Abstract

As e-Science and Grid computing technologies mature from research prototypes to production infrastructure underpinning real research, the necessity for education and training in the development and use of these technologies grows. A workshop was held at NeSC to foster interaction between the funders, providers, and consumers of e-Science education and training in the UK, and to discuss possibilities for collaboration in such activities. The workshop reviewed the wide range of activities already underway, from awareness-raising events with participants from the commercial sector through MSc-level courses to workshops training more experienced scientists to use e-Science technologies. This revealed a willingness within the community conducting those activities to collaborate in their development and delivery, and, more concretely, several specific requirements were identified as necessary for the future success of education and training in UK e-Science.

In this background paper we review the issues raised during the workshop, to inform and stimulate discussion within the community leading to the development of a coordinated plan for education and training within UK e-Science. We also outline the reasoning behind our four specific recommendations, namely that:

- A coordinated programme for e-Science education and training is needed in the UK;
- A dedicated testbed Grid infrastructure should be provided for education and training purposes;
- A repository should be established to facilitate the sharing of materials amongst those undertaking education and training activities; and
- A study should be funded to develop the detailed specifications for the proposed testbed Grid and repository and, more generally, to assess the match between the requirements for education and training within UK e-Science and activities underway or currently planned.
1. Introduction

The technologies underpinning e-Science are starting to mature, and develop from the research prototypes developed by the few into, potentially, the computational infrastructure used by the many. This is clearly the moment at which education and training in e-Science and Grid computing need to commence in earnest, and a range of programmes are now underway in the UK, under the aegis of the Regional e-Science Centres and Centres of Excellence, through e-Science projects (notably EGEE) and within a number of universities. While e-Science and Grid technologies are maturing, it is still the case that this is a rapidly developing field, and expertise in different aspects of it is distributed throughout the UK e-Science community. This motivates the idea of collaboration in the development and delivery of education and training in e-Science, to ensure that up-to-date material is developed with the minimum duplication of effort. At the same time, some level of coordination between its funders, providers and consumers is necessary to ensure that the content delivered meets the requirements of the e-Science community, within both academia and the commercial world.

A workshop on “Education and Training in UK e-Science” was held at NeSC on 1&2 November 2004, with the primary goals of fostering interaction between those undertaking or planning education and training activities within the UK e-Science community, and assessing the scope for collaboration in their development and delivery. The programme for workshop is presented in the Appendix, and all presentations can be downloaded from the NeSC WWW site, at the following URL:
http://www.nesc.ac.uk/action/esi/contribution.cfm?Title=487.

This document summarizes the salient points that arose during the workshop, and its purpose is to stimulate and inform further discussion of this topic within the community. Its structure is as follows: Section 2 lists the current and planned education and training activities within the UK e-Science community represented by workshop attendees; Section 3 outlines the requirements for education and training; Section 4 discusses a number of issues relating to the delivery of education and training, notably the computational infrastructure required, the sharing of materials, the role of e-learning and distance learning, and financial considerations; and Section 5 presents conclusions and recommendations.

For the most part we make little distinction between “education” and “training” in this document. Several distinctions could be made. Education could be defined to be the study of underlying principles and training to be the development of practical aptitude in applying them, but most e-Science courses will contain a mixture of these two activities, albeit in a ratio which will vary, depending on the requirements of the particular audience. Training tends to be used to describe activities undertaken by those in employment, but the notion of Continuing Professional Development (CPD) blurs the boundary between education and training for such people. Education tends to include more assessment than training, but that is not necessarily so.

2. Current and Planned Activities

A range of education and training activities are currently underway within the UK e-Science community, including the following:

2.1 NeSC Training Team

The NeSC training team is developing and delivering courses for EGEE, the National Grid Service and OMII. Core courses are targeted at applications developers, systems administrators and users.

3 Properly “e-Research”, as these technologies can be beneficial beyond the scientific domains where they have typically been deployed to date.
Under its EGEE role, the team leads 22 partners in the EGEE training activity and facilitates and delivers courses across Europe. Under its national role the team provides a series of courses at the NeSC, advertised widely for anyone in the country to attend, and will deliver training elsewhere in partnership with universities, regional e-science centres and e-research projects. Within NeSC's regional role, courses are given both at the NeSC and at research institutes in this region. Further details are accessible via http://www.nesc.ac.uk/training.

2.2 University of Edinburgh: MSc/Diploma in e-Science

The University of Edinburgh is preparing a new MSc/Diploma programme in e-Science, to admit its first cohort of students in the autumn of 2005. The taught component of this degree programme is intended to provide students with an appreciation of the principles underlying e-Science, as well as developing their aptitude in the use of current e-Science technologies. Two semesters of taught courses will comprise the Diploma, while MSc students will complete their year of study with an extended individual research project which applies what they have learnt to a real e-Science research problem. Initially, the MSc/Diploma will be delivered by staff from the Schools of Physics and Informatics, plus the NeSC training team, but it is anticipated that further Schools will become involved in due course, meeting the multidisciplinary goals for the programme. Further details are available at the following URL: http://www.ph.ed.ac.uk/postgraduate/degrees/msc_escience.html.

2.3 University of Edinburgh: MSc in High Performance Computing

EPCC, the Edinburgh Parallel Computing Centre, offers a one-year taught MSc in High Performance Computing (HPC), which includes coverage of e-Science topics. This well-established programme provides an excellent grounding in HPC technologies and their practical application. The MSc in HPC will appeal to students who have a keen interest in programming and would like to learn about HPC and parallel computing. It aims to attract students who hold degrees in the physical sciences, engineering, computer science or mathematics, or who have equivalent work experience. The course has a strong practical focus and students have access to world-leading HPC platforms and technologies. EPCC is an institute within the School of Physics at the University of Edinburgh. It has an international reputation as a world-leading centre of expertise in HPC and Grid computing. The MSc is supported by EPSRC. For full information see www.epcc.ed.ac.uk/msc/.

2.4 University of Glasgow: MSc/Diploma in Advanced Computing Science

The Grid computing course in the Advanced Computing MSc at Glasgow University attracted 16 full time students in its first year, exceeding expectations. It is planned to repeat this course on a yearly basis. The next course will be offered in January 2006. The course itself focused predominantly on the underlying principles and challenges associated with Grids. That said, students were expected to develop significant software assignments based upon technologies such as Globus toolkit version 3.3, Condor and PERMIS. Currently four advanced MSc students are undertaking their dissertations in Grid-related areas under the supervision of Richard Sinnott. The lecture material for this course is available at www.nesc.ac.uk/hub/projects/dyvose. Discussions are on-going regarding rolling out these Grid courses to a wider audience.

2.5 Training and Education: Current & Planned Activities for WeSC.

The Welsh e-Science Centre continues to develop its training and education activities. The current portfolio of 8 training courses will be maintained and it is intended that at least two
more will be added in the next year, while supporting literature and case study material will be reviewed and updated where appropriate. Work is in progress to add a new Centre case study, in both paper and DVD format, and additional project case studies. WeSC continues to run a regular programme of e-Science & Grid awareness raising events, at the Centre and at venues in Wales, for both industrial/commercial and academic audiences. It will also be increasing our existing pool of expertise in High Performance Computing, Grid Technology and Immersive Visualization to include High Capacity Data Storage and Advanced Pattern Matching. WeSC is participating in an increasing number of weekly seminar/short talks with other Centres and Universities via AccessGrid. Discussions and trials are also taking place to involve other Schools at Cardiff University in utilizing Access Grid for collaborative working, and, where appropriate, additional functionality will be incorporated for Access Grid. Further information is available at http://www.wesc.ac.uk/training/.

