Report on the Consultation Workshop

Skills and Human Resources for e-Infrastructures within Horizon 2020

30 May 2012

European Commission
DG Information Society and Media
Unit F3 'Géant & e-Infrastructures'
Brussels, Belgium

This workshop was organised by the European Commission, Information Society and Media Directorate-General (Kirsti Ala-Mutka of Unit F3).

The report was produced with the assistance of the workshop rapporteur, Niobe Haitas.

Table of Contents

Executive Summary

European Commission DG INFSO e-Infrastructure activity aims to support Horizon2020 objectives by working towards a single and open digital European research area where researchers can access the expertise, resources and instruments they need online. In addition to technical challenges, this requires considering the development of necessary human capital base for e-Infrastructures development, service provision and efficient scientific usage. The skills and human resources aspect has been recognised as an important part of Research Infrastructures programme under Horizon 2020. This report summarises the outcomes of a consultation workshop that was organised by DG INFSO "GÉANT and e-Infrastructures" unit to consult the stakeholders on their views of approaching these challenges.

The workshop discussions highlighted cross-cutting challenges of i) new and changed skills needs which combine technical and scientific skills and require interdisciplinary thinking and communication; ii) recognizing new job profiles and tasks rising from the emergence of computing intensive and data-driven science with integral role of e-infrastructures; iii) need for effective European level collaboration and coordination to avoid duplication of efforts and join the forces for developing high quality human capital for e-infrastructures. Furthermore, there are specific challenges relating to the skills and human resources for i) e-infrastructures development, ii) digital research service provision, iii) scientific usage of e-infrastructures, and iv) the institutional strategies for effectively tackling the human resources challenges.

During the consultation, several suggestions for promoting skills and human resources for e-infrastructures were raised. It is necessary to better map the current situation and future needs, and support recognizing and establishing new job profiles and career paths. Specific investment should be addressed to innovating new education and training approaches: creating new specific curricula for e-infrastructures personnel; integrating technical and scientific aspects into cross-disciplinary curricula; encouraging collaborative mindset and computational thinking into education already from early age; on-the-job learning through knowledge transfer from and between experts in terms of short-term courses and staff exchanges. These should be supplemented by encouraging 'communities of practice' between e-infrastructures actors, including developers, operators and scientific users. Training development should be supported by developing accreditation mechanism for certifying high-quality courses, and by implementing collaborative repositories for best practices, training materials and supporting e-science and e-infrastructure tools and approaches. These require institutional strategies, which should be aligned with European level strategies developed in multi-stakeholder collaboration, including participants both from academia and industry.

Several concrete recommendations for supporting the suggested development aspects with e-infrastructures activities under Horizon 2020 were devised. It was considered important to have both specific and integrated activities to support skills and human resources aspects within the e-infrastructures projects. The suggested measures include several concrete topics for project calls, specific studies and support activities, conditions and elements to be integrated in project selection and reviews, and overall leadership and advocacy activities to be implemented by the EC on this field, in collaboration with stakeholders.

1 Introduction

Horizon 2020 priority on Excellent Science aims to support the best ideas, develop talent within Europe, provide researchers with access to priority research infrastructure, and make Europe an attractive location for the world's best researchers. This requires ensuring that Europe has world-class *research infrastructures* (including e-Infrastructures) accessible to all researchers in Europe and beyond and fully exploiting their potential for scientific advance and innovation.

The aim of e-Infrastructures activity is to achieve by 2020 a single and open digital European research area where researchers can access the expertise, resources and instruments they need online. Every researcher should become digital, know how to benefit from technologies for scientific purposes, use relevant tools for tackling grand challenges of today through computing and data-driven research approaches, and benefit from worldwide connections and collaborations. However, making this a reality for all researchers relies on access and availability of suitable e-Infrastructures as well as on their skills to deploy e-Science approaches. This requires considering the development of necessary human capital base for e-Infrastructures development, service provision and efficient scientific usage.

1.1 Workshop objectives

In view of the above, the European Commission e-Infrastructures unit is organising in spring 2012 consultation workshops with key stakeholders in order to investigate their needs, expectations and prospects and integrate these into the Horizon 2020 (H2020) work programmes. In the series of consultations, the event on 30th May concentrated on the skills and human resources aspects of e-Infrastructures. The workshop aimed to recognise and elaborate the needs for **training and human resources for effective e-Infrastructures development, provision and deployment** for European excellence in science, and to discuss how e-Infrastructures activities under H2020 could contribute to addressing these needs. In particular, the goals of the workshop were:

- To **bring together key actors** to discuss skills and human resources aspects for infrastructures in order to advance their development and contribution to excellence in science
- To study the **state-of-art and development needs** for skills and human resources in order to promote effective e-Infrastructures development, service provision and deployment
- To **propose concrete recommendations for e-Infrastructures activity under Horizon2020**, both as a programme wide strategy and specific actions to be addressed in the work programmes to be launched.

The workshop was organised to approach the skills and human resources aspects from three main perspectives, as illustrated in Figure 1, with the objective to cover aspects of:

- Human resources for e-Infrastructures operation and maintenance
- ICT and data skills in short supply for e-Infrastructures development
- ICT and computational skills for scientific communities and industry.

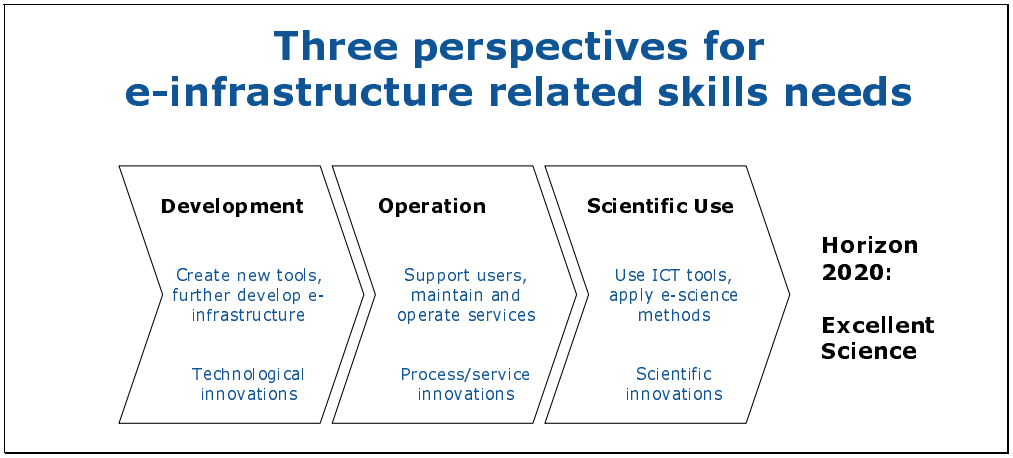

Figure 1: Starting perspectives for the workshop discussions on e-Infrastructure related skills needs.

