H2020-FETHPC-2014

Coordination of the HPC strategy

EXDCI

European eXtreme Data and Computing Initiative

Grant Agreement Number: FETHPC-671558

D5.3
Report on Promotion of HPC as a Career

Final

Version: 0.2
Author(s): Catherine Inglis, EPCC
Date: 17.02.2017

Project and Deliverable Information Sheet

EXDCI Project	**Project Ref. №: FETHPC-671558**	
	Project Title: European eXtreme Data and Computing Initiative	
	Project Web Site: http://www.exdci.eu	
	Deliverable ID: < **D5.3**>	
	Deliverable Nature: <DOC_TYPE: Report>	
	Dissemination Level: PU	**Contractual Date of Delivery:** 28 / February / 2017
		Actual Date of Delivery: 27 / February / 2017
	EC Project Officer: Béatrice Marquez-Garrido	

* - The dissemination level are indicated as follows: **PU** – Public, **CO** – Confidential, only for members of the consortium (including the Commission Services) **CL** – Classified, as referred to in Commission Decision 2991/844/EC.

Document Control Sheet

Document	**Title: Report on Promotion of HPC as a Career**	
	ID: D5.3	
	Version: <0.2>	**Status:** ***Final***
	Available at: http://www.exdci.eu	
	Software Tool: Microsoft Word 2013	
	File(s): EXDCI-D5.3-v0.2.docx	
Authorship	**Written by:**	Catherine Inglis, EPCC
	Contributors:	Maria Ribera, BSC David Henty, EPCC Blaž Rodič, University of Ljubljana
	Reviewed by:	François Bodin, IRISA Iñigo Yenes, PRACE
	Approved by:	MB/TB

Document Status Sheet

Version	Date	Status	Comments
0.1	07/February/2017	Draft	First draft for internal review
0.2	17/February/2017	Draft	Revised after internal review
0.2	24/February/2017	Final version	Approved by EXDCI Management Board and Technical Board

Document Keywords

Keywords:	PRACE, HPC careers, young talent, skills, training, community, staff retention

Table of Contents

List of Figures

References and Applicable Documents

List all external documents referenced in this document

[1] https://ec.europa.eu/digital-single-market/en/news/staff-working-document-implementation-action-plan-european-high-performance-computing-strategy

[2] https://www.epcc.ed.ac.uk/discover-and-learn/resources-and-activities/what-is-a-supercomputer/wee-archie

[3] https://www.mathe-im-advent.de/fotos_cray/

[4] https://www.hpcwire.com/2016/11/09/bof-boost-supercomputing-outreach-skills/

[5] https://www.epcc.ed.ac.uk/discover-and-learn/resources-and-activities/what-is-a-supercomputer/design-and-manage-your-own-supercomputer

[6] http://www.prace-ri.eu/daretothinktheimpossible/

[7] http://www.prace-ri.eu/daretothinktheimpossible/shootingstars/

[8] http://sc16.supercomputing.org/sc-archive/bof/bof_pages/bof189.html

[9] https://www.hpcwire.com/2009/07/23/big_ideas_in_hpc

[10] https://www.ted.com/

[11] http://bigthink.com/

[12] https://www.top500.org/

[13] https://exdci.eu/jobs-and-training/hpc-career-case-studies

[14] http://www.womeninhpc.org/

[15] http://www.hpc-diversity.ac.uk/

[16] https://exdci.eu/jobs-and-training/job-portal

[17] http://www.archer.ac.uk/community/champions

[18] https://www.xsede.org/campus-champions

[19] http://edison-project.eu/edison/education-and-training-champions

[20] https://www.software.ac.uk/

[21] http://www.prace-ri.eu/prace-advanced-training-centres/

[22] http://www.fib.upc.edu/en/masters/miri.html

[23] http://www.epcc.ed.ac.uk/msc/

[24] http://ec.europa.eu/education/resources/european-credit-transfer-accumulation-system_en

[25] http://www.epcc.ed.ac.uk/online-courses/

[26] https://developer.nvidia.com/educators

[27] https://developer.nvidia.com/teaching-kits

[28] http://bcw.ac.upc.edu/PUMPS2017/

[29] https://eventi.cineca.it/en/hpc/26th-summer-school-parallel-computing

[30] http://www.prace-ri.eu/csc-summer-school-in-high-performance-computing-2017-jun-27-jul-5/

[31] http://www.imperial.ac.uk/computational-methods/news-and-events/hpc-2016/

[32] http://www.ihpcss.org/

[33] www.anl.gov

[34] http://extremecomputingtraining.anl.gov/2016-videos

[35] https://www.futurelearn.com/courses/supercomputing

[36] http://bjc.berkeley.edu/; https://www.youtube.com/watch?v=eqbgqSe_hvU

[37] https://code.org/
[38] http://csunplugged.org/
[39] https://euraxess.ec.europa.eu/jobs/hrs4r
[40] https://www.bsc.es/news/bsc-news/lets-stay-connected-bsc-folks
[41] http://www.software.ac.uk
[42] http://www.rse.ac.uk/
[43] http://edison-project.eu/edison
[44] http://edison-project.eu/data-science-model-curriculum-mc-ds

List of Acronyms and Abbreviations

Below is an extensive the List of Acronyms used in previous deliverables. Please add additional ones specific to this deliverable and delete unrelated ones.