2.6 National Cancer Research Institute Informatics Initiative

The National Cancer Research Institute (NCRI) Informatics Initiative was created to maximise the impact of cancer research. To achieve this, a Task Force composed of key individuals across diverse disciplines was convened and a series of data-sharing principles were developed then communicated to the research community. In addition the Initiative has identified and supported a number of high impact demonstrator projects that illustrate the utility of an integrated approach to cancer research. If the common aspects of data sharing are to be realised efficient and effective training in the underlying technologies and utility of distributed data and computational grids is required. The developments in training and education at the NeSC have been recognised as a valuable resource in e-science and Grid technology and it is acknowledged that the Initiative will greatly benefit from such a programme. Further information on the Initiative is available at http://www.cancerinformatics.org.uk/.

3. Requirements for e-Science Education and Training

The mix of skills needed for e-Science is dependent on what role the e-Scientist will play; she may be a member of a small (local or geographically distributed) research team working on a specific e-Science project, or a member of a larger centralised team providing e-Science services to many different projects. For the moment we assume that the former case is more common than the latter, and that a typical, small e-Science project team is a heterogeneous group which may comprise:

- A principal investigator who orchestrates the project;
- A co-investigator who is an expert in the application science of the project;
- A computational scientist who has formal training in the application science and is skilled in some aspects of using computer to advance the application science;
- A computing engineer who has formal training in computer science and may have some knowledge of the application science.

The principal goal of this team is to generate efficiently new results in the application science domain, not to advance computer science. Hence, the computing engineer plays an important but supporting role in the team. Also, e-Science education and training should not seek to bring every team member to the same level of computing skills: each team member should bring distinct, complementary skills and knowledge to the team. Rather, the goal is to bring every team member into a ‘common space of discourse’ where team members can effectively exchange ideas, learn and develop new science. The common space of discourse does not require that all team members be able to communicate in an equally effective manner with every other team member. For example, statements made by the application scientist may be restated in other terms by the computational scientist so that the computing engineer can grasp the significant concepts of the communication.
Consider the role of the computing engineer in an e-Science project team. The engineer is likely to carry primary responsibility for all phases of the software development cycle, as follows:

1. Gathering requirements
2. Investigating possible known software solutions
3. Integrating software components
4. Deploying and maintaining solutions
5. Back to phase #1

The engineer’s formal computer science education should provide a solid background for these tasks: a typical undergraduate curriculum in computer science defines a set of fundamental courses (covering computational logic, programming skills, software engineering methods, database programming, scripting languages, data structures and algorithms) following which students are likely to choose a few from of a selection of possible specialisations. However, we note that:

1. The computing engineer in an e-Science project will need skills from many areas of specialisation. For example, e-Science computing tasks could require knowledge of networking, middleware, service-oriented architectures, distributed systems, parallel systems, database systems, language processing, rule-based systems, human-computer interaction and knowledge systems. Typically a student will study courses in only one or two areas resulting in basic knowledge gaps due to specialisation.
2. Even topics included in the curriculum may not be covered in a practical or applied manner. Basically this is knowledge that needs to be transformed to skill.

To identify requirements for education and training to fill these knowledge gaps and turn textbook knowledge into skills, we consider typical activities in each software development phase.

1. Gathering requirements. It is essential that the computing engineer understand ‘enough’ of the science application domain to communicate effectively with the rest of the project team. The engineer must also be able to answer questions and provide information to other team members intelligibly. Vocabulary in the application domain is critical to this phase (and all phases). Typically the engineer will not be versed in the application domain, and developing the necessary vocabulary takes time and effort.

2. Investigate solutions. There may well be only one computing engineer in the project team. For this reason (and others) the engineer should strive to reuse existing solutions. To do so, the engineer must know how to search for, obtain, install and evaluate existing applicable software technologies, often specific to the particular application domain. Computer science students are trained to create new solutions ‘from scratch’, with little emphasis placed on reusing existing technologies. Also, the engineer must define metrics to evaluate the technologies for applicability. This requires an open mind (with respect to using another engineer’s work) and ability to carefully read a manual. These skills are not emphasised in traditional computer science training.

3. Integrate solution components. The engineer must identify technology gaps, develop new software to fill those gaps, and test and deploy a complete system. It is important that the solution be sufficiently robust for the application science research to progress using the software. Although commercial quality software is not required, the necessary level of robustness requires rigorous application of software development methodologies. The typical computer science curriculum does include a course in software engineering, but, unfortunately these methodologies are rarely (if ever) used again in other parts of the curriculum: this is a serious problem in all software development projects (not only e-Science). The engineer should also have strong skills in systems analysis, but typically this is not part of the basic CS curriculum, but rather a specialist topic taken by the few, not the many.

4. Deploy and maintain the system. The engineer must document the system and train team members to use the software for productive scientific research. A ‘customised software vocabulary’ may be required to effectively communicate computer science concepts to other
team members: this is the inverse of learning application domain vocabulary. Finally, system maintenance (which is probably never taught in the computer science curriculum) requires a ‘stick with it’ attitude.

We have discussed the role of the computer engineer so far; but what of the application scientists? What skills must they possess to aid the success of the project team? Their first requirement is to contribute to the “common space of discourse”, which will require their learning some of the vocabulary and concepts familiar to the computing engineer, just as much as the converse. A technically savvy applications scientist can be a great bonus for an e-Science project team, but, equally, a scientist (especially a PI) who thinks he knows enough to set the technical choices for the project can be a disaster. The different members of the project team contribute their complementary expertise, and should be respected for doing so.

This clear separation of roles may seem overstated and artificial. Writing in New Scientist, Andrew Davies, vice-president of Discovery IT at GlaxoSmithKline noted that “The boundary between the workbench scientist and the IT support staff is increasingly blurred. Nowadays scientists have to be pretty IT savvy to know how to do their job. They may not have the know-how to build applications, but there is so much data to be analysed that many scientists are being turned into IT experts, information analysts and scientists all combined.” The technologies used in e-Science and Grid computing may not have yet matured to a point where the typical applications scientist can be an expert in them, and so the separation between scientist and engineer may still be valid. Even if that is so, the situation will soon change and it is surely time to include exposure to e-Science technologies into the education of at least the most IT-enthusiastic of the coming generation of scientists. That is certainly the motivation behind the new e-Science MSc/Diploma at Edinburgh, which aims to prepare scientists from a range of backgrounds for a future research career using e-Science technologies, as well as producing the trained computer scientists and software engineers who will further develop those technologies.

4 Delivering Education and Training

It is clear from the previous section that there are numerous requirements that need to be met to ensure that the investments made in the UK in e-Science are fully exploited. These requirements are broad in scope, but the overall goal is clear: to educate future computer scientists to engineer improved Grid middleware and to educate e-Scientists to use existing middleware to solve scientific problems.

