryan-Saha, presently employed by the Scottish Datalab facility, gives an overview of data science related activities and skills requirements to be delivered by different societal or stakeholder groups. The Datalab in Scottland is an example of an innovation effort funded through The Innovation Centre Programme, whose objectives are to enhance innovation and entrepreneurship across Scotland's key economic sectors, create jobs and grow the economy. To achieve this, the programme's main objectives are:

- To offer collaborative knowledge exchange and research activities to help solve industry defined problems and co-create innovative opportunities for growth.
- To enhance two way knowledge exchange between universities, industry and others towards realising tangible benefits for businesses while also stimulating and challenging the Scottish research base.
- To provide an environment that supports the development of the next generation of business innovators, academics and entrepreneurs in Scotland and a culture change towards greater and more effective university / industry collaboration.
- To simplify the innovation landscape in Scotland through creating conduits to the university knowledge and expertise for all businesses in Scotland and being complementary to and exploit existing initiatives such as Interface and the planned Single Knowledge Exchange Office.

Jusua's presentation is based on two recent and officially presented reports:

- http://www.nesta.org.uk/sites/default/files/analytic_britain.pdf and
- http://www.nesta.org.uk/sites/default/files/skills_of_the_datavores.pdf.

He identified three categories of stakeholders:

1. Schools and Colleges; 2. Universities and Vocational education; 3. Labour Market and Industry. Each to be allocated responsibilities in data science education:

1.1. Stronger teaching of mathematics and statistics in schools and colleges;
1.2. More and better information about analytical career prospects and role models in schools and colleges;
1.3. Embed data analytics in other subjects;
1.4 Support development of extracurricular data activities.

2.1 increase the visibility of strong data analytics courses;
2.2 Embed quantitative analysis acrosss disciplines;
2.3 Increase the supply of high-end analytical talent;
2.4 Foster interdisciplinary research and skills development programmes.

3.1 Create a cross-cutting taskforce around data analytics;
3.2 Actively convene industry and community analytics networks;
3.3 Support innovative interventions enabling local authorities to boost local analytical skills;
3.4 Raise awareness of the value of data for business and local public services;
3.5 Develop innovative solutions for data analytics training.

The underlying reports are very suitable reading for this topic. They will be put on the PLAN-E website for reference.

For the workshop the participants were split into two groups which both discussed both topics:

- eSkills for research, courses at post doc/PhD level;
- eSkills in education: what has to brought into the university curricula and how to achieve this.

After that the output from both groups was plenarily shared.

Courses material for PhD students, post-docs and staff

Issue 1: there are courses, perhaps enough in many domains, but how to get the interest of scientists and students?

- ☞ Provide information along the lines of existing gremia, such as committees and councils;
- ☞ Require regular skills updates of staff

Issue 2: What kind of courses are needed?

- ☞ Any kind. Traditional courses, two or three day courses, physical courses and electronic courses;
- ☞ The intermediate teaching role is considered indispensable, irrespective of existing digital and electronically accessible courses;
- ☞ Consider high resolution digital courses. They will have more impact and effect than traditional video quality.

Issue 3: what about sharing courses?

- ☞ Consider a platform of curricula/training courses in English (something for PLAN-E?)
- ☞ Software Carpentry (http://software-carpentry.org/) is a good example. (NLeSC joined this group last year)
- ☞ Centers are often protective towards freely distributing course material. In contrast in the US, all types of courses (MPI was mentioned explicitly) are freely available. There one is *proud* to present these courses;
- ☞ In Italy the EduOpen organization was set up (see www.eduopen.org/en/), involving 14 universities sharing their (electronic) courses. One can follow the on-line courses and earn certificates. The courses are free, the certificates require a fee to be paid;
- ☞ A limited scope set of courses on using e-infrastructures is provided through the H2020 funded sci-gaia project (www.sci-gaia.eu) which took over activities in that domain from the earlier FP7 supported chain-project (www.chain-project.eu), which was about co-ordination and Harmonization of advanced e-Infrastructures for Research and Education Data Sharing;
  Sci-gaia is a project with full name “Energising **Sci**entific Endeavour through Science **Ga**teways and e-Infrastructures in **A**frica” and the following work packages:

- ☞ WP1: Promote the uptake of Science Gateways and e-Infrastructures in Africa and beyond
  WP2: Support new and already emerging Communities of Practice
  WP3: Strengthen and expand Science Gateway and e-Infrastructure related services
  WP4: Training, dissemination, communication and outreach.
- ☞ Example syllabi can also be examined at the open syllabus project. http://opensyllabusproject.org/.