ACM	Association for Computing Machinery
ATPESC	Argonne Training Program on Extreme-Scale Computing
BoF	Birds-of-a-Feather session
CFD	Computational Fluid Dynamics
CS	Computer Science
CSC	Finnish IT Center for Science
D	Deliverable
EC	European Commission
ECTS	European Credit Transfer and Accumulation System
EPCC	Edinburgh Parallel Computing Centre
EU	European Union
EXDCI	European Extreme Data and Computing Initiative
H2020	Horizon 2020 – The EC Research and Innovation Programme in Europe
HLRS	Höchstleistungsrechenzentrum Stuttgart (High Performance Computing Centre in Stuttgart)
HPC	High Performance Computing
IEEE	Institute of Electrical and Electronics Engineers
ISC	ISC High Performance – the Event for High Performance Computing, Networking and Storage
IT	Information Technology
M	Month
MOOC	Massive Open Online Course
MSc	Master of Science
ORNL	Oak Ridge National Laboratory
PRACE	Partnership for Advanced Computing in Europe
RSE	Research Software Engineer
SC	The International Conference for High Performance Computing, Networking, Storage and Analysis
SSI	Software Sustainability Institute
TED	Technology, Entertainment and Design (non-profit media organisation)
UK	United Kingdom
US	United States
WP	Work Package
XSEDE	Extreme Science and Engineering Discovery Environment

Executive Summary

The shortage of personnel with expertise in HPC skills has been identified as a barrier to increasing the uptake of HPC in industry, which in turn represents a major risk to European competitiveness. This issue is the focus of EXDCI Work Package 5 – Talent Generation and Training for the Future, the goals of which are to support talent generation, facilitate HPC staff recruitment, and identify and meet future training needs

This deliverable, *D5.3 – Report on Promotion of HPC as a Career Choice*, looks at how HPC is currently promoted as a career option to the workforce of the future, and examines what should be done to increase the appeal of careers in HPC to young people. It also looks at the problem of how to retain valuable staff within the HPC arena, as the failure to keep hold of individuals with relevant skills also contributes to the shortage of expertise in the workforce.

1 Introduction

In ever more areas of academic research, industry, commerce and government, large-scale computational approaches are used for simulation and optimisation, resulting in increased productivity in research and innovation. However, while the demand for software development has risen, the availability of scientific software developers has not, largely due to poor long-term planning and a lack of succession policies. The shortage of personnel with expertise in HPC skills in particular continues to be a barrier to increasing the uptake of HPC in academia and industry, and this has been identified as a major risk to European competitiveness[1].

In part, this shortage is due to the relatively low profile of HPC as a career choice – many young people are not sufficiently aware of the variety of opportunities that working in HPC can present to consider it as a possible career. Some do not even realise that HPC exists outside a few ultra-specialised areas (such as weather prediction).

The problem is further exacerbated by poor retention of skilled staff within the sector. HPC needs to be an attractive career choice not only for those who have yet to embark on their professional career, but also for those who are already in it.

Another contributing factor is increased competition from the Big Data sector, which is also rapidly expanding and faces a similar shortfall of staff with appropriate skills. In the absence of suitably qualified Big Data engineers, HPC engineers may be seen as "the next best thing", and this represents an additional threat to the sector, and one which is unlikely to diminish in the near future.

Task 5.1, Supporting Talent Generation, aims to examine ways to promote the importance of HPC skills to students, and to raise the profile of the many job opportunities open to those with HPC skills, both within the HPC community and in professions where these skills are required. This report looks at how to attract and retain skilled and enthusiastic people into careers in HPC.

[1] See https://ec.europa.eu/digital-single-market/en/news/staff-working-document-implementation-action-plan-european-high-performance-computing-strategyimplementation-action-plan-european-high-performance-computing-strategy

2 Attracting young talent to careers in HPC

The HPC sector is suffering from a shortfall of potential employees with suitable skills. In order to address this, we need to understand why this is happening. Is it that careers in HPC are not attractive to new graduates? Do graduates lack appropriate skills to apply for the vacancies? Perhaps they are not even aware of the opportunities which HPC can offer?

In fact, a combination of all of these factors is probably to blame, and a multi-pronged approach is needed to tackle these various issues.

2.1 Public engagement: raising the profile of HPC

2.1.1 *Outreach to the public: a different focus*

HPC centres are often not very good at telling the world at large about what they do. In publicising their services and the research to which they contribute, most centres aim their efforts primarily at existing or potential funders. This might mean publicising research results arising from use of the facilities, in order to justify to the research councils their current funding or bid for future funding. It may be publicity aimed at new communities, to increase the number of users and so ultimately further justify funding from the research councils. It may be publicity aimed at potential industrial partners who would pay to use the facilities.

Yet few efforts are directed towards the general public. It is even often the case that researchers working in universities which have their own HPC facilities are only vaguely aware of their existence.

Public engagement activities may not be seen as core business activities for HPC centres and manufacturers. However, the prevailing approach to HPC recruitment involves letting people find their own path to HPC, recruiting people who happen to have become sufficiently specialised on their own initiative, or who have followed a specific interest in their earlier career choices. This includes those who have chosen to do a specialised postgraduate degree focusing exclusively on HPC, and those who have taken some optional courses while studying for an undergraduate or postgraduate qualification in a subject area with a strong reliance on HPC methods for their research goals.

By undertaking public engagement activities, HPC stakeholders are brought into contact with a completely different population, and often this is an opportunity to reach young people, from schoolchildren to students, and to inform and inspire these people who will be the workforce of the future. Anecdotally, we know of students whose first point of contact with HPC at a science festival event inspired them to pursue an MSc in HPC and then go on to work in HPC jobs after their studies.

2.1.2 *Capturing the attention of the wider public: mini-supercomputers and competitive games*

The ground-breaking nature of much of the research carried out on HPC systems means that it should be relatively easy to produce attractive outreach material and demos to use at public engagement events. Not everyone will be interested in technical specifications of individual supercomputing systems, but most people can relate to case studies showing, for example, how supercomputers have helped to make cars or planes more aerodynamic and therefore faster or more efficient.

Due to the diverse nature of the audience at public engagement events, outreach material needs to be designed carefully and ideally should be adaptable in order to stimulate people of different

ages and with different skills.To have maximum impact, there should be highly-visual displays and hands-on demos, to attract attention and then keep people engaged. Ideally, demos should work on many levels, starting at a simple level but providing opportunities for detailed discussion. While children may enjoy the "game" aspect of a demo, their parents may be more interested in extending the principles of the demo to things that affect their daily life, or in learning more about the hardware.