At the moment, these requirements must largely be met from academia, since not many IT training companies provide training courses in e-Science and Grid technologies. The demand for web services training is significant however, as there is major corporate interest in web services, and in the adoption of XML interfaces for distributed applications. In terms of business interest in Grid services, there is currently only really a demand for overview courses of the type “what is the grid and how could it benefit my business”. The financial sector in the UK is beginning to use Grid services in earnest - e.g. using large clusters to do all their end of day calculations – but the general perception however is that while many technologists in the mainstream commercial sector may have come across Grid services, they are often currently not really sure what they are, nor do they often have a clear idea how they could be used within their organization. If web services technology is anything to go by, the initial demand for e-Science technologies will be for high level training seminars as team leaders and senior technologists get to grip with the concepts, such as those organised for life sciences.

4 New Scientist, 2 August 2003, pp. 48-49.
5 Based on feedback from Nick Todd from Conygre Consultants.
informaticians by the PRISM Forum⁶. Then as the technology matures, and senior people begin to grasp the significance of the technology, then roll outs begin of grid projects, and training requirements ramp up to implementation classes.

4.1 Content
One of the greatest challenges in delivering materials for educating and/or training future Grid engineers and e-Scientists is the fluidity of the technological landscape. Grid technology and associated standards are perpetually evolving with new recommendations and software from standards bodies and solutions providers. This has been exemplified in the last year with the move from Grid infrastructures, to OGSI-based Grid services and then towards WSRF web/Grid services. The evolution of OGSA is also a key issue that makes the development and delivery of any form of education or materials difficult. Trainers and educators need to be sure that they are developing materials which has some expectancy of life time. Developing and delivering educational materials based upon explicit technology - e.g. Globus toolkit version 3 - is fraught with dangers associated with a moving technology base.

In educating the future generation of Grid engineers, a balance between concepts and principles associated with Grids and e-Science is needed. Taking the Grid Computing module at the University of Glasgow as an example, the lecture material developed has been focused upon the underlying principles and challenges associated with Grid technologies. Thus whilst there might be numerous technologies say for job scheduling (Condor, Sun Grid Engine, OpenPBS, Maui,…), the basic principles of job scheduling and the specific challenges of large scale, wide area job scheduling remain the same.

However, a course based solely upon basic principles is not appropriate for what is essentially a practical subject, and hands-on experiments and investigations using current state-of-the-art technologies are needed. At Glasgow this has been through looking towards Globus toolkit version 3, but this technology has provided a vehicle through which many of the basic principles have been demonstrated, rather than itself being the cornerstone of the course. The assessment of the Glasgow module comprises a written examination (70%) and marked coursework (30%). The marked coursework is divided up into three problem sets and one large programming assignment. This assignment is expected to take up to 30 hours of student effort, from design, through implementation to final write-up.

A key requirement on Grid education is a broad scope and balance. Grid technology touches on many areas from security, usability, job scheduling and data management etc, and care must be taken in matching the scope of a course to its target audience: high level overviews of the Grid can be provided to undergraduate students, but such topics are probably better delivered to computer science students that have the necessary grounding in related subjects. Numerous prerequisites were required of students wanting to take the Grid Computing module at Glasgow. The students had to be competent in Java, knowledgeable in internet technologies, have experience of distributed algorithms and systems, and done some work on databases. (As it happened numerous students did not meet all requirements hence extra lecture material had to be provided, e.g. on XML based technologies and standards, and on web services.)

4.2 Computational Infrastructure
To illustrate a number of the infrastructural issues concerned with delivering education and training in e-Science, we consider some practical demonstrations from the National Institute for Environmental eScience (NIEeS) and the Grid Computing module in the Advanced Computing Science MSc at Glasgow. NIEeS has a remit from NERC to provide training in

⁶ The PRISM Forum: http://www.prismforum.org
e-Science for the environmental sciences community and the topics it covers include the use of grid computing, data grids, and tools to support virtual organisations comprising individuals located at geographically distributed sites. The example of the NIEeS is a standard case where members of the community who want to be educated in e-Science need to see facilities and tools in action. The most satisfactory way to show these is both from lecture demonstrations and classroom practicals. To illustrate what is possible, here are a few key lecture examples from recent courses:

Figure 1: Screen shot of the live demonstration on computing a phase diagram. The two plots are displayed on a web browser with live updates. The lower plot is built up point by point as more computations are performed on a Grid.

1. Demonstration of workflow within a grid environment: a trivial example
This is a very simple talk-through example. Individual calculations need to be quick in order that the demonstration doesn't pause for too long to wait for tasks to finish, and the calculations need to be sufficiently simple to be able to be understood by the audience. The NIEeS example uses Condor's DAGMan tool to create simple workflows involving three jobs,
where the third job depends on the output from the first two. In fact these jobs involve simple additions of numbers to make it very transparent.

2. Demonstration of high-throughput computing in a grid environment
This demonstration involves calculating the stability of the two most important mineral phases at pressures and temperatures within the Earth's upper mantle. A few hundred short calculations are launched onto a set of Condor pools that are flocked together. The idea is that when each job is completed, the result of which phase is more stable is added to a data file. This data file is then inspected every 30 seconds and an SVG graph is updated on a web browser. Bit by bit the audience can see the graph being populated with data points (see Figure 1), but also during the time of the demonstration there is a tool that enables the audience to see which resources are being used. The condor flock that is used involves the computers of 5-6 independent institutes, and it is instructive to see jobs migrating between pools. Moreover, the job uses XML data management tools to make it work, and if appropriate it is possible to show some of the inner workings.

3. Demonstration of integrated compute and data grids
The eMinerals project has developed an integration of a compute infrastructure based on the use of Globus and the Condor-G tool, and a data grid infrastructure based on the Storage Resource Broker. The demonstration shows the use of a DAGman workflow run from within Condor-G in which data input files are placed in the SRB, the job then begins to execute on a remote resource by downloading the data from the SRB, and on completion the job puts the output file back onto the SRB. The resultant files can then be viewed from a laptop in the lecture room. The demonstration also shows how to launch a grid proxy using the Globus toolkit with an e-Science digital certificate.

NIEeS also have classroom demonstrations, and some examples are:

1. Running a computational fluid dynamics code on a grid environment using the globus job submit tools and running many instances to investigate parameter sweeps. This program gives a nice visualisation at the end (see Figure 2).

![Figure 2. Student running a computational fluid dynamics simulation on the UK Level 2 Grid.](image-url)

These demonstrations from NIEeS, which are analogous to those which can be run in other settings, have identified a number of infrastructure requirements:

1. The need to generate temporary **digital certificates** and have them recognised by providers of resources. In practice this has not been a problematic issue for NIEeS, but it does place
demands on the owners of resources. It would be useful to have individual institutes set up as
recognised CAs.

2. The need for **guaranteed available resources** for all times during a lecture demonstration. A
lecture demonstration will fail completely if the grid job doesn't execute immediately. For the
simple workflow demonstration, NIEeS need 2-3 free computers in a condor pool. For the
larger phase diagram, a set of free resources within a Condor flock is needed. At the present
time NIEeS borrows these resources from the eMinerals minigrid and from the developing
CamGrid infrastructure. When these are being used in production mode, it will either be
impossible to run a demonstration, or else to stop job queues. Similarly, there is a need for
guaranteed available resources for classroom practical sessions. For example, NIEeS can
obtain access to classrooms of computers for one per participant as an interface to a wider
grid, but it does need to make arrangements to have access to external grid resources.