1.2 Reporting approach

The consultation was implemented by inviting individual presentations and by facilitating joint interaction and discussion. The consultation participants are listed in Annex, and the contributions of all of them are very much appreciated. Individual presentations can be consulted online at http://cordis.europa.eu/fp7/ict/e-Infrastructure/home_en.html.

This report aims to summarise the outcomes and resulting recommendations from the consultation in a synthetic and concise manner. In particular, during the discussions it was realised that the institutional change aspect was an important perspective as such, and was therefore added as a separate dimension in this report. Furthermore, several cross-thematic issues were recognised. Therefore, the resulting report does not follow the structure of the sessions or presentations, but aims to present the synthetic outputs in a logical manner. The report is structured to first highlight overall common challenges (Section 2), discuss the specificities on different areas (Section 3) , describe desired developments and solutions (Section 4) and finally list concrete recommendations for actions to be taken by the European Commission and to be supported under e-Infrastructures activities in Horizon 2020 (Section 5).

2 Cross-cutting challenges for e-Infrastructures human capital

The overarching theme of the consultation was the need to improve and develop education and training approaches to better support relevant skills and jobs relating to e-Infrastructures development, service provision and scientific usage. Especially given the 'data deluge' challenge and the growing needs for data-driven research approaches, university education and short-term training schemes are often out-paced by technology. The main challenges lie in developing lifelong, training approaches that respond to continuously changing skills requirements, in developing human resources policies which take into account the new roles and jobs needed in institutions, and in ensuring and supporting European level development through efficient coordination of efforts and sharing of results.

2.1 Addressing new skills needs

As new scientific methodologies emerge and are increasingly taken up, the gaps in skills and in numbers of skilled people are becoming more and more evident, as is the need to create new roles. It is estimated for instance that by 2019, there will be a shortage of 190,000 data scientists in the US (McKinsey Global Institute, 2011). Scientific data management has so far often been carried out by librarians, however, they no longer have the necessary skills to respond to the emerging needs of managing big and more complex data created by dynamic research workflows. Even fewer have the necessary specialist scientific knowledge to interact meaningfully with data themselves or even to provide support to researchers to organize their data and research work. Figure 2 shows data discussed by Liz Lyon, with areas recognised as having potentially significant skills gaps, since these skills needs are already seen essential by many, and their importance is seen to grow in 2-5 years.

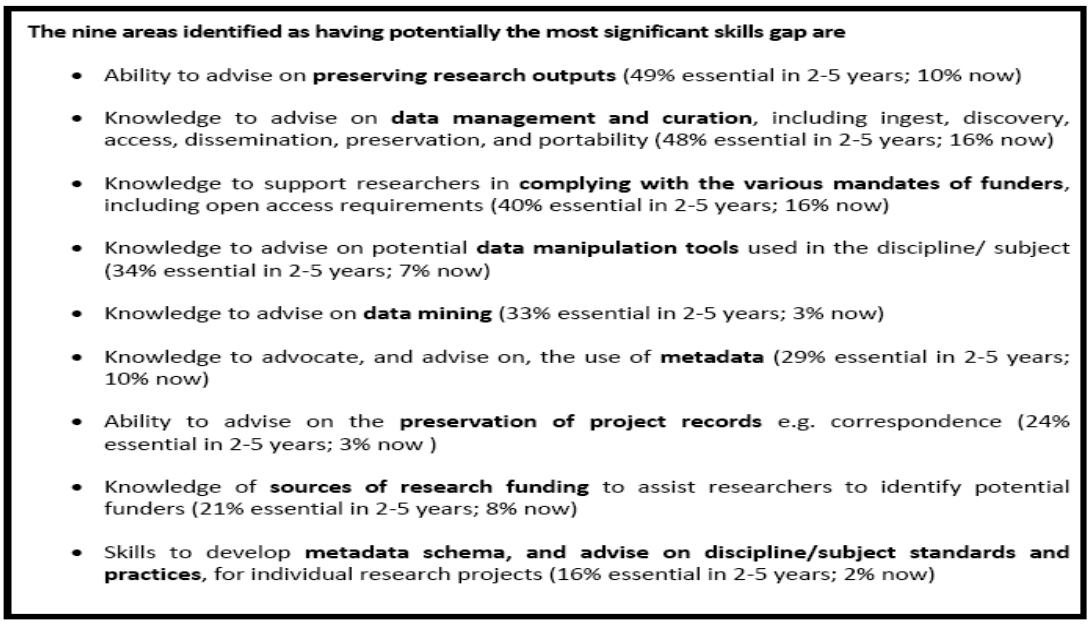

Figure 2: Areas with potential skills gaps from Auckland (2012).

Bert van Pinxteren highlighted that a specific issue in skills and human resources for e-infrastructures is the required high specialisation in a dynamic very fast moving area where knowledge outdates quickly. Furthermore, many of the new skills relating to e-Infrastructures are context-dependent, however their common denominator is the need for cross-disciplinary understanding and communication ability. Developers, operators and scientific users need some level of understanding of each others' roles, in order to work together towards excellent scientific results enabled by high-quality e-Infrastructure services. Developers need to be embedded in the research process in a regular manner and similarly, scientists need to be embedded in the development process, from the definition of requirements to the final user testing. To optimise efficiency at all levels and continuously update and improve services, regular interaction is also necessary between developers and the staff who provide services and maintenance support.

2.2 Acknowledging new jobs and tasks

Everyday research practices are continuously changing due to take-up of computational and data-driven science approaches. Jobs and roles at the institutional level are therefore affected and challenged, with new necessary activities and completely new job profiles emerging. However, such changes in human resources and job profiles requirements are not necessarily recognised by the institutions. Even when specific needs are acknowledged, they may only be considered as project-based short-term tasks, instead of permanent change needs. Skills for e-Infrastructures are often developed through specific courses and on-the-job training, which can mean that by the time an employee reaches high specific skill levels, their contract terminates. This may result in lack of motivation in developing one's skills and job performance and higher and more frequent training investments by the institutions that do not benefit from the cumulative skills development of their staff.