One fun way to demonstrate the concepts of supercomputing to a general audience is to create a "mini-supercomputer". Oak Ridge National Laboratory (ORNL) in the US led the way with their Tiny Titan, a portable parallel computer (named after ORNL's Titan supercomputer), built from 9 Raspberry Pi units and designed to teach aspiring scientists the basics of parallel computing. Following in their footsteps, EPCC in the UK has built Wee Archie, named after the UK's national supercomputer ARCHER[2]. Wee Archie is composed of 18 Raspberry Pi units, each with an LED display that lights up when in use, providing a visual display that helps demonstrate how multiple processors work in parallel to solve complex tasks. In Germany, HLRS has created "Konni"[3], a model of the HLRS supercomputer Konrad, which is also used in outreach activities.

Figure 1 Wee Archie, a suitcase-sized supercomputer made of Raspberry Pi units

In the case of Tiny Titan, a local TV station ran an item on how the "tiny supercomputer could come to schools" to raise awareness in the local area, and such "suitcase supercomputers" could also be taken to summer schools in relevant computational science disciplines. The portability and relative novelty of these systems gives them a high impact factor.

EPCC has gone a stage further and developed Wee Archlet, composed of just five Raspberry Pi units – held in a Lego case, to increase the appeal to the younger generation. Wee Archlet is designed to be cheap and easy to build while still demonstrating the key concepts of parallel computing. Detailed instructions are available online for schools and groups to build the machine themselves. There is a software suite, combining the models which run on Wee Archlet

[2] https://www.epcc.ed.ac.uk/discover-and-learn/resources-and-activities/what-is-a-supercomputer/wee-archie
[3] https://www.mathe-im-advent.de/fotos_cray/

with a client visualisation application for all major platforms so that the public can simply "point and click" to run demos. Demos include dinosaur racing, weather simulation, a molecular dynamics demo (based on mouse urine), and a Computational Fluid Dynamics (CFD) demo illustrating air flow over aerofoils.

In the view of Nick Brown, an Applications Consultant at EPCC, "demos and gamification are a really powerful way of engaging the public in our field"[4]. A successful technique is first to grab the public's attention with something "cool", and then have some kind of competition (a leader board or race) to keep them engaged – and ideally give them a sense of achievement and progress. The dinosaur racing used on Wee Archie and Wee Archlet is a good example of this and has proven very popular with schoolchildren at outreach events. One of the key messages that this gets across is that parallel simulation can be used where experimentation is difficult or impossible to do.

It should be noted that it is not necessary to develop complex demos in order to demonstrate the basic concepts of parallelism. It is easy to set up a simple ball-sorting exercise, where participants have to sort coloured balls into different boxes as quickly as possible, firstly individually and then in pairs or groups. This activity is highly effective in demonstrating concepts such as parallel speed-up and overheads, is great fun, and is always popular with the public. Puzzle sheets, colouring-in, and even the use of cartoon mascots for outreach material aimed at younger audiences can all be effective at drawing people in and getting them interested.

It is very difficult to measure the impact of general outreach activities – realistically you might not attract these people to HPC directly, but you may succeed in enthusing them about science in general and spark some interest in the general area of HPC. Impact will generally not be realised until some years after the event, making it difficult to measure beyond the occasional anecdotal success story. While it gives no meaningful estimate of the final impact, the best measure of success is the number of people who have interacted with the presenters.

2.1.3 *Inspiring interest through online activities*

Attending public science festivals and events can be expensive due to registration fees, cost of transporting outreach materials, and the staff time that must be invested in preparation and attendance. It is therefore also important to have online material targeted at the general public, with a view to sparking their interest in HPC. This should address straight-forward questions such as:

- What is a supercomputer?
- Why are supercomputers important?
- How do you use a supercomputer?

The content should be written in simple language, avoiding technical jargon, and focusing on simple concepts and issues of direct relevance to a general audience, with real-life examples to which people can easily relate.

Online activities can also be developed, such as the "Build your own supercomputer" challenge[5] developed by EPCC. Working within a monetary budget, participants select different types of hardware and see how many jobs they can run in two minutes. Completed jobs bring in money which can be spent on upgrading or repairing hardware, or making the system more energy-

[4] https://www.hpcwire.com/2016/11/09/bof-boost-supercomputing-outreach-skills/

[5] https://www.epcc.ed.ac.uk/discover-and-learn/resources-and-activities/what-is-a-supercomputer/design-and-manage-your-own-supercomputer

efficient and thus reducing costs. This teaches some basic hardware concepts – what the constituent parts of a computer are and how they affect performance and cost – as well as a bit about the sort of jobs that run on HPC systems.

The PRACE "Dare to Think the Impossible"[6] outreach initiative used language designed to engage children through presenting HPC as a challenge which requires daring and imagination, which was neatly encapsulated within the name of the campaign. With an emphasis on "solving mysteries" and statements such as "The possibilities are limited by only one thing. Your imagination!", "dream up solutions that will make the world a better place", and "Don't be afraid to explore", the campaign appealed directly to the enthusiasm and optimism of young people, addressing their hopes and dreams, and challenging the commonly-held misconception that computing is dull.

A physics puzzle game, Shooting Stars[7], was developed for the "Dare to Think the Impossible" campaign. This game simulates a star system with its orbiting planets and gravitational forces, and the aim is to keep as many planets as possible in orbit in different scenarios. Social media were used to try to connect more easily with the target group, so the game was also developed as a mobile app, and research highlight videos were edited to be more accessible and made available on YouTube.

2.1.4 *Expanding horizons: targeting teachers and other adults*

"Dare to Think the Impossible" was targeted not only at young people, but also at their teachers, with material available for teachers to download to use in class. The ARCHER Outreach team have also developed teachers' packs to help them to build their own Wee Archie systems and run real HPC codes on them. This helps to expand the impact of the outreach material developed, reaching classes anywhere in the country, who may not have the opportunity to benefit from organised school visits or trips to science festivals.