3. Teaching **middleware installation** is difficult. For example, with a dedicated set of
resources it could be possible to teach people how to set up, configure and manage a condor
pool. This would need more than one machine per participant. One could imagine running an
intense hands-on course where sets of PCs are arranged into clusters, with condor and globus
being installed and built into a functioning grid structure.

4. Demonstrations are limited to software that is available within the public domain, and, for
example, NIEeS have had some problems extending its range of demonstrations due to **IPR
constraints**.

Another significant requirement is the establishment of the necessary infrastructure. This is
non-trivial and the effort involved should not be underestimated: Grid technologies, and
dependencies between packages, still pose a challenge for installation and general usage. This
includes getting students equipped for working on the Grid, e.g. through X.509 certificates
(issued from the UK e-Science Certificate Authority at RAL), or as was the case at Glasgow,
establishing a local certificate authority.

For the Grid computing module at Glasgow, the following list of software was used.

- **PERMIS Privilege Allocator 1.5+, PERMIS API 1.3** requires:
  - Java SDK 1.4.2+ (includes JDNI)
  - XALAN Java 2
  - IAIK JCE (contained in Privilege Allocator 1.5)
- **Globus Toolkit 3.3**, which requires:
  - Apache Ant 1.5+
  - Junit 3.8.1
  - gcc 3.3+
  - YACC or Bison
  - GNUtar
- **Condor 6.6.5**
- **OGSA-DAI 4.0**
- **Tomcat 4.1.24+ (with GT3.3)**
- **JDBC Database (Postgresql)**

This software was initially built on a single machine and the disk image duplicated across the
training cluster. In addition, a lab server was set up with OpenLDAP 2.1.22 for the attribute
certificates generated by the PERMIS Privilege Allocator (PA1.5).

Another problem which was addressed in Glasgow to make Grid training realistic and
applicable to current experiences was security. To be realistic training should be based upon
usage of resources such as the National Grid Service (NGS), or an implementation of a very
similar software stack, since the NGS is representative of best practice in the UK Grid
community. However, to allow such resources to be used requires appropriate X.509 certificates, and there are scalability issues with central certificate authorities such as the UK e-Science CA at RAL. A more realistic approach is to have local educational/institutional certificate authorities issue local certificates and have these used by students to address security aspects, e.g. restricting the resources that students may access and use. Such local certificate authorities need to be recognised by others - e.g. owners of resources such as the NGS - since there is no root domain of trust. Bridging technologies can be used to overcome this, but there is an explicit requirement that these are recognised and accepted by the NGS.

4.3 Sharing materials through a repository

The practice of e-Science is inherently collaborative, as is the development of e-Science technologies, given the geographical distribution of the relevant expertise. For the same reason it is natural that cooperation should be evident in e-Science education and training, and for that cooperation to be successful will require some sort of mechanism for sharing materials to be used in these activities. These materials might include documents, PowerPoint files, videos, software and data for practicals, and in this Section we discuss issues relating to the use of a repository to facilitate the sharing of such digital materials within the UK e-Science community. Many of these issues have legal aspects, so we quote at some length from relevant sources in what follows.

4.3.1 Concepts

To illustrate some of these issues we consider the example of “Induction to EGEE” course prepared by NeSC: EGEE is an EU funded project, its acronym standing for “Enabling Grids for e-Science”. This course can be found on the NeSC and EGEE websites, and in the NeSC-EGEE repository. In the course, and indeed in one presentation, may be found slides taken from commissioners of the European Commission, colleagues in the EGEE project, the Globus Alliance, and the European Datagrid project. Some are modified, some are used “as-is”. Issues that have arisen include:

- **Attribution:** Wherever possible attribution is given, in an acknowledgements slide and, where the originator is known, on the slide itself. Some slides have a heritage of several years and the originators are not known: by the time slides from this presentation are re-used and modified for purposes - maybe by trainers in other regions of the EGEE project - the history of some of the slides can be lost. Metadata concerning the provenance of individual slides may be desirable, but in many cases is not available, and is certainly not readily implemented.

- **Permission for re-use:** Where the creator of a re-used slide is known, permission has been sought for re-use. However there is also a possibility that a slide is used out of its intended context, to illustrate a theme that was not the originator’s intent. As commercial interests in e-Science technologies evolve, so these issues might become more complex: currently good will and the commitment to cooperation simplify the issues that arise.

- **Ease of re-use.** Web-sites typically hold PowerPoints files matching talks given at events. In the induction for EGEE course, within one talk there will be sets of slides on several themes. These sets of slides might be a resource that an educator or trainer could re-use, but unless there is prior knowledge of the contents then these might not be discovered from the web-site. Rather than considering a PowerPoint presentation to be the most common molecule of training material, a presentation can be considered to be built from several smaller components, and these components also need to be held and discoverable in a repository.

The above illustrates a number of the facets of sharing material: attribution, alteration and re-use imply moral and the legal rights, and ease of re-use the more evident technical, metadata and functional requirements.
Amongst concepts that commonly occur in the creation of repositories are:

**Learning object (LO)**

“A learning object is any resource that can be used to facilitate learning and teaching that has been described using metadata” (JORUM 2004)

**Learning object repository**

“A learning object repository is a collection of LOs having detailed information (metadata) about them that is accessible via a network or the Internet. In addition to housing LOs, repositories can store ‘locations’ for objects that are held elsewhere…. Such locations are given various names, including virtual objects and resource stubs”. (JORUM 2004)

**Digital Rights Management (DRM)**

“Digital rights management is commonly about offering resources and reaching agreement on the uses to which the resources may be put. However, there is not often much discussion about what it means to use a resource. The Copyright, Designs and Patents Act, 1988, grants exclusive rights to the copyright owner, who may, in turn, grant licences to other to carry out the restricted acts”. (Duncan et al., 2004) (page 16).

These “are the reasons why DRM is needed:

- to allow staff and students in HE and FE to make use of digital resources in confidence that they are adhering both to what is permitted by law and to the rights-holders permitted uses;
- to enable self-publication of resources supported by the ability to declare the permitted uses;
- to enable users to work within the confines of copyright (such as being able to copy an insubstantial part of a work and working within the parameters of fair dealing) knowing they are not engaging in infringement;
- to ensure that all of the above can operate in an internationally connected, digital environment.