More and more sources exist to generate bigger and bigger volumes of data and more skills are required to manage data efficiently, including to use, reuse and generate new knowledge out of them Iryna Kuchma suggested that there is currently high heterogeneity of data management maturity in Europe – in some countries the libraries are paving the way, but some countries have none or very few activities. Figure 3 presents data from Liz Lyon's presentation, comparing differences in time spent on 'big data' instead of 'normal data'. Required increase in work effort should be taken into account when planning work and defining respective job profiles. Additional

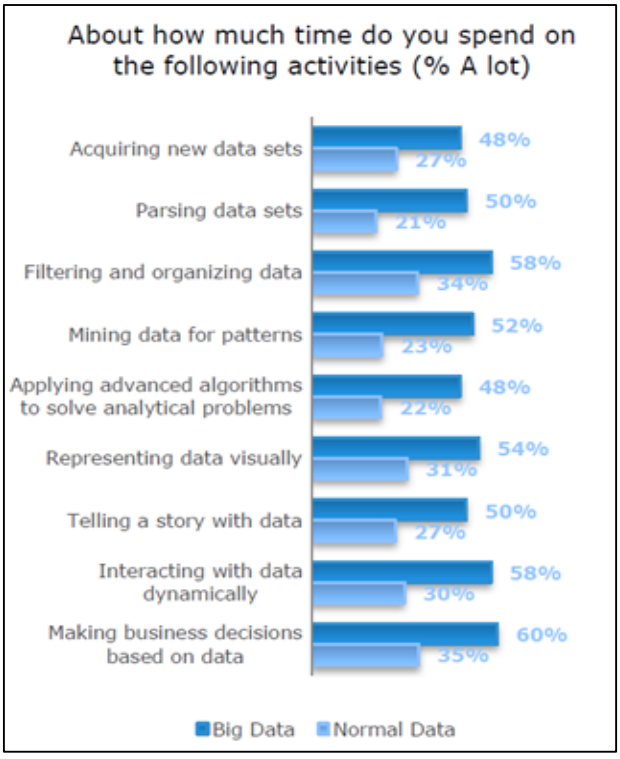

Figure 3 : Big Data effects on research tasks (EMC 2011).

roles that emerge to support e-Infrastructures include interaction and engagement with the public, working with social media and networking tools, as well as teaching and sharing expertise. Figure 4 extracted from the presentation of Nuket Yetis summarises the range of technical and non-technical skills, as well as the personal qualities needed to respond to the work requirements of e-Infrastructures by 2020.

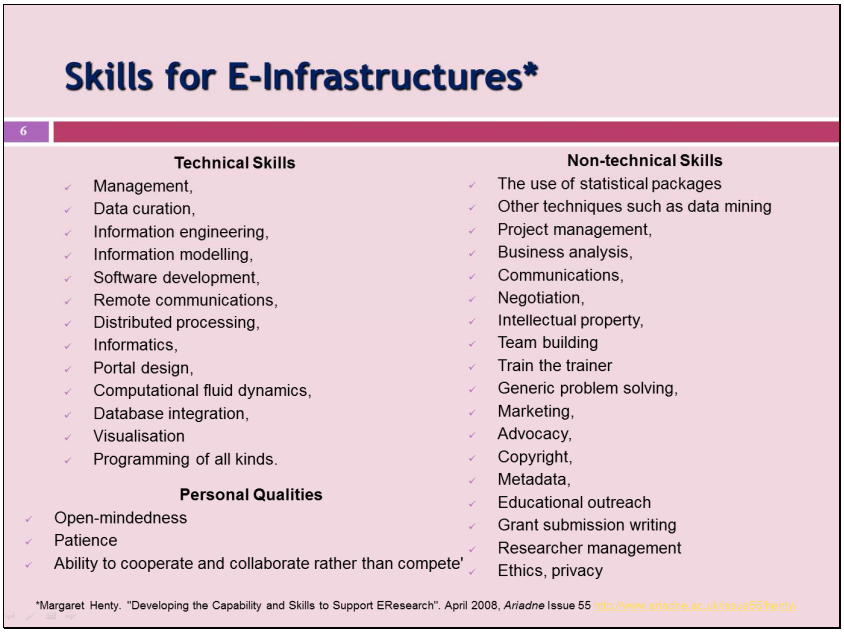

Figure 4: Technical and non-technical skills for e-Infrastructures, based on Henty (2008).

2.3 Effective collaboration and coordination

Throughout the workshop, various aspects relating to dispersed developments and lack of coordination and harmonization of efforts were emphasized. Ineffective knowledge transfer between existing e-Infrastructures initiatives and projects can lead to unnecessary duplication of efforts and waste of resources, 'reinventing the wheel'. This applies both to projects focusing on e-Infrastructures development and provision and those providing specific training approaches.

Various training efforts relating to e-Infrastructures development, operation and scientific usage do exist, and some of them are listed in the Annex of this report. However, the workshop participants considered that university curricula, short-term courses and online trainings are very diverse and often developed in isolation to meet the needs of specific projects. In this dynamically developing area there are no commonly agreed degrees, training contents or even terminologies to enable common recognition of given training. Some training programmes do not provide certificates useful for career development which could be used locally or for facilitating mobility of trained people between institutions or countries.

There are currently no effective means to share best practices and training related resources between e-Infrastructures projects and institutions at the European or even the national levels.

Roberto Barbera pointed out this as an important barrier for personnel based at a decentralized, local support node without resources to build their own regularly updated training modules to keep up with fast moving new knowledge. Similarly, some countries do not have sufficient expertise or funding to provide cutting-edge knowledge resorting to less than optimal training provided by volunteers without specialized engineering or teaching skills.

Huge heterogeneity in job profiles, roles and responsibilities exist not only across countries in Europe but even across institutions in the same country. For example, although the 'data scientist' job profile is increasingly recognised as important, there is no common definition. Liz Lyon gave as an example 0 job titles found to describe related roles, including Science Data Librarian, Data Management Librarian, Research Data and Digital Curation Officer, Data Curation Librarian, Data Officer, Research Data Management etc. Such heterogeneity also applies to career structures across institutions in Europe, and to approaches and incentives for professional development.