It has been recognised that the overheads for producing material are high if there is no specific funding available, but one suggestion from the SC'16 BoF "HPC Outreach: Promoting Supercomputing to the Next Generation"[8] was to develop a shared Europe-wide repository of material, perhaps working with existing networks such as Inspiring Science Education and European Schoolnet to spread the message and disseminate any materials that can be downloaded and used in the classroom.

Looking at communications aimed at a slightly older age group, it has been noted[9] that HPC visionaries are almost entirely absent from arenas such as TED talks[10] and Big Think[11]. Michael Feldman concludes in his HPCWire article, "This may be because most of the top-end academicians in high performance computing are used to attending the same ACM and IEEE sponsored events every year. At the other end are the vendors, which go mostly to trade shows. HPC users, like Makram, may venture further afield, but are usually focused on talking about their applications rather than the wonders of supercomputing."

Overall, HPC stakeholders could be much better firstly at making the general public aware of their very existence, and then at communicating the importance of the work which HPC

[6] http://www.prace-ri.eu/daretothinktheimpossible/

[7] http://www.prace-ri.eu/daretothinktheimpossible/shootingstars/

[8] http://sc16.supercomputing.org/sc-archive/bof/bof_pages/bof189.html

[9] https://www.hpcwire.com/2009/07/23/big_ideas_in_hpc

[10] https://www.ted.com/

[11] http://bigthink.com/

facilitates. This would not show great immediate results, but could over time inspire people to make career choices at various stages that would lead them to HPC.

2.2 Promoting HPC as a career

In order to consider HPC as a career, one must first know that such an option exists. People are perhaps aware that HPC is used in very specialised areas, such as weather forecasting, but few outside the HPC ecosystem have any idea how many supercomputers there are in the world, where they are, who uses them, and what they are used for.

For example, almost everyone working in the HPC sector will be aware of the Top 500 list of the most powerful HPC systems[12]. Yet most other people might be surprised to know that there are enough such systems in the world to warrant not only producing this list, but to update it every six months. Many people are unaware that such systems may be located in nearby research organisations, and that they have hundreds or even thousands of users. HPC is more prevalent than most people think – for many people it would simply not occur to them to consider this as a career choice, unless they had already had some exposure to it.

Therefore, HPC centres and manufacturers of HPC hardware and software need to raise their profile in a more general way. As their products are not aimed at general consumers, it is not common to see general awareness-raising advertisements on TV or at the cinema, in the printed press, at sports events, or even in online advertising in places such as YouTube. The return on investment for advertising in such places is considered too low, as the target audience is such a small percentage of the total audience reached. But without ever taking a more general approach, there is a danger of largely only "preaching to the converted" when advertising to a more targeted demographic.

As discussed in section 2.1, public engagement can help raise some general awareness among a very general audience, but can only have impact in the long term. In order to recruit talented people today, a more targeted and immediate approach is needed. Clear information on the range of opportunities needs to be made available in a highly-visible part of an organisation's webpages, and printed material should be developed for face-to-face events. This should aim to answer very general questions that people new to HPC may have: What do HPC centres do? What sort of work do their staff do? Why should I want to work there? How can I prepare myself to get a job there?

Personal case studies are an ideal way to highlight the variety of work that exists in the HPC world, and how individual career paths have developed. It is important to emphasise that having experience in HPC is not a prerequisite, and that many people come from domain-specific backgrounds where they have acquired only some general programming skills.

Including informal photos of the staff featured can be a useful way to grab people's attention, and can also go a long way to breaking down preconceived ideas of the type of person who works in computing – e.g. that careers in HPC are only for men, or for people who have had a love of programming since they were six years old.

EXDCI is producing a series of careers case studies[13] which attempt to address this. These case studies show a mix of people from different backgrounds, doing different jobs, and include some mention of their interests outside of work, to give a more personal feel and emphasise that people working in HPC are not necessarily all about computers all of the time.

[12] https://www.top500.org/

[13] https://exdci.eu/jobs-and-training/hpc-career-case-studies

Many women think that computing careers are not for them, and while statistics show that the sector is still very much male-dominated, this is slowly changing, and it is very important to show that the opportunities in HPC are equally open to women. The same can be said for almost any of the protected minorities – they may be under-represented, but they are not completely unrepresented, and should in no way feel excluded. Efforts must be continued to emphasise this.

Two excellent initiatives have been established to encourage people from different backgrounds to get involved: Women In HPC[14] and Diversity in HPC[15].

Women in HPC strives to recruit and retain women in the international HPC workforce by bringing together women working in HPC and technical computing and encouraging them to engage in outreach activities and improve the visibility of inspirational role models. The initiative has been recognised in the annual HPCwire Readers' and Editors' Choice Awards 2016, coming top in both the Readers' Choice Workforce Diversity Leadership Award (for the second successive year) and the Editors' Choice Workforce Diversity Leadership Award, with Director Toni Collis also gaining the Readers' Choice Outstanding Leadership in HPC Award.

Diversity in HPC aims to improve the participation of under-represented groups working in the High Performance Computing and supercomputing community, by encouraging participation by all and showcasing that HPC is a career path available to everyone. In addition to publishing Best Practice Guides aimed at improving diversity in the HPC community, the Faces of HPC feature highlights the careers of a wide variety of people, some famous and some less well-known. Many of them identify with one or more of the protected characteristics, while others reflect diversity in a different way, such as one who was the first member of his family to go to university.

To promote careers in HPC, suitable material must therefore be developed to showcase the opportunities available and the different paths that can lead to HPC. Emphasis must be placed on the fact that previous HPC experience is not required, and that diversity is welcomed, in order to prevent suitable candidates from thinking that HPC is not for them.