DRM in an academic environment should be an ‘enabler’ not a ‘preventer’. Its purpose is to let people work as freely as possible in the knowledge that they are both working within the bounds of the law of copyright and respecting the rights of others.” (Duncan et al., 2004) (page 7)

“Over the last few years, the education sector has begun to identify and address a need for DRM systems designed for managing rights in LOs. A simple, supportive DRM solution for use in education (and beyond) is the Creative Commons (CC). Creative Commons provides tools for generating licences to allow discovery and use of content that an author wishes to offer for widespread distribution free of charge. Even when an author wishes to distribute her/his work free of charge, it is important that s/he asserts ownership of IPR and clearly expresses the rights that s/he grants so that the content may not be misappropriated, e.g. for commercial use, and to ensure that her moral rights are upheld. A CC licence is provided in both human-readable and machine-readable forms. DRM requires standardisation of rights expression for machine-to-machine communication and for presentation to users (e.g. teachers and learners). A digital rights expression language (DREL) is a standard means of expressing

---


the rights associated with a digital information object. The DREL is a key component of a sophisticated DRM system”. (JORUM, 2004, page 12)

Copyright ownership.
Who has the ‘right’ to give permission to allow use of material? The author of the material is the first owner, except where the author is an employee and acting in the course of employment. In this case, the employer is the owner. In many cases, the IPR are held by the originator’s employer. Copyright comprises a bundle of exclusive rights, where “exclusive” means that only the copyright owner has the right to exercise these unless he/she assigns or licenses to another. The rights are:

- To copy a work
- To issue copies of the work to the public
- To rent or lend the work to the public
- To perform, show or play the work in public
- To communicate the work to the public
- To make an adaptation of the work

If a third party wishes to carry out any of these activities, then authorisation is required from the copyright owner. This can be by way of assignation or licence. Assignation basically means an outright transfer of the rights to a third party. Licence is more limited, and would require terms to cover both ‘in rights’ (to the database) and ‘out rights’ (to the user).

Although there has been much harmonisation of copyright law both at international level within Europe, there remain differences in the law between Member States, and often still greater differences with countries beyond the EU. Where works protected by copyright are to be exploited abroad, then some thought should be given to these differences.

Database rights
The development of a repository of learning materials within Europe is likely also to result in the subsistence of what is called the ‘sui generis’ database right. This right is given to the maker of a database, where that maker invests in the obtaining, verification or presentation of the contents of the database. Investment can be financial, technical or human resources. Any investment that goes towards the creation of materials for placing in a database is not counted towards the database right. So, for example, the resources needed to develop e-Science training materials is not relevant. But when those materials exist, then the investment that goes towards collating (obtaining) them for placing in the database, the verification of those materials once they are in the database, and the investment in the presentation of the materials once in the database are all relevant for the existence of the right.

Given the resources that are necessary to collate, verify and present materials within an e-Science training repository, it is likely that the right will exist. The right then gives to the maker the right to prevent the unauthorised extraction and re-utilisation of a substantial part of the contents of the database. This test is both quantitative and qualitative. Thus the amount taken out of the database will be important for the infringement of the right, as will the quality of the extraction. Any third party may take out insubstantial parts without infringing the right. That does not, however, prevent the maker of the database from using technical controls to limit access.

It should be remembered that the database right is in addition to any copyright that might subsist in the underlying materials. Further, ownership may vest in different people. Thus, you might have an individual who owns the copyright in a power point presentation, but it may be an Institution who owns the database right.
**Fair dealing**
Quoting from (JISC, 1998):
“'fair dealing': any copying activity that is permitted under the terms of the [Copyright] Act not to require permission from, or to pay fees to the copyright owner”
“It is not fair dealing to post part or all of an electronic publication on a network or WEB site open to the public. Clearly posting part or all of an electronic publication on a network, or WEB site open to the public, has a much wider purpose than the creation of a single copy for the purpose of individual's research or private study. It therefore falls outside the scope of fair dealing”.

**Licenses**
Quoting from Duncan et al. (2004), pages 14-15:
“Many resources are made available within the existing legal framework but without any consideration given as to contractual provisions that may govern use. It is often preferable to provide a statement of permitted uses and conditions of use thus making it clear which exclusive uses have been licensed and for what purposes. This is usually done by licensing use of a resource”.
“[The] GNU General Public Licence (GNU, 1991) [is] commonly used for open source software. There are no options, though there are many variants, with this licence (also known as copyleft) but many people are willing to make resources available under its fixed conditions. It provides a legal framework rather than trusting to none at all. It offers an irrevocable licence to copy, modify, and redistribute the work freely but only under the condition that any redistributed material must be distributed under the same licence conditions”.
“More recently Creative Commons has introduced a selective licence approach allowing authors to choose from a limited number of conditions which are then combined to form a licence agreement. One feature of Creative Commons is that, like the GNU General Public Licence (GPL), it can include a “share-alike” condition. Creative Commons originated in the US and is finding some adoption in the UK through Oxford University, although there is some question about its applicability under English law and JISC is investigating an equivalent for use in the UK”.

**Z39.50**

### 4.3.2 Examples of Repositories

(a) The JISC Information Environment Service Registry (IESR)
Quoting from (Apps, 2004):

---


The JISC Information Environment Service Registry (IESR) (http://www.mimas.ac.uk/iesr/) contains quality descriptions of collections of resources available to researchers, learners and teachers within UK Higher and Further Education. Alongside the collection descriptions, which enable resource discovery, are technical details of the 'informational' services that provide access to them, enabling determination of the best access option to a collection of interest. The IESR also contains descriptions of 'transactional' services that are not based on an explicit collection but provide a significant service, for example an institution's OpenURL resolver. Additionally the IESR contains details of the parties (agents) that own the collections and administer the services.

“The IESR is primarily a machine-to-machine middleware shared service within the JISC Information Environment (http://www.jisc.ac.uk/index.cfm?name=ie_home).”

“The service and agent metadata properties are based on Dublin Core (http://www.dublincore.org) where possible. The description of all entities includes some IESR-specific properties. Data is supplied to the IESR as separate entities, but data within, and output from, the IESR is composite. A composite collection record includes: the collection metadata; the metadata for all the services that provide access to it; the metadata for its owner agents; and the metadata for the agents that administer its services. IESR metadata is covered by a Creative Commons (http://creativecommons.org) licence: non-commercial, share-alike, attribution required.”

(b) JORUM
The JORUM, the JISC online repository for learning and teaching materials, will be a repository service available to all Further and Higher Education Institutions in the UK from August 2005, providing access to materials and encouraging the sharing, re-use and re-purposing of them between teaching staff. The JORUM repository will be multi-disciplinary and cross-sectoral and will contain learning and teaching resources e.g. learning materials that lecturers can provide to their students, and resources that support teaching staff in doing their job, for example lesson plans, tutor guides and schemes of work. Teaching staff will be able to locate resources using simple search facilities in the repository, download the resources for use locally, and deposit new resources using JORUM's deposit licence procedures, as long as their institutions own the IP rights and agree to deposit them. Given the increase in student to lecturer numbers, and teaching work that is done between teaching teams, or across institutions, the JORUM repository offers an easy way for staff to collaborate with colleagues on sharing resources throughout the UK. EDINA and MIMAS (the JISC-designated national data centres) are leading the JORUM Project. For more details and to follow developments see: www.jorum.ac.uk

Service model
JORUM will have site subscriptions i.e. Further and Higher Education Institutions in the UK interested in taking the service will have to subscribe and sign site licences. The service will be free to institutions and free at the point of use for end users.

Licensing system and workflow
The JISC have decided that JORUM shall launch next year open only to deposits of institutional-owned, and not individually-owned, content. Project teams and other teams wanting to deposit institution-owned content in JORUM will have to sign an institutional deposit licence currently being drawn up by the solicitor.