3 Specific aspects to e-Infrastructures skills and human resources

3.1 e-Infrastructures development

Context dependent needs for specific skills

Several specific skills required for e-Infrastructures development were mentioned during the consultation. The discussion highlighted that e-Infrastructures should be developed to address both grand challenges communities and increasingly various long-tail communities with specific needs. This implies that e-infrastructures developer skills needs vary from generic software skills to domain-specific application developments. The Figure 5 from the presentation of Krzysztof Kurowski illustrates the areas for software engineering skills. Andrew Smith suggested that most difficult profiles to find include software developers (java, database developers), software engineers, data scientists/curators and developers of parallel software code. Gabriella Cattaneo presented IDC study results showing needs for parallel software skills, and proposed new specific skills needs also rising from the takeup of clouds-based approaches. Participants also discussed the need to improve experts' teaching skills to be able to teach other e-Infrastructures experts.

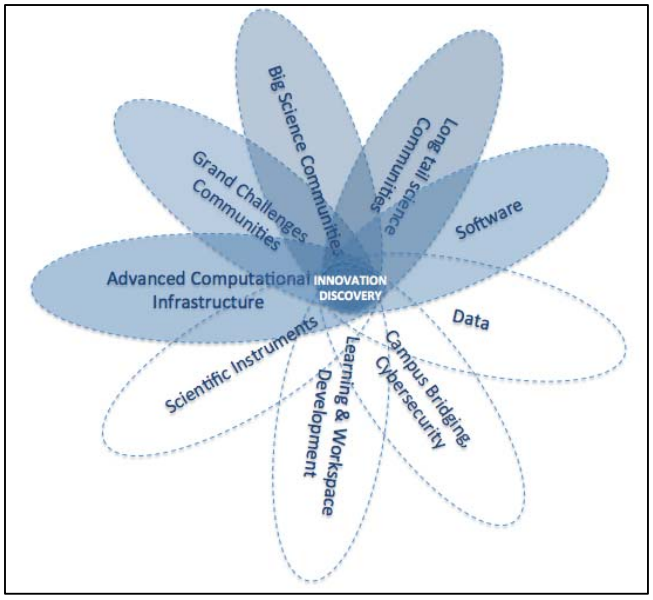

Figure 5 : Aspects of software engineering skills needs. NSF Cyberinfrastructure Framework for the 21st Century (CIF21), extended by two components representing Big Science and Long Tail Science Communities (by Krzysztof Kurowski)

Skills and practice to communicate with scientific users

The participants agreed that usability, ease of use and also quality and security of e-Infrastructures must be improved, together with customized and standardized interfaces, configurations and services to simplify researchers' access to them. Optimally, software developers' skills should be complemented with 'interface developers' and also with 'scientific awareness' skills to create an understanding of the research fields for which the e-Infrastructures are being developed. There is "lack of biology-aware computer scientists and of informatics-aware biologists", as put by Andrew Smith. This is a big challenge when research is becoming more and more interdisciplinary, cross-cutting and collaborative. The lack of understanding, communication and collaborative work between IT professionals and scientific users can limit the process followed to design and develop research e-Infrastructures to a mere collection of users' requirements rather than following a dynamic interactive design model based on communication and user feedback. Krzysztof Kurowski suggested that to prevent building e-Infrastructures that fail to meet users' needs and expectations to solve complex problems, potential end-users need to be involved in the design and deployment phases but also in the verification processes that follow, so as to improve future services.

3.2 e-Infrastructures service provision

Skills for understanding service context

Challenges for e-Infrastructures service provision can relate to insufficient skills in technical aspects such as maintaining and supporting software, providing user support, data processing services and products, but are also related to a lack of general understanding of users' research practices or the bigger picture of e-Infrastructures. Focusing only on specific aspects of e-Infrastructure operation and maintenance without broader view can result in not being able to proactively address risks and improve service provision. It is important to ensure skills and human resources for making user support available on a permanent basis as well as being able to provide customized user trainings (e.g. beginners, intermediate, advanced). Gudmund Host also suggested that green sustainability policies for e-infrastructures should be included among the new issues to be considered in training.

Low appreciation of the job profile

There is a general lack of recognition of the contribution of e-Infrastructures operators to the overall field of research; as opposed to researchers and even developers, service provision staff rarely get credited, recognized, cited in publications or participate in conferences and trainings. e-Infrastructure service provision does not always get sufficient attention or investment in the scientific organizations. This may result in unmotivated employees failing to improve services from a 'business as usual' mode to excellence and posing a challenge for long-term operation, maintenance and user support of e-Infrastructures. The unclear valorisation of this job and recognition of its contribution to research output is also reflected in the unclear career path development for professionals in the area of e-Infrastructures services.

3.3 Scientific usage of e-Infrastructures

Skills to benefit from the potential of e-Infrastructures

Pekka Manninen categorised the skills needs for researchers in academia and industry into domain-specific, general methodologies and transferable skills (Figure 6), all of which have links to e-infrastructures. Big gaps in researchers' skills in using e-Infrastructures currently exist in specific aspects such as high performance computing, cloud-supported research and in silico simulations. Andrew Smith highlighted that scientists may also feel under-qualified to make the most of publicly available data. Systematic and standardized training for using e-science methods and tools is typically not available. This is further enhanced by the fluidity of the technical landscape which requires that scientists need to invest time and effort during their whole career to keep themselves updated with new developments. However, it was highlighted in the discussions that although technical skills are currently needed by researchers to be able to efficiently use e-science approaches, more important is their capability for computational thinking and understanding the potential of e-Infrastructures for their specific field of research. Roberto Barbera introduced Science Gateways as a user-friendly approach that put users at the centre and in control of e-infrastructures, highlighting the need to develop new dissemination and training formats to progress with e-science takeup.