2.3 Better promoting the opportunities that are available

Once prospective HPC workers have decided that this might be a career option of interest, and have established that they have the necessary skills to get started, they need to know where to find job opportunities. Such jobs are still somewhat specialised and may not be advertised in the standard places. However, to identify all the potential employers and search their websites for suitable job vacancies would be time-consuming and frustrating, as most employers will only occasionally be seeking new staff.

To address this, EXDCI has created a Jobs Portal[16], which is a central place for HPC stakeholders from academia and industry to advertise their job vacancies. The portal also includes a list of links to other sites that advertise HPC jobs, and a link to the EXDCI Career Case Studies series. The need for a "European centralised HPC job advertisement site" was recognised at a BoF at the SC'16 conference, shortly before the EXDCI Jobs Portal was launched, so it is to be hoped that this proves to be a useful resource that helps to unite employers with suitably-skilled job-seekers.

[14] http://www.womeninhpc.org/

[15] http://www.hpc-diversity.ac.uk/

[16] https://exdci.eu/jobs-and-training/job-portal

2.4 Building a community

Champions-style initiatives are increasingly used to build networks of people who act as a local source of specialist knowledge within their own community. Examples include:

- **ARCHER Champions**[17]

  A network of local Champions in UK HPC centres and Higher Education Institutes who are able to provide local support for researchers in preparing proposals for ARCHER, the UK National Supercomputing Service. ARCHER Champions support activities to broaden the UK HPC user base to new disciplines and communities, and promote activities designed to provide career development to Research Software Engineers seeking a career in HPC.

- **XSEDE (Extreme Science and Engineering Discovery) Campus Champions**[18]

  A network of US campus representatives who act as a local source of knowledge about high-performance and high-throughput computing and other digital services, opportunities and resources. There are three types of specialised Champions: Student Champions, Regional Champions, and Domain Champions.

- **EDISON Education and Training Champions**[19]

  EDISON is a 2-year H2020 project aiming to accelerate the creation of the Data Science profession. EDISON Education and Training Champions are teams at universities, colleges or other teaching or training establishments engaged in developing and / or delivering the courses that Data Science professionals need.

The Champions approach brings together volunteers with a passion for the topic of interest (e.g. HPC, Data Science) and a commitment to sharing their knowledge to help their own community. The Champions receive initial training in order to become a local point of contact within their own community (this may be a regional community, e.g. the science faculty of their university, or a domain-based community, such as computational biologists). The initial training is followed up by ongoing support from the organising centre, including information packs and further workshops, either in person or online.

In the case of ARCHER Champions, there are two types of Champions that are particularly sought: staff at other HPC centres in the UK, who can help Tier-2 users progress to Tier-1 facilities, and early-career researchers interested in developing the skills needed for a career in HPC, who also gain valuable experience of networking and communication by taking on this role.

Clearly only a relatively small number of people can benefit from the opportunities offered by becoming a Champion, but such initiatives help expand the user community by encouraging users to progress to larger facilities, and by bringing in entirely new users and user communities. Anything that enlarges the community of HPC users consequently expands the pool of HPC-skilled people who are potentially available to fill HPC posts.

A similar initiative is run by the Software Sustainability Institute (SSI)[20] in the UK. SSI runs Software Carpentry and Data Carpentry workshops, which teach software development and data handling skills respectively, to enable researchers to be more productive and to make their research robust and reproducible. The demand for these workshops greatly exceeds SSI's

[17] http://www.archer.ac.uk/community/champions

[18] https://www.xsede.org/campus-champions

[19] http://edison-project.eu/edison/education-training-champions

[20] https://www.software.ac.uk/

capacity to deliver them, so in order to increase the pool of certified workshop instructors, the Institute runs a two-day intensive instructor training workshop. In addition to scheduled runs, the workshop can be run on demand for specific groups, and an on-line version also exists.

There is certainly an overhead in setting up a Champions-style initiative, and in keeping it running – training more people, and staying in touch with the existing Champions. However, once established, it is a relatively efficient way of increasing both the support that is available for existing HPC users, and the number of new users and potential future users of the systems. A greater number of such networks could be an excellent way to expand the user community and enhance communication within the community to make it a more supportive – and hence attractive – environment.

2.5 Integrating HPC training into undergraduate and postgraduate courses

To make a clearer path into HPC in the first place, a vital step forward would be to integrate HPC training into undergraduate and postgraduate courses. This would increase exposure to HPC and its possibilities to a far wider population at an earlier stage, and equip students with some basic skills which could lead them into careers in HPC. This issue is discussed further in D5.4, the HPC Training Roadmap.

One example of a successful collaboration between an HPC centre and a university is the incorporation of the BSC PATC[21] courses as accredited seminars in a compulsory module on the Barcelona School of Informatics (FIB) Master Program in Innovation and Research in Informatics[22].

An example of a highly-successful MSc programme dedicated to HPC, which also opens its courses to final-year undergraduates, is the 12-month full-time MSc in High Performance Computing, offered by EPCC and accredited by the University of Edinburgh[23]. In 2014, the MSc was extended to include a new specialism: the MSc in HPC with Data Science. Although students can take options from any relevant programme at the University, all the core content – comprising 12 courses, each worth 5 ECTS[24] – is taught by EPCC. The material from the MSc is used to refresh the courses run for the UK national supercomputer service ARCHER and by EPCC's PATC (and vice-versa). EPCC also runs two fully-accredited online courses (each equivalent to 10 ECTS): Practical Introduction to HPC, and Practical Introduction to Data Science[25].

Despite substantial fees being charged – around 12K euro for the MSc and 1.7K euro for each of the online courses – there were 26 MSc students in 2017 and 32 studying online courses. MSc graduates go on to a range of careers including industrial HPC, software development, and PhD study.