Issue 4: accreditation. What about the quality of material and allocation of credits to those that took knowledge of the material?

- ☞ It is suggested to organize a system with (i) eScience learning results; (ii) criteria for eScience courses quality so these could be checked against or used when compiling new course material.

Educational curricula

Issue 5: The importance of eSkills in early education

- ☞ e-Science: not just a tool but a discipline, from *kindergarden* to researchers
- ☞ Early education → project-based work that introduces e-science concepts
- ☞ Standing on the shoulders of others (software data) speeds up science and discovery, so teach students about this in an early stage
- ☞ EUNIS (http://www.eunis.org/) body to address on this topic. EUNIS is the European University Information Systems cooperative body. It's focus is on education.

Issue 6: What topics should be part of the education?

- Basic skills widely taught - general upskilling
- Best practice software engineering methods are not generally taught, e.g. software carpentry
- Humanities and Social Sciences: *in some countries* need to get technically up to date including data science skills and technologies
- Need to know about basic data handling/analytics
- Essential Maths/Stats as foundation course to broad areas
- Advanced Excel
- Validation methods
- Parallel/efficiency software development
- Data efficiency and knowledge of reuse strategy
- Tools and knowledge of data management tools
- Consider reproducibility as basis for sound science
- Open science considerations
- Data Management Planning and Software Sustainability
- Software language choices for different types of problems

Issue 7: where to find information on educational curricula?

- ☞ The ACM published a list of courses on parallel processing, known as the NSF/IEEE-TCPP Curriculum Initiative on Parallel and Distributed Computing – Core Topics for Undergraduates. See for details: grid.cs.gsu.edu/~tcpp/curriculum/. There one can also find a pdf file on the

curriculum initiative for parallel and distributed computing. http://grid.cs.gsu.edu/~tcpp/curriculum/sites/default/files/NSF-TCPP-curriculum-version1.pdf

- ☞ Publish an overview of curricula in the same fashion as the ACM/IEEE did for Computer Science: https://www.acm.org/education/CS2013-final-report.pdf. The report is named Computer Science Curricula, December 2003; Curriculum Guidelines for undergraduate Degree Programs in Computer Science.
- ☞ There exists a European Credit Transfer and Accumulation System (ECTS) which is "a standard for comparing the study attainment and performance of students of higher education across the European Union and other collaborating European countries"[1].

Issue 8: what practices exist in different countries (participating in the discussion)?

- ☞ UK: Centres for Doctoral Training. Centres for Doctoral Training (CDTs) (and sometimes also abbreviated to DTC) are one of the three main ways by which EPSRC provides support for Doctoral Training. The other routes are the Doctoral Training Partnerships (DTP) and Industrial CASE Studentships. Many universities participate in these trainings.
- ☞ SWE: once entered in the system, students at undergraduate level are able to get input courses from other institutions.
- ☞ NL: similar to SWE in that respect.
- ☞ SA: Master level funding for the student not the institution in SA.
    - Call out to form consortia to provide coherent skills and coverage
    - Curricilum then developed and agreed by ministry
    - National masters program in eScience
    - Mobile students who will travel around lecturers
- ☞ CH: Development of interconnected courses to ensure that technical employees can link between scientists and e-infrastructure. Supporting these is a longer term challenge. Presently building the support from admin IT -> supporting science. Each system operating has SLA with different levels available.
  It is objected by one of the participants, that there is a serious danger of having e-Science engineers too closely linked to HPC-centres is that they may unduly focus on HPC systems and support, rather than providing the bigger picture.
- ☞ Fr: Central centres with research engineers to support research, 24-7 support available.
- ☞ RO: There are both public and private universities. Public universities are of high quality. Funds are given to institutions directly. Students can move but the money doesn't move with them.
    - Staffing shortage in computing so extending curriculum extremely difficult
    - Non-computing areas has excess staff.
    - MSc with a particular centre but much higher mobility.
    - PhD difficult to do computing within other area PhDs.
    - Challenging conditions for both teaching and learning due to market requirements
    - Industrially required skills already being taught to the extent that students starting to already have jobs before they finish
    - Departments generally provide support to e-infra including high availability.

[1] From https://en.wikipedia.org/wiki/European_Credit_Transfer_and_Accumulation_System