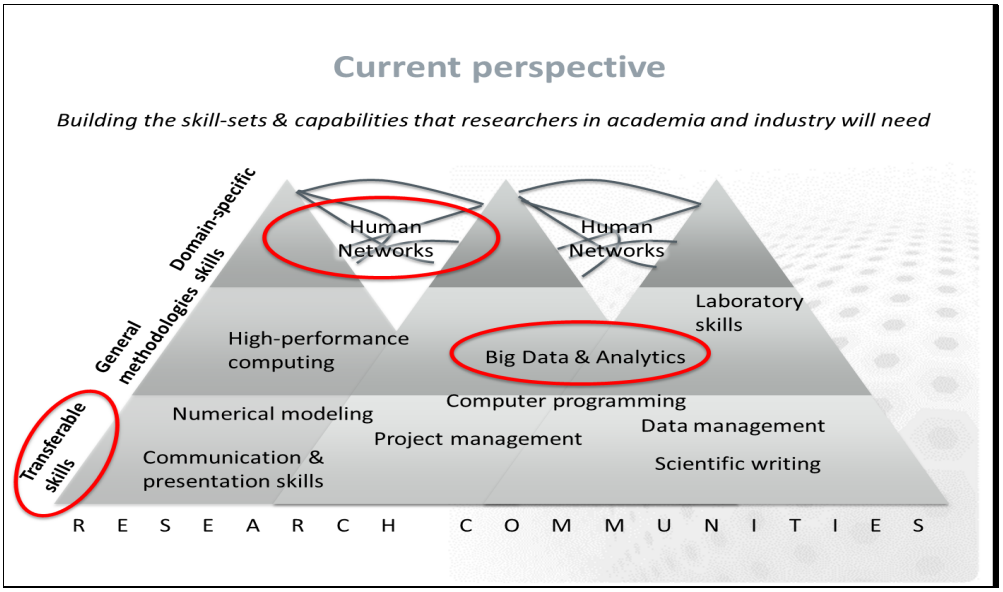

Figure 6: Skill-sets and capabilities required for the industry and the academia by 2020 (Pekka Manninen).

Embracing open e-science culture

The share of e-Infrastructures' users is currently relatively small, especially among the social sciences communities, due not only to the technical challenges posed by e-Infrastructures but also to a general lack of awareness of e-Infrastructures capabilities to transform and benefit everyday research practice. Furthermore, many of the e-Infrastructures aim at supporting collaborative research approaches which rely on opening and sharing one's research data and participating in virtual collaborations. This requires a cultural change for scientific disciplines and research practices where only final results have been published in journals, without openness during the research process. The lack of exposure of scientists to e-Science practices often leads to misperceptions and prejudice, perceptions about their technical difficulties, fears about security and cyber-attacks, concerns about IPR issues and future use of data by others, further hindering their uptake, as discussed by Nuket Yetis.

3.4 Changes in research institutions

Need for institutional leadership

Big data science and related e-Infrastructures are bringing forward several issues requiring action on the level of research institutions. Institutions need to transform organizational practices and adapt themselves to respond to the emerging needs of the e-science context and optimally manage time and resources for all staff. New questions have emerged regarding research data, including access, use and reuse, storage, sharing, ethical issues, IPR, annotations and metadata. Many institutions follow their usual decision-making mechanisms which may result into sometimes implicit, sometimes explicit policies, differing between different institutions and even between departments in the same institution. This can result into uncoordinated, uneven progress and non-interoperable solutions, thereby hindering national and European-wide development in e-Science. Participants agreed that

there is a need to coordinate existing policies at the institutional levels, and align these to the national and European levels.

Need to support professional development

Institutions need to pay specific attention to support the e-Infrastructures-related actors in view of the changes in the scientific landscape and create long-term perspectives for jobs and professional development. Many institutions measure scientific excellence through traditional peer-reviewed publications as the main indicator considered for career development. However, e-Infrastructures developers and operators do not necessarily participate in publications despite their very important contributions to the scientific output. Workshop discussions highlighted that incentive systems should take into account different types of job contributions and trainings and should enable appreciation and meaningful career development for all actors, not only traditional researchers. As suggested by Peter WIttenburg, also administrative regulations and procedures can also pose challenges in Human Resources management for e-Infrastructures. Institutional, national or funding organisation rules for contract durations, recruitment procedures, and renewal of contracts, salaries and number of reportable working hours per day that can be claimed for costs can challenge the emerged needs to organize human resources in the most efficient manner in order to be competitive.

4 Suggested developments

Better understanding of the problem and current status

The main difficulties in supporting skills and human resources for e-Infrastructures relate to the emergence of a data-driven science paradigm and the recently increased importance and contribution to science of developers and service providers. The overall landscape of the problem and scale of supply and demand of the human resources needs is not yet fully clear. Mapping of roles, job titles and respective tasks with the aim to reach common definitions of job profiles was suggested as a means to draw a picture of the current and future needs of human resources relating to e-Infrastructures. This should include mapping the number of current professionals but also future scenarios for increased demand. Gudmund Host suggested that clarifying roles and responsibilities and harmonizing terminology through common job descriptions with required skills will help improve recognition of new professions emerging in their own right, such as the ones of data scientists and data librarians. Establishing common job descriptions will also support developing respective education and training approaches and harmonise human resources policies at the institutional, national and European levels. This mapping and forecasting effort should consider respective efforts and European level frameworks on ICT industry, such as the European e-competence framework, in order to observe synergies and compatibilities.

Recognition of new jobs and careers

e-infrastructures development and service provision effort needs to be better recognized and valued for their contribution to the modern science. New job profiles need to be created and their importance should be made visible through permanent positions, recognition in institutions and open science merits. These novelties will need to be conceptualized, designed, and new salary scales and career paths will need to be envisioned and respective costs invested. In research communities, these actors should be recognized as part of contributing team members, career paths with envisioned promotions and personal evolution should be developed and general public awareness should be improved. Becoming recognized as a contributor and part of a scientific team will help boost motivation to develop in the profession, and attract new people and students to the jobs. This should be supported through top-down actions and at the same time encouraged through bottom-up means which enable current and future developers, operators and users from different scientific fields to interact, learn and innovate with each other, with an open and collaborative mindset. Importance of these new jobs and roles has been highlighted e.g. in the UK , as pointed out by Torsten Reimer (Figure 7).

Figure 7: Importance of new job profiles highlighted in the UK e-Infrastructure vision (Torsten Reimer)

Supporting Education and Training development

It was strongly highlighted that more funding and effort should be allocated for innovating existing training schemes and developing new ones. Gudmund Host suggested also that specific investments should be targeted to doctoral and post-doctoral training of future leaders who will take forward the development of ERA's capacity in this field. e-Competencies must go beyond ICT skills and trainings need to respond to this trend, stimulating the demand for e-Infrastructures. All actors in the entire chain of e-Infrastructures should acquire a certain level of understanding about technology, infrastructure and research processes across disciplines to be able to identify innovation opportunities. Cross-disciplinary curricula and training modules combining technological and scientific expertise should be encouraged -- E&T strategies should target both students in computer science (for in-depth operational knowledge of e-infrastructure) and students in other disciplines (to use e-infrastructure to enhance their research or work capabilities). New professional curricula need to be developed to reflect the changing needs for skills and human resources. University curricula should reflect the dynamic nature of e-science and e-Infrastructures by being flexibly complemented with courses provided by experienced practitioners on the latest technological developments. Bert Van Pinxteren highlighted that learning should be lifelong, multi-channel, adapted to users' level, combining on-the-job training with short intensive courses and complemented with professional exchanges including with the industry. All trainings should provide certificates for the participants and be systematically assessed for quality to identify best strategies and content.