An additional benefit of running a focused Master's course in HPC is that it serves as a good recruiting ground, with students able to get a feel for the working environment and opportunities before applying for a job, and employers having a better chance than usual to evaluate the candidates prior to recruitment. For many MSc in HPC graduates, doing the course is a vital stepping stone in their career, allowing them to move on to work in HPC or continue on to PhDs

[21] PATC – PRACE Advanced Training Centre: http://www.prace-ri.eu/prace-advanced-training-centres/

[22] http://www.fib.upc.edu/en/masters/miri.html

[23] http://www.epcc.ed.ac.uk/msc/

[24] ECTS - European Credit Transfer and Accumulation System

[25] http://www.epcc.ed.ac.uk/online-courses/

in the field. Dissertation projects are very much related to real-world applications in HPC and challenges in the field, and industrial projects forge links between students and companies which may lead to job offers.

The NVIDIA GPU Educators Program[26] is another highly successful initiative. It provides teaching materials and GPU resources for use in university classrooms and labs, with the aim of equipping students with the skills in deep learning and accelerated computing that they are likely to need in the future. The flagship GPU Teaching Kits[27] are co-developed with academia for use in any academic discipline that benefits from accelerated computing. The three comprehensive packages (Deep Learning, Accelerated Computing, and Robotics) contain everything an instructor needs to teach a full-term curriculum course with GPUs, including lecture slides, lecture videos, hands-on labs and coding projects, as well as source code solutions. The global community of academics who use the Teaching Kits is supported by an online Teaching and Curriculum Forum. The Educators Program shows how industry can work with academia for mutual benefit, and offers a good example of how technology providers can help to foster in the next generation of workers the skills that are needed to exploit their technology.

2.6 Increasing the range of additional opportunities

2.6.1 *Summer schools*

For many current students, HPC is not a part of their undergraduate training, and committing to a Master's course is not always an option. Summer schools represent an excellent opportunity for students to make use of their long summer break to gain some HPC knowledge and / or experience in a more flexible way.

The PRACE Summer of HPC is a summer placement programme for senior undergraduates and early-stage postgraduates, with a one-week training course followed by 8 weeks spent at an HPC centre abroad, working individually on a specific HPC project. The placements are attractive to students of all backgrounds because they offer the chance to work on a real project, guided by a mentor, and do not require previous HPC experience; reasonable financial funding is provided, and living and working abroad for the summer offers much in the way of career and cultural enrichment – and hopefully is a fun experience too.

Outreach is a strong focus of the programme, and students are expected to keep up a lively presence on social media during their placement. They are required to write blog articles for the Summer of HPC website, and are encouraged to make use of sites such as Facebook and Twitter to broadcast their experiences. Final presentations are uploaded onto YouTube and are publicly available. The extent of the emphasis on outreach and dissemination is clear from the fact that two prizes are awarded each year, one of which is for the best HPC Ambassador.

The Summer of HPC has also been a positive step for advancing participants' careers in HPC, and several participants have gone on to do PhD studentships in their host centres, or to work in an HPC-related company. Being immersed in the field and being mentored by somebody who works in HPC is unquestionably a great way of promoting interest, building up some HPC related skills and looks great on their CV.

[26] https://developer.nvidia.com/educators
[27] https://developer.nvidia.com/teaching-kits

On a smaller scale, there are a variety of summer schools which offer one week of intensive HPC training, such as PUMPS[28], a well-established parallel programming summer school run jointly by the NVIDIA GPU Centre of Excellence at BSC-UPC and the BSC PATC. Other summer schools are run by HPC centres including Cineca[29] and CSC[30], and also by other institutes with a strong HPC focus, such as the HPC summer school run by Imperial College London[31]. PRACE runs a series of seasonal schools, though these usually do not take place in summer due to commitments with the Summer of HPC and International HPC Summer School[32], a summer school jointly organised by PRACE, XSEDE, RIKEN and Compute Canada.

Argonne National Laboratory[33] has also recognised that using supercomputers for computational science and engineering (CSE) requires expertise that is not always covered by formal education. The Argonne Training Program on Extreme-Scale Computing (ATPESC) was created to address this gap in professional training. ATPESC is a two-week programme for early-career researchers which focuses on the key skills, approaches, and tools needed to conduct CSE research on today's supercomputers and the extreme-scale systems of the future.

2.6.2 *Online learning*

Although there are only 65 places available for ATPESC each year, most of the lectures are openly available on the Argonne YouTube Training Channel[34]. Anyone can, in their own time, watch the 76 hours of lectures from some of the world's foremost experts and pioneers in extreme-scale computing.

Putting such content online makes it available for anyone who was interested in the programme but was not selected to participate, was unable to schedule the time for it, or found out about it too late to apply. Such an approach vastly increases the potential reach of the course, but if it is not advertised to a relevant audience, it will not fulfil its potential.

PRACE has developed two MOOCs, which will run for the first time from March 2017. One of these, "Supercomputing"[35], is a 5-week course designed for anyone interested in leading-edge computing technology, supercomputers, or the role that computer simulation takes in modern science and engineering. It does not assume any prior knowledge or computing programming ability. The course aims to explain what supercomputing is all about: how supercomputers are built, how they are programmed, and how they are used to advance our knowledge of the world around us through large-scale computer simulations. It also tries to explain the workings of a modern laptop or home PC: you may know that your machine has a quadcore processor, but what does this mean and why is it useful?

The MOOC will be offered via FutureLearn. FutureLearn courses are designed to be easily accessible, fun to study, and to promote the maximum possible interaction between learners. The course uses a combination of articles, videos, quizzes and discussion topics to explain the fundamental concepts of supercomputers and parallel computing. Although it is split into five weeks, learners have access to all the material from the outset so they can study at their own

[28] http://bcw.ac.upc.edu/PUMPS2017/
[29] https://eventi.cineca.it/en/hpc/26th-summer-school-parallel-computing
[30] http://www.prace-ri.eu/csc-summer-school-in-high-performance-computing-2017-jun-27-jul-5/
[31] http://www.imperial.ac.uk/computational-methods/news-and-events/hpc-2016/
[32] http://www.ihpcss.org/
[33] www.anl.gov
[34] http://extremecomputingtraining.anl.gov/2016-videos
[35] https://www.futurelearn.com/courses/supercomputing

pace. The course runs for the first time on 6th March 2017; as of 9th February it had 1200 registrations.