The JORUM team have received legal advice that licences signed by institutional representatives would have to have electronic digital signatures to be acceptable in law. Because this kind of system is not open to us to implement at the present time, the institutional deposit licences will have to be administered off-line and in print.
There will only be one possible set of permissions granted by depositing institutions for each of their objects, allowing the end user to modify, excerpt and aggregate content i.e. this allows full re-use and re-purposing for educational use.

The JORUM service will not launch with any kind of community deposit function i.e. it will not be possible for individuals to deposit the content they own themselves in JORUM, nor will it be feasible for an individual to re-publish re-purposed content in JORUM that they have downloaded from the service. Research into the legal issues involved in implementing this functionality will form part of the R&D work and it is hoped this will be implemented at a later date.

Depositors of content into JORUM will complete metadata fields asking for details of permissions for use and IP owners. Each object exported will contain the content, plus metadata file giving details of IP owners and permissions, plus either a licence or a URL to a place holding the licence, to be determined.

End users will have to register for the service, using their Athens accounts, to enable certain functionality in the system and allow the JORUM team to collect useful statistics for JISC and the institutions. End users must also click "I Agree" to a general set of terms and conditions as they log in to the service.

Licensee
The Licensee of all the JORUM licences will be HEFCE. The JISC is setting up a not-for-profit company from the end of 2005, which will take over Licensee duties for JISC licences.

(c) MIT's OpenCourseWare

This is a “free and open educational resource for faculty, students, and self-learners around the world” and further details can be found at http://ocw.mit.edu/index.html. Key points concerning IP and licensing\(^\text{11}\) are to:

- "Obtain agreement that open courseware is premised on making materials available to end users under "open" license terms that allow use, reuse, adaption, and redistribution.
- “Preserve faculty rights of their course materials to allay faculty concerns.
- “Develop a sound copyright review and clearance process, and build awareness around intellectual property requirements. Faculty often use third-party materials in teaching; these materials must be cleared for publication with copyright owners, or removed. This is a manageable process.
- “Ensure that permissions from third-parties are compatible with the open terms of open courseware publication. Commercial publishers are least likely to grant liberal license terms, but there are alternatives.”

Quoting from http://ocw.mit.edu/OcwWeb/HowTo/tech-metadata.htm on metadata standards:


“This standard specifies the syntax and semantics of Learning Object Metadata, defined as the attributes required to fully describe a learning object which are defined as any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning. The MIT OpenCourseWare (MIT OCW) project adopted the use of this standard for organizing metadata. Online at http://ltsc.ieee.org/wg12/”...“It is recommended that

\(^\text{11}\) From http://ocw.mit.edu/OcwWeb/HowTo/IP-Licensing.htm
opencourseware implementors adopt the IEEE LOM metadata standard. It is flexible and can be extended to accomplish a wide range of metadata requirements”.
MIT’s OCW initiative adopted a slightly modified version of the Creative Commons model license effective with its January 2003 publication. (http://ocw.mit.edu/OcwWeb/HowTo/IP-Enduserlicense.htm)

(d) Repository for EGEE
One of the NeSC training roles is to enable training across the EGEE project, and to this end a light-weight repository is being created. It currently (January 2005) holds presentations, primarily, and can be searched via a web page (http://egee.nesc.ac.uk/trgmat/index.html). The repository was designed by David Fergusson, implemented by Susan Andrews, both of NeSC, on a best-efforts basis without dedicated resources.
The term “light-weight” refers to a) the lack of explicit management of IPR; b) access is solely by the web page - there is no interoperability with other repositories. The IPR issues for EGEE material are relatively simple because rights of contributors who are partners in EGEE have been signed over to the EGEE project as a part of the partner’s contract with EGEE. Material is accepted and held in the repository on trust; EGEE relies on being able to remove material if problems are reported (for example, if a slide contains an image from a third party organisation, without copyright issues being resolved). Whether a licence for re-use of training material will be implemented in future remains to be resolved, but if so then it would be a licence to support re-use, for example EGEE embraces open source for non-profit use of its software.

4.4 The Role of e-Learning and Distance Learning
A recurring theme in this paper is that the distributed nature of the expertise in the fast-moving field of e-Science motivates collaboration in education and training. So far, our discussion has focused on collaboration in the development of education and training, but in this Section we consider collaboration in the delivery of education and training – i.e. distance learning, where participants in a course are geographically distributed. This Section also discusses e-learning, but we stress that, while e-learning and distance learning often go together, they are far from being synonymous: many distance learning programmes include a significant e-learning component, but they usually include other activities, while e-learning can be provided as a way of supplementing conventional teaching on-campus, equally as well as it can contribute to distance learning.

4.4.1 AccessGrid
Within the e-Science community, much of the distance learning on offer is provided via AccessGrid. Here we summarise experiences using AccessGrid in two situations – a training and outreach programme and an educational programme.

Deployment and experience of AccessGrid at Welsh e-Science Centre, Cardiff
The Welsh e-Science Centre (WeSC) was an early adopter of Access Grid (AG). WeSC’s first AG suite was set up in October 2001, to the Argonne specification, including Virtual Coffee Room. The room accommodated up to 10/12 people, and consisted of three projectors, four microphones, four cameras, and four machines to drive the audio and video capture machines. Initially there was a programme of “test cruises” to test the facilities in different configurations and collectively learn from one another’s experiences.

In February 2004 WeSC moved into a new Centre building, which contains a purpose built AG suite. This room accommodates 25/30 people, and employs four projectors, four microphones, four cameras, and four machines to drive the audio and video capture. Recent
acquisitions include: a Visualizer, a camera mounted on an illuminated stand which allows small objects and hand drawn material to be displayed as part of an AG session; IGPIX, distributed presentation software, to allow synchronized presentation of PowerPoint material; Electronic whiteboard software; and an echo-canceling microphone. Software has recently been upgraded from AG1 to AG2.3 which offers a more automated set-up for sessions and is less reliant on specialist knowledge.

Mike Daley and John Oliver from WeSC have also been trying out mobile AG configurations (typically a single camera and two projectors for exhibition stand use) outside the Centre, using Internet connections at several different venues in Wales. In each case the exhibition stand became an AccessGrid studio, with images projected on to a large screen created at the back of the stand. A 3-node Access Grid session was set up, with one node being the AG suite at Welsh e-Science Centre at Cardiff University, one node being the exhibition stand and the third node set up using a PIG (laptop and web cam) adjacent to the exhibition stand. Broadband Internet connection at the hotel was used in each case to achieve the network connectivity.

The principle behind this deployment is to attract delegates to the stand to engage with them in conversation. By demonstrating AG and explaining how the global e-Science community use it as a tool for collaborative working allows the WeSC representatives manning the stand to introduce delegates to the concepts behind e-Science and Grid computing. In this way we are using the AG demonstrations as part of the overall WeSC e-Science and Grid Computing awareness raising strategy. This type of demonstration has allowed WeSC to include colleagues who are not physically present at the exhibition to interact with the delegates and talk about different aspects e.g. Dr Roger Philp gave short presentations about the computing resources available at WeSC and via the UK e-Science community. Short Powerpoint presentations can also included using distributed Powerpoint facility, and visualization examples utilizing Vizserver have also been included where there is sufficient bandwidth.