Certification and accreditation for e-Infrastructures training

The discussions highlighted the need to develop standardized training contents, especially on standards and interoperability of e-Infrastructures in order to provide coherent training at the European scale. A jointly agreed scheme should be developed to accredit and certify e-Infrastructures related trainings so that courses given by different providers can be compared and acknowledged by different actors, including both learners and institutions. This would contribute to the mobility of scientific and technical staff and help share high-quality training materials and programmes. This certification scheme should be developed in multi-stakeholder partnerships with actors from the academia and the industry. This work should take into account relevant efforts, such as the EU e-skills strategy (http://ec.europa.eu/enterprise/sectors/ict/e-skills/index_en.htm) and e-competences (http://www.ecompetences.eu) with already established frameworks and multi-stakeholder collaboration with common motivation, as introduced by Andre Richier. Creating a harmonized accreditation system for trainings and respective certification will support recognizing service providers' skills at the international level and advance their careers.

Open collaborative repository of resources and best practices

Participants agreed that sharing trainings and resources at the European level is essential for effectively building capacity for e-Infrastructures. Existing courses and summer schools across Europe from undergraduate to PhD levels (e.g. EGI Training Marketplace, PRACE and EUDAT) need to be mapped out in order to make existing good quality training for e-Infrastructures actors visible. Open, accessible repositories where training material can be deposited and shared together with annotations for excellence and descriptions of trainings will improve overall coordination and sharing of resources. Iryna Kuchma also suggested exploring possibilities for collaboration with the UNESCO Open Educational Resources platform. As a continuous effort, systematic collection of best practices

from e-Infrastructures projects should be facilitated in order to promote knowledge transfer between projects and actors and promote e-infrastructures takeup for potential new users. Information of relevant tools for e-science take-up and management should also be gathered, such as DCC Data Management Planning Tool (DMP Online) to support users in developing research data management, PURE Research Information System to allow management of research funds and link to publications and data, as well as user-friendly science gateways.

Incentives and support for on-the-job learning

Lifelong learning must be supported to allow e-Infrastructures developers, operators and other personnel to constantly remain up-to-date on the state-of-the-art. Therefore, even experienced staff should be encouraged and rewarded to follow trainings to optimize their skills or acquire new ones in enterprise architecture, strategy and innovation, a practice that needs to be supported by respective supervisors. Career paths need to give the opportunity to developers to envision their professional development and satisfy their aspirations, creating an environment conducive for outstanding performance for e-Infrastructures. Offering variations in their everyday work routines, such as exchanges with other groups, challenging situations offering the opportunity to innovate, professional flexibility and sufficient salaries should contribute in sustaining employees' motivation to 'invest' in their job. Peter Wittenburg suggested RAMIRI/CLARA type of education, exchange and hosting program with hands-on courses and forums for concept, solution and code debates with guidance by proven seniors. Specific funding should be allocated to organize skills' building upgrades with short intensive courses on technological evolutions by highly specialised experts. Staff exchanges, hosted placements and secondments between various e-Infrastructure operators, research communities but also between the academic field and industry at the European level should also be facilitated as additional training patterns. It was also suggested to set up a pool of specialised experts to teach other experts in a collaborative manner across Europe and fill in national gaps.

Institutional and European strategies

It was considered important to have a shared vision, based on solid statistical evidence on supply and demand and agree on a common agenda for actions (including a roadmap) by all stakeholders. Coordinating bodies, task groups, fora and policy think tanks need to be set up with the aim to create the required leading force, policies and advocacy at national levels. The discussions highlighted the need to develop institution-wide policies relating to e-Infrastructures and data-driven science, which would be aligned with national and European developments in the process. Through coherent institutional strategies for research and data management and collaboration among the various stakeholders concerned, the institutions can systematically steward research projects in their scientific, technical and human resources planning for e-Infrastructures. This would ensure understanding of the importance of research e-Infrastructures and help allocate required resources for training and job profiles. Leadership is required by the EC so that member states align towards the common goals. Conditional funding was suggested as a measure to ensure compliance of infrastructures with standards towards the vision for a pan-European research infrastructure, open science and demand for transparency. EPSRC (Engineering and Physical Sciences Research Council) provides a good practice example in the UK by having requested the institutions it funds to develop a roadmap that aligns their policies and processes with EPSRC's expectations by May 2012 and to be fully compliant by 2015.

Awareness raising and early engagement

Starting to stimulate attraction to e-Science from an early age is necessary to create a critical mass of future experts and e-Scientists in Europe. Computational thinking and data management considerations should become part of the regular school curricula, which would require support from national ministries of education. Specific courses relating to e-science and e-Infrastructures could be promoted and teachers could be used as multipliers following e.g. the lessons learnt from 'Simulated Worlds' project presented by Ruediger Berlich, with experience of teachers dedicated and willing to work as "HPC ambassadors". It was suggested to consider utilising highly-skilled, recognized and influential enthusiasts with excellent communication skills and experience using e-Infrastructures' to act as 'champions' as done e.g. in the SeIUCCR project mentioned by Claire Devereux, in order to improve the awareness of e-infrastructures. Providing real-life examples of how e-infrastructures can support and transform every research practice and help discover new knowledge in all fields of science is necessary to convince funding bodies, policy makers and potential users into further exploring how e-Infrastructures could benefit them.

Supporting communities for e-Infrastructures

Constant interaction and co-evolution is necessary among people working with e-Infrastructures to ensure their best design, operation and usage. Specific investments and efforts are required to organize interdisciplinary collaboration and coordination. Through education and supported by the required leadership and policies, it is important to build a culture of openness and communication within and across research teams actors. Furthermore, institutions should create and support fora, professional networks and communities of practice in order to enable e-Infrastructures developers, operators and users to interact, share practices and ideas, and to learn from each other at the national, European and international levels. Such communities for interaction and innovation at the national, European and international levels can help prevent duplication of efforts and enhance knowledge exchange and pioneer e-Infrastructures. The DevCSI (Developer Community Supporting Innovation) presented by Torsten Reimer and the WENMR project presented by Claire Devereux consist of two positive examples of community platforms and engagement.