Anything which increases the reach of training material in an accessible and affordable fashion can only boost the numbers of the HPC-ready workforce of the future. By offering more summer programmes to fit in with the availability of enthusiastic students, and making more material available freely available online, so that people can study in their own time, more people have the opportunity to gain their first experience of HPC – or deepen their existing knowledge.

2.7 Promoting computing as an essential skill to university and school-age students

Integrating HPC training into undergraduate and postgraduate degrees, providing more summer schools and increasing opportunities for online training in HPC all go to the heart of tackling the lack of HPC readiness in the future workforce. But a commonly-held view among HPC professions is that strong software development skills are every bit as important as HPC skills. There is an increasing view that all researchers must be exposed to best practices in software development, and that the fundamentals of good software engineering should form part of every researcher's basic training. Giving researchers a sound grounding in software engineering practices could create more competent programmers who could more easily make the leap into HPC.

There is also a strong push to increase general computing skills in school-age children – computing is sometimes referred to as the new literacy. Every initiative to increase the younger generation's competence and confidence in programming can only have a positive effect on the future HPC workforce.

Many new initiatives to broaden participation in computing are focused not only on learners, but also on educators and policy-makers, recognising the multiple directions in which efforts must be directed. These often include a "train the trainers" aspect, providing training and resources for teachers to then deliver to their classes, widening the impact of the material.

One such American initiative, The Beauty and Joy of Computing[36], is a non-majors university course which is also offered in many high schools across the USA. It includes HPC concepts such as concurrency and distributed computing. It has also now been redeveloped as an edX MOOC, and there is an associated six-week professional development workshop which trains high school teachers to teach the course to their pupils.

Code.org[37] is a non-profit organisation dedicated to expanding access to computer science, and increasing participation by women and underrepresented minorities. It has expanded efforts to include the development of high quality curricula and classroom tools, ongoing professional learning for teachers, and outreach efforts to school districts and the CS education community.

CS Unplugged[38] ("Computer Science without a computer") is a collection of free learning activities that teach computer science concepts through games and puzzles that use cards, string, crayons and lots of running around. The activities are suitable for children of all ages and introduce them to computational thinking through concepts such as binary numbers, algorithms

[36] http://bjc.berkeley.edu/; https://www.youtube.com/watch?v=eqbgqSe_hvU

[37] https://code.org/

[38] http://csunplugged.org/

and data compression, separated from the distractions and technical details of having to use computers – no programming is involved.

Although teaching general computing skills to school age children and undergraduates may not fall within the remit of HPC centres and technology providers, it is clearly something which should be supported whenever opportunities arise. Ensuring that tomorrow's workforce is comfortable with and competent in the use of computers is an essential foundation for equipping that workforce with the necessary HPC skills to meet demand.

3 Retaining talent within the HPC workforce

The lack of suitably qualified people in the HPC workforce may be caused largely by a lack of awareness of the available opportunities, and insufficient preparation through school and university studies to equip the new generation of workers with the appropriate skills. However, the problem is exacerbated by the failure to retain talented people with the correct skills within the sector.

An example of an action taken not only to capture but also to retain talent is the Professional Development Programme (PDP) at BSC, which provides each researcher with a personalised development plan, identifying annual objectives and an individual training plan.

The BSC Extra Diploma in Research Skills is targeted at young researchers and is related to the BSC HRS4R[39] (Human Resources Strategy for Researchers) award. It is influenced by the specific needs of the young researchers themselves, as well as by the need within the EU HPC ecosystem for well-trained specialists. The Diploma is the focal point for training and career development actions in different areas of competence, including technical, personal and leadership skills, as well as skills relevant to the industrial sector.

Another BSC initiative is the BSC Folks[40] project, which acts as a meeting point for researchers who have studied and worked at BSC. At any given time, the institution has more than 100 PhD students and a number of early-stage post-docs from across the world in its research departments, so BSC Folks provides this fluid body of HPC specialists with opportunities to stay in touch, participate in new projects and share information about training and other initiatives.

Although examples of good practice exist, the initiatives are few and not sufficiently far reaching. While jobs within the commercial sector are generally better remunerated and benefit from clear career structures, with opportunities for promotion and specialisation, many HPC-related jobs are in the academic sphere, and here people can fall uncomfortably between the roles of academic researcher and computing support staff, with little prospect of career progression, and a lack of job security due to being tied to fixed-term research grants.

Two current initiatives, described below, are examining how to establish a career path in order to retain skilled workers in job roles that have not previously been well defined. While these look at the jobs of Research Software Engineers and Data Scientists, the general principles also apply to HPC.

[39] https://euraxess.ec.europa.eu/jobs/hrs4r

[40] https://www.bsc.es/news/bsc-news/lets-stay-connected-bsc-folks

3.1 Defining the role: Research Software Engineer

The Software Sustainability Institute (SSI)[41] is driving a movement to recognise and define the role of Research Software Engineers (RSEs). The need to define this role arose out of the recognition that most research is now powered by software, and a growing number of people in academia combine expertise in programming with an intricate understanding of research – but that these people lack a formal place in the academic system. Researchers are generally judged on their research output, but there is no formal way of measuring software output within this framework, and there is no reward structure for those who spend most of their time producing code, who are also often simply acknowledged on publications, instead of being credited as a named author.

RSEs combine a professional attitude to the exercise of software engineering with a deep understanding of research topics. Understanding research literature is a major part of the life of the research software engineer. They lead the design and construction of increasingly complex research software systems, and play an important part in the co-design of research requirements, understanding and addressing software engineering questions that arise in research planning. They also need to develop pedagogic skills to teach and train colleagues, and may be required to write papers and grant proposals, as well as technical reports and manuals.