A PIG (Personal Interface to the access Grid) has also been employed for awareness-raising sessions. This consists of a laptop computer with AG software installed, a web cam and a microphone. This can be used at a desk (using a headset) to connect to an AG session from virtually anywhere an Internet connection is available. This configuration has been used as a demonstration on board the Welsh Information Society exhibition trailer. This trailer is a mobile IT presentation theatre which travels around Wales promoting ITC issues to small and medium sized businesses. It is equipped with digital projectors, a sound system, plasma screens and a satellite broadband link. It is relatively easy to connect the PIG into the trailer’s on-board systems and to connect to the AG suite at Cardiff over the satellite broadband link. John Oliver has achieved this on a number of occasions to support e-Science mini-seminars held on the trailer. Once initial technical issues had been resolved the configuration proved to perform well over the broadband satellite link and the demonstrations have been well received by the audiences.

AccessGrid usage is increasingly part of day-to-day life at WeSC. It is regularly used by Centre staff for project meetings, and staff from a number of other schools within Cardiff University have used the WeSC AG suite for meetings with colleagues at remote locations. AG is currently not used for teaching at Cardiff, although trials during 2005 will see weekly computer science seminars transmitted outside the University using AG.

Most people who have used AG at Cardiff have responded positively and having tried an initial session returned to book further ones. Many people have commented favourably on AG in comparison to other videoconferencing technologies, with which they have noted the following problems: unreliable technology; high cost of deployment; poor and inflexible camera and screen positioning; sound cut-off/clipping; and a high dependence on a skilled technical operator. However, for AG most people report how much more flexible it is in the
way images can be ‘dragged and dropped’ anywhere on the screen and revised/repositioned at any time during a session as and when required. Also the sound facility does not switch between speakers, which with conventional systems can cause loss of speech and be a source of irritation. With AG, where it is possible for everyone to talk at once, this rarely happens as most meetings have a chair and normal rules of conduct for a meeting are applied.

In conclusion, the use of AccessGrid for collaborative working and awareness-raising at Cardiff has been steadily increasing since its introduction in 2001. It has not been deployed a great deal for more formalized teaching and education, although from experience gained it has great potential. John Oliver and Mike Daley from WeSC continue to investigate new facilities to integrate into AG sessions to make them more interactive. Currently Mike Daley is evaluating Qwizdom, and interactive classroom response system which allows interactive question and answer sessions using keypads and voting facilities with instant on-line display for results.

Using AccessGrid for distance learning in Australia

The Australian analogue of the UK e-Science programme is APAC, the Australian Partnership in Advanced Computing\(^\text{12}\). APAC has three roles: research support; education; and technology diffusion. Its educational activities focus on the sharing of educational modules between universities, at both undergraduate and postgraduate levels.

A good example of a shared educational module is the 6-lecture Advanced Visualisation course run in 2004 by staff at the University of Queensland and the University of Western Australia, which are more than 4000km apart. This course was put together by one lecturer from each university, and was taken by about ten students at each location. Overall, this experiment was a success: students gained university credit for attending the course; they found the experience interesting and valued the chance to attend lectures given by leaders on the other side of Australia; and they were enthusiastic about attending the lectures.

This course did reveal some clear requirements for a successful educational experience over AccessGrid. Firstly, lecturers need to consider their presentation style, to encourage student interaction with them: this is especially true for the students at the remote location, although, after the first lecture, students did feel comfortable to ask and answer questions with a remote lecturer. There is a clear need for a “whiteboard” – a means of making impromptu explanations or diagrams of course content – and there is a need for a way of running programs easily at each venue. There were also some teething problems in the use of OpenOffice with AccessGrid, but these are not insurmountable.

Therefore from this example it can be seen that AG has great potential for education and training in e-Science, provided that care is taken with the supporting technologies, the choice of topic and the style in which the material is presented. Such a mode of teaching may be particularly relevant to smaller universities, who want their students to be exposed to topics in which they don’t have local expertise.

4.4.2 e-Learning for e-Science

In assessing the role that e-learning may play in e-Science education and training, it is important to realise that content does not make a course. As mentioned above, MIT have placed a lot of content online through their OpenCourseware (OCW) initiative, but they emphasise that OCW “is not a degree- or certificate-granting program [nor] an MIT education”. Online content must be supplemented with communication between students and

\(^{12}\) APAC web site: [www.apac.edu.au](http://www.apac.edu.au)
teaching staff for a successful educational experience; how this communication is provided – synchronous vs asynchronous, face-to-face vs remote – will depend on the nature of the course and of the people taking it, but the communication is the key.

The experiences of the Technology-Assisted Lifelong Learning (TALL) unit at Oxford University suggests that there are two possible models for successful online courses – a High Investment and a Low Investment model - which can be summarised using the following table:

<table>
<thead>
<tr>
<th></th>
<th>High invest model</th>
<th>Low invest model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costs per course</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>TALL involvement</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Author costs</td>
<td>High</td>
<td>Lower [?]</td>
</tr>
<tr>
<td>Number of students</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Cost of delivery per student</td>
<td>Low</td>
<td>Medium</td>
</tr>
<tr>
<td>Online tutors</td>
<td>Less specialised</td>
<td>More specialized (= the authors)</td>
</tr>
<tr>
<td>Tutor numbers available</td>
<td>Many</td>
<td>Few?</td>
</tr>
<tr>
<td>Scalability to large numbers</td>
<td>Yes (if no residential)</td>
<td>Limited</td>
</tr>
<tr>
<td>Study skills needed by students</td>
<td>All abilities, potentially</td>
<td>Well established [?]</td>
</tr>
<tr>
<td>Delivery time</td>
<td>Medium to long</td>
<td>Immediate to short</td>
</tr>
<tr>
<td>Flexibility to change materials</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Risk</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Payback period</td>
<td>Longer</td>
<td>Short</td>
</tr>
<tr>
<td>Potential payback</td>
<td>High</td>
<td>Low</td>
</tr>
</tbody>
</table>

Different parts of a hypothetical e-Science curriculum may be suitable to different models. Some topics – e.g. the principles and problems of generic distributed computing – are relatively stable and could have a shelf-life justifying the High Investment approach, while others – e.g. those covering specific technologies – are likely to be more ephemeral, as well as more specialised, and the Low Investment model is probably more appropriate for them.

TALL has devised a standard methodology for more long-lived courses, a cyclic process of Specification-Design-Production-Evaluation-Specification, etc, as well as a standard technological infrastructure for implementing them. Many of the issues in the design of online
courses mirror those in general course design, e.g.: understanding the existing expertise of the target audience; identifying the key points to convey; and thinking about appropriate mechanisms for assessment. However, there are further concerns which are more specific to the design of online material. How and where will the audience learn? – are they students in a lab, staff in an office, professionals learning on the move or at home? One important consideration is establishment of a feeling of community within the student body: research and experience suggests that dropout rates for online learning courses can be reduced significantly if a feeling of community can be engendered from the start.

The Low Investment model is more suited for fast-moving subjects, where courses may only be used once or twice before they have to be discarded. In this model editorial and postproduction activities are reduced to a minimum. Often all that can sensibly be done is to capture lectures (voice, PowerPoint, annotations, handwritten notes) and make them available to students online, relying on a simple discussion board mechanism for communication amongst the students and between them and the teaching staff.