Encouraging collaborative e-science mindset

Trainings should not only focus on acquiring technical knowledge and skills but also aim to challenge and improve mindsets towards open and collaborative practices. A mindset for open e-science culture should be instilled in scientists of all disciplines early on through education and training -- use, reuse and innovate as formulated by David de Roure. David Gauckler highlighted the need to encourage digital research culture which efficiently benefits from the digital media for communication and collaboration. Open science culture includes communication and collaboration with other scientific actors and also with society, as well as sharing resources and best practices for the benefit of the community. The objective should be to get the traditional boundaries out the way around disciplines, companies, machines, ages, citizens, etc. This could be supported by providing good examples and integrating them into training and education, considering adequate IPR and licensing schemes on all forms and stages of data (from raw to processed). This requires inaugurating new reputation and incentive structures by funders and research organisations, rewarding and encouraging scientists to share and collaborate.

5 Concluding Recommendations

As a response to the development needs recognised during the consultation, following concrete recommendations for European actions were devised:

Specific studies and support actions

- Studying the current status and estimate future needs for e-Infrastructures professionals in different tasks on European level to provide concrete information on the scale and nature of the problem
- Mapping of existing situation in the Member States regarding policies and tools adopted across institutions regarding skills and human resources on e-Infrastructures
- Mapping of roles, job titles and respective tasks with the aim to reach common definitions for job profiles, roles and responsibilities
- Mapping of existing university curricula and short-term courses in a 'who does what where' database to highlight existing training provision schemes
- Support establishing an inventory and repository at the European or even international level for e-science support tools for institutions, e-Infrastructures developers, operators and scientific users
- Support pilot projects to explore innovation in education and training in developing new content for training and new training approaches for lifelong learning and on-the-job training
- Support the creation of a global training context comprising joint online technical workshops on specific topics to share best practices
- Support specific curricula development for "new" jobs and profiles
- Support integration of cross-disciplinarity and computational thinking in all scientific curricula at all levels, including secondary schools
- Support establishment of bottom-up communities of specialists (e-Infrastructures developers, operators and scientific users) through workshops (charrette meetings) for cross-disciplinary communication and collaboration, together with the necessary platforms
- Support studies and projects to develop and collect best practices for institutional change in acknowledging new job profiles, developing career paths and incentive systems

Integrating skills and human resources aspects into Horizon 2020 projects

- Highlight the need for project proposals to present how they comply and take into account interoperability with other existing efforts and sustainability planning of the e-Infrastructure development, including for related human resources (e.g. through conditional funding)
- Integrate plans for training of project personnel and project training outreach with quality assessment of training outcomes as compulsory elements of e-Infrastructures project planning
- Give positive credit to e-Infrastructures whose project planning shows collaboration in training provision and participation, considers best practices published in the field and complies with the EC guidelines and strategies for research organisations
- Require that all training material developed in e-Infrastructures projects are made available under open license in a common open repository
- Require that project proposals specify responsible team members and their experience for e-Infrastructures development and operation
- Require that project proposals include specific data-driven science issues highlighting new roles and tasks, such as data management strategy

- Fund training both as separate activity and as part of projects with broader aim, including dedicated funding for users', and short intensive courses for developers by a European pool of specialists to upgrade skills. Support expansion of training to include skills for additional tasks, such as teaching, using social networking tools, advocacy, communication, 'green IT'
- Include into project reporting identification of transferable best practices and tools that can ease the takeup of e-infrastructures and data-driven research practices.

Advocacy and leadership

- Facilitate development of pan-European e-Science policies, strategies, and development of standards for the institutions, disciplines and national levels whereby research e-Infrastructures are developed as sustainable services and not only as projects, with respective sustainability considerations to their human resources.
- Maintain a continuous dialogue with member states to encourage national investments in e-Infrastructures and related policies, education and training
- Establish and lead collaborative multi-stakeholder group including academia and industry for developing European wide schemes for recognising e-Infrastructures related skills and competence profiles and accreditation for related training
- Facilitate professional mobility and recruitment at European level for experts with skills in e-Infrastructures development, operation or training provision
- Recommend actions to Member states and research organizations to deal with e-Infrastructures related skills and human resources issues
- Coordinate systematic collection and showcase regularly best practices relating to e-Infrastructures skills, human resources management, support tools and related institutional practices
- Establish and promote e-Infrastructures community champions to advocate new jobs and skills needs at schools, universities and scientific communities

Optimised relevant funding tool usage

- Increase funding for e-Infrastructures programme and earmark specific part of funding for training and human resources related activities
- Better promote the possibility to use Structural funds for e-Infrastructures, together with best practices for using these resources in engaging new users and facilitating user-developer-operator communities and relevant training.
- Coordination of funding schemes is necessary as well as conditioning investments to meeting specific requirements in the long-term. Cross-reference and advocate both Marie Curie and e-Infrastructure calls to potential applicants of e-infrastructures skills and human resources development.

Bibliography

Auckland, M. (2012). *Re-skilling for research.* London: RLUK. Retrieved from http://www.rluk.ac.uk/files/RLUK%20Re-skilling.pdf

Ball, A., & Duke, M. (2011). *How to Cite Datasets and Link to Publications.* Digital Curation Centre, Edinburgh. Retrieved from http://www.dcc.ac.uk/resources/how-guides

Conway, E., Matthews, B., Giaretta, D., Lambert, S., Wilson, M., & Draper, N. (2012). Managing Risks in the Preservation of Research Data with Preservation Networks. *International Journal of Digital Curation, 7*(1).

e-IRG. (2008). *Education and Training Task-Force Report.* Retrieved from http://www.e-irg.eu/images/stories/publ/task_force_reports/ettf_long_report_final_july08.pdf

e-IRG. (2010). *Blue Paper to ESFRI.* Retrieved from http://www.e-irg.eu/news/news/276/esfri-and-e-irg-endorse-blue-paper-on-e-infrastructure.html

e-IRG. (2011). *White Paper.* Retrieved from https://gridlab.man.poznan.pl/Meetings/Zakopane2012/e-IRG%20WhitePaper_2011.pdf

EMC. (2011). *Data Science Revealed Community Survey.* Retrieved from http://www.emc.com/collateral/about/news/emc-data-science-study-wp.pdf

Henty, M. (2008, April). Developing the Capability and Skills to Support EResearch. *Ariadne*(55). Retrieved from http://www.ariadne.ac.uk/issue55/henty

JISC. (2008). *Data scientists study.* Retrieved from http://www.jisc.ac.uk/news/stories/2008/10/datascientist.aspx

Kurowski, K., Nabrzyski, J., Katz, D., Merzky, D., Gabrielle, A., Glinos, K., & Seidel, E. (2012). *Grid Computing: The Next Decade - Report and Summary. Zakopane Workshop.* Zakopane, Poland. Retrieved from http://www.gridlab.org/Meetings/Zakopane2012

Lyon, L. (2012). The Informatics Transform: Re-Engineering Libraries for the Data Decade. *International Journal of Digital Curation, 7*(1).