Having established the role of Research Software Engineers and founded a strong community (UKRSE Association[42]), SSI is now focusing on ensuring that career paths are put in place, to stop the outflow of talented postdocs to industry, or to entirely different careers. RSEs cannot be fairly compared to academics, as they do not produce many publications, but nor can they be compared to lab technicians, as they need to keep abreast of the cutting edge of scientific methodologies. By recognising the contribution of RSEs, the hope is that there will be a greater willingness to share code and best practice in software development for the good of research.

The growing emphasis on sustainability in research is a problem for software development in the academic environment. Development focuses on the short-term goals of specific research grants, and the software often dies with the grant. The majority of software developed in academia never leaves the research group or gets released in a formal manner. Making scientific software more open, reproducible, and sustainable needs to be recognised as a separate endeavour from the original scientific publication. If other research projects can be seen as future stakeholders, the goal then becomes to produce reusable software that can last beyond the scope of any one project.

Few research groups are large enough to support a dedicated full-time Research Software Engineer, but nearly all research groups require help from one. To solve this problem, a new model was proposed: the Research Software Group. This pools RSEs into a single group and allows researchers to hire them only when they are needed. It provides a flexible and cost-effective service for researchers, and it focuses the huge, but fragmented, demand for Research Software Engineers, allowing the Research Software Group to sustain long-term careers. It also allows the group to grow to a size where a hierarchy is required, thus providing a path for promotion.

This model has now been put into practice successfully at University College London, where a group of RSEs works closely with researchers around the university to help them develop better

[41] www.software.ac.uk

[42] www.rse.ac.uk, @ResearchSoftEng

software, through projects where they effectively become a temporary member of the research group. They also deliver training for PhD students and postdoctoral researchers.

3.2 Defining the role: Data Scientist

Data Science is an emerging field that requires multi-disciplinary knowledge, with strong links to Big Data. It is having an impact on nearly every aspect of how research is conducted, how scientists think, and how research data are used and shared. There is an urgent demand now for Data Scientists with the right skills and competencies to efficiently adopt the big data approach in their daily work, and the gap between the number of vacancies and the number of qualified people is expected to grow rapidly.

The EDISON Data Science Framework[43] is an H2020 project tasked with identifying the skills and competences required in order to establish the Data Scientist as a profession. It will speed up the increase in the number of competent and qualified Data Scientists across Europe and beyond. This will be achieved by aligning industry needs with available career paths, and supporting academies in reviewing their curricula with respect to expected profiles, required expertise and professional certification.

EDISON is developing a reference Data Science Model Curriculum, a draft version of which is now available under a Creative Commons licence[44].

[43] http://edison-project.eu/edison

[44] http://edison-project.eu/data-science-model-curriculum-mc-ds

4 Conclusions

The shortage of HPC-skilled people in the workforce is preventing the full uptake of HPC in academia and industry, threatening Europe's competitiveness. The reasons for this shortage are two-fold: firstly, too few people enter the workforce with the necessary HPC skills, and secondly, not enough is done to ensure staff retention. There is no single solution to address these issues, but rather several different approaches must be taken.

At a very basic level, public engagement activities should be given more attention, in order to create a general awareness among the wider population of HPC and its impact, how it is relevant to every one of us, and the sort of job opportunities it presents. Some HPC centres have put a strong emphasis on public engagement and have had great success connecting with the public using attention-grabbing displays and fun demos. Developing suitable material requires creative thinking, which some staff find rewarding as it is something totally different. While such outreach activities are likely to have only a low impact and over the long term, they may reach people who might otherwise not have come into contact with HPC. The more people know about HPC, the more mainstream it becomes. All HPC stakeholders should have a public engagement strategy to raise the profile of HPC beyond the existing community.

For those people who are already aware of HPC and might consider it as a career, there needs to be clear information on how to move into the area if they are not already in it, about the sorts of jobs that are available, and about the type of people who work in HPC, to ensure that nobody feels that they are not the right sort of person. Developing and publicising career case studies, along with the work of initiatives such as Women in HPC and Diversity in HPC, can help to showcase what working in HPC can offer and present it in an exciting way.

Current job openings can be difficult to find; to this end, the EXDCI Jobs Portal collates job offers from academia and industry and makes them available in one place, along with a list of links to other job sites and other resources. This makes it easier for job-seekers to find jobs, but also benefits the recruiters by making vacancies more visible.

Creating a Champions-style community is a more indirect way of promoting HPC as a career. One specific role of Champions is to promote activities which provide career development to research software engineers seeking a career in HPC, but other objectives are to assist users to move from Tier-2 to Tier-1 facilities, and to engage with new communities. By expanding the use of HPC, it naturally follows that the HPC skill-set in the workforce also grows.

Increasing learning opportunities at undergraduate and postgraduate, and even school, level is another way to create more of an HPC skills base and therefore increase the chance of being able to fill future posts with suitable candidates. This can be done by integrating HPC courses into undergraduate and postgraduate courses, as well as through summer schools and placements, and the reach of training can be greatly expanded by making online recordings available and through new online learning methods including MOOCs.

Most of the work to address the shortfall of HPC skills in the workforce lies in raising awareness of HPC, job opportunities in general, and currently open posts, as well as in equipping the future generation with solid computing skills from an early age, and with additional HPC skills from undergraduate level. But to address the issue of retention of staff with suitable skills, lessons should be learned from the establishment of the Research Software Engineer role by SSI, and the EDISON effort to define the role of, and model curriculum for, the Data Scientist. It may be time to clearly define an HPC specialist role – perhaps as a category of Research Software Engineer – and develop a career path with opportunities for recognition and reward, in order to

make HPC as a career an attractive option not only for those choosing to enter it, but for those who are already working within it.