McKinsey Global Institute. (2011). *Big data: The next frontier for innovation, competition, and productivity.* Retrieved from http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

NSF 07-28. (2007). *Cyberinfrastructure Vision for 21st Century Discovery.* Retrieved from http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf

Voss, A., Mascord, M., Fraser, M., Jirotka, M., Procter, R., Halfpenny, P., . . . Anderson, S. (2007). e-Research Infrastructure Development and Community Engagement. *UK e-Science All Hands Meeting.*

Workshop agenda

CONSULTATION WORKSHOP (30TH MAY 2012):

"SKILLS AND HUMAN RESOURCES FOR E-INFRASTRUCTURES WITHIN HORIZON 2020"

Venue: Avenue de Beaulieu 25, Room 0/S9, 1049 Brussels

Agenda

09:30 **Opening session**

Welcome and introduction,
Kostas Glinos (Head of Unit, INFSO.F3 GEANT and eInfrastructures)

Objectives of the workshop,
Kirsti Ala-Mutka (INFSO.F3)

Keynote: Policies and strategies for eInfrastructures Human resources development in Europe
Gudmund Høst (e-IRG Chair)

Keynote: Building capacity and capability for data: Requirements, Challenges, Opportunities
Liz Lyon (Digital Curation Centre, UK)

10:45 *Coffee break*

11:00 **Human resources for e-Infrastructures operation and maintenance**

- Andrew Smith, EMBL EBI
- Claire Devereux, STFC UK
- Iryna Kuchma, EIFL
- Bert van Pinxteren, TERENA

Questions and discussion

12:15 *Lunch break*

13:15 **ICT and data skills in short supply for e-Infrastructures development**

- Torsten Reimer, JISC
- Peter Wittenburg, Max Planck Institute
- Krzysztof Kurowski, PSNC
- Gabriella Cataneo, IDC
- David de Roure, Oxford e-Research Centre

Questions and discussion

14:30 **ICT and computational skills for scientific communities and industry**

- Pekka Manninen, CSC
- David Gauckler, Université de Strasbourg
- Nuket Yetis, Tubitak
- Roberto Barbera, Univ of Catania, INFN
- Ruediger Berlich, Karlsruhe Institute of Technology

Questions and discussion

15:45 *Coffee break*

16:00 **Synthesis discussion**

17:15 **Conclusions and closing**
*Niobe Haitas (*rapporteur)
Kostas Glinos (HoU, INFSO.F3)

List of Participants

Name	Surname	Organisation
Roberto	Barbera	University of Catania and INFN
Ruediger	Berlich	Steinbuch Centre for Computing / Karlsruhe Institute of Technology
Gabriella	Cataneo	IDC
David	de Roure	Oxford e-Research Centre
Claire	Devereux	STFC/SelUCCR
David	Gauckler	LERU
Gudmund	Host	e-IRG
Iryna	Kuchma	EIFL
Krzysztof	Kurowski	Poznan SuperComputing & Networking Centre
Liz	Lyon	Digital Curation Centre
Pekka	Manninen	PRACE training centre
Torsten	Reimer	JISC
Llorba	Rossend	Netherlands Organisation for Scientific Research
Andrew	Smith	EMBL/EBI
Bert	van Pinxteren	Terena
Peter	Wittenburg	RAMIRI
Nuket	Yetis	Tubitak (formerly)
Martin	Lange	European Commission DG EAC
Andre	Richier	European Commission DG ENTR
Francesco	Fusaro	European Commission DG RTD
Athina	Zampara	European Commission DG INFSO
Wim	Jansen	European Commission DG INFSO
Kostas	Glinos	European Commission DG INFSO
Kirsti	Ala-Mutka	European Commission DG INFSO

Related projects and resources

These projects and resources were mentioned by the workshop participants as good examples:

CERIF for datasets: http://www.sunderland.ac.uk/research/ceriffordatasets/

CLARIN (Common Language Resources and Technology Infrastructure): http://www.clarin.eu/

DARIAH (Digital Research Infrastructure for the Arts and Humanities): http://www.dariah.eu/

DCC (Digital Curation Centre): http://www.dcc.ac.uk/

DCC Data Management Planning tool: http://www.dcc.ac.uk/dmponline/

DevCSI (Developer Community Supporting Innovation) :http://devcsi.ukoln.ac.uk/

Devops: http://devops.com/

EMBL-EBI (European Bioinformatics Institute) for User training: http://www.ebi.ac.uk/training/

EPIKH training and secondment programme: http://www.epikh.eu

EPRSC Expectations http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx

EUDAT (European Data Infrastructure): http://www.eudat.eu

European e-Competence Framework: http://www.ecompetences.eu

Geant: http://www.geant.net/pages/home.aspx

NSF ongoing strategic program: http://www.nsf.gov/pubs/2012/nsf12090/nsf12090.jsp

OpenAIREplus (Open Access Infrastructure for Research in Europe): http://www.openaire.eu/

Pure Research Information System: http://www.st-andrews.ac.uk/staff/research/pure/

SeIUCCR (Supporting e-Infrastructure Uptake through Community Champions for Research): http://www.ngs.ac.uk/seiuccr/home/

SIM4RDM (Support Infrastructure Models for Research Data Management): http://www.sim4rdm.eu

TERENA TRANSITS courses http://www.terena.org/activities/transits/

Training Marketplace customisable events calendar and resource repository: http://www.egi.eu/services/support/training_marketplace/index.html

UNESCO OER platform: http://oer.kengasolutions.com/

XSEDE (Extreme Science and Engineering Discovery Environment): https://www.xsede.org/