TALL’s experience suggests that two of the most serious problems in providing online material is discovering relevant existing resources, and having access to them in a changeable form: both of these motivate the idea of a repository operating a Creative Commons or similar licencing model.

4.4 Financial Considerations

The earlier Sections of this document have discussed the need for a coordinated education and training programme for e-Science in the UK, but who should pay for it? A detailed answer to that question is beyond the scope of this document, but it seems clear that a mixed funding model is likely to be appropriate, given the mixture of different types of education and training activity which appear to be required.

As e-Science technologies migrate into the mainstream the notion of a distinctive “e-Science community” will weaken, and so, probably, will the role of the e-Science Centres in education and training: as the coherence of a distinctive e-Science community weakens, so will its coupling to the Centres, and, hence, so will their ability to respond to the needs of that community for specific training activities. But if the Centres do not provide the training, who will? - and how can a cohesive body of knowledge be maintained if not through a coordinating structure like the e-Science Core Programme?

This migration into the mainstream will also change the nature of the training to be given. From providing core skills to a limited number of people working primarily in e-Science or Grid computing, it will develop a wider market, providing what is essentially Continuing Professional Development, through teaching staff working in a wide range of disciplines, covering technologies which underpin their work but which only constitute part of the skills set that they require for it. As the market widens, “training the trainers” may be the most cost-effective way of disseminating expertise.

As e-Science standards and technologies stabilize, it will become easier for universities to offer courses covering them – and a number of universities are starting to do this already. These will be funded via the same routes as the universities’ other courses, and there will be the same competitive market for applicants as exists for other degree programmes. If there is only a free market, though, we risk the duplication of effort discussed earlier, as well as incomplete coverage of the more cutting-edge topics, in which expertise is likely to continue be distributed across the country, so the benefits of collaboration will remain.
5. Conclusions and Recommendations

This paper has discussed many issues relating to e-Science education and training in the UK, but it remains partial, in both senses of the word: it is both biased and incomplete because it represents the views only of the attendees of a single workshop who are already engaged in e-Science education and training. These are clearly the people who will form the foundation of any future programme in this area, but, unsurprisingly, the workshop attendees did not include many applications scientists not already active within the e-Science community and who will come to require knowledge of e-Science technologies in the next few years. Furthermore, despite the fact that a number of representatives from companies engaged in e-Science/Grid-related activities were invited to the workshop and expressed interest in its goals, none were able to attend. The views of both of these groups need to be taken into account in planning a coordinated e-Science education and training programme for the UK.

However, despite the partial view of the subject provided by the workshop attendees, a number of conclusions are apparent:

- A wide range of education and training activities is underway within the UK e-Science community, addressing the requirements of a number of different target audiences.
- The range of requirements for education and training will only broaden as e-Science technologies mature and migrate into the mainstream, necessitating a variety of kinds of education and training, probably funded from a number of sources.
- While some parts of the hypothetical e-Science curriculum will stabilize, the expertise in the more cutting-edge topics will remain distributed throughout the country, so that collaboration in the development and delivery of e-Science education and training would be beneficial both to reduce duplication of effort and to ensure that relevant, up-to-date material is delivered.
- There is a willingness within the community currently delivering education and training activities in the UK e-Science community to collaborate, and this should be exploited.
- Unstable software and the difficulty of obtaining guaranteed resources in a realistic Grid environment cause significant problems for those running education and training activities now, and this situation could be ameliorated by the provision of dedicated resources with associated staff expertise.
- The discovery of relevant learning objects is currently a major problem for those developing courses, so the existence of a repository with an appropriate licencing model would be a great advantage.
- Online learning modules can have a significant role to play, but they must be supplemented by communication amongst the students, and between the teaching staff and the students.
- Interactive distance learning – e.g. performed over AccessGrid – can provide an excellent means of providing e-Science education and training.

From these conclusions we derive four recommendations.

1. **A coordinated programme for e-Science education and training is needed in the UK.** The UK e-Science programme was funded by reason of its strategic importance to both industry and academia in the UK. To ensure that the promise of the technologies and expertise developed within the programme is fully exploited, it will be necessary to coordinate the work of the funders, providers and consumers of e-Science education and training.

2. **A dedicated testbed Grid infrastructure should be provided for education and training purposes.** The availability of guaranteed resources is a necessary prerequisite for successful education and training sessions. This requirement is unlikely to be met with the same production Grid infrastructure used to support research projects, since the latter should not be exposed to the level of disruption that is inevitable in systems used for training. This testbed should be as similar as possible
to the production system, and should have associated staff expertise in the installation and configuration of e-Science and Grid software, as well as its use.

3. **A repository should be established to facilitate the sharing of materials amongst those undertaking education and training activities.** This would be a great help to those developing new education and training programmes, but for that help to be maximized the repository will need not only the metadata required to aid discovery of relevant material, but also a licencing model which recognizes effort through attribution, but which also allow the modification of materials. The JISC JORUM repository may provide the necessary features, and this should be investigated further.

4. **A study should be funded to develop the detailed specifications for the proposed testbed Grid and repository and, more generally, to assess the match between the requirements for education and training within UK e-Science and activities underway or currently planned.** This background paper has presented a partial view on many of the issues relating to education and training in UK e-Science, but a proper study will be required to develop these ideas into a coordinated programme. This should address the interests of all the potential stakeholders in this programme – the funders, providers and consumers of e-Science education and training – and should endeavour to ensure that the full range of requirements from education and training can be met. This process should include the review of possible solutions for provision of a testbed Grid infrastructure and a learning object repository, and deliver a costed proposal for their implementation.
Appendix

The table below presents the programme from the NeSC workshop on “Education and Training in UK e-Science”. Copies of all presentations may be downloaded from the following URL: http://www.nesc.ac.uk/action/esi/contribution.cfm?Title=487

Monday, November 1
10.00 Welcome to NeSC (Malcolm Atkinson)
10.05 Workshop Intro (Bob Mann)

<table>
<thead>
<tr>
<th>Session 1: Current and Planned Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.15 Edinburgh e-Science MSc (Bob Mann)</td>
</tr>
<tr>
<td>10.40 the Grid computing module of the advanced CS MSc at Glasgow (Rich Sinnott)</td>
</tr>
<tr>
<td>11.05 Edinburgh HPC MSc and EPCC training (Judy Hardy)</td>
</tr>
<tr>
<td>11.30 NeSC training activities (Mike Mineter)</td>
</tr>
<tr>
<td>11.55 WesC training and aware-raising activities (John Oliver)</td>
</tr>
<tr>
<td>12.20 Experiences in the National Institute for Environmental eScience (Martin Dove)</td>
</tr>
<tr>
<td>12.45-1.30 lunch</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 2: Requirements for education &amp; training in e-Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.30 NCRI Informatics Initiative (Max Wilkinson)</td>
</tr>
<tr>
<td>13.55 Experiences from edikt (Denise Ecklund)</td>
</tr>
<tr>
<td>14.20 Astronomy in the Virtual Observatory era (Bob Mann)</td>
</tr>
<tr>
<td>14.55 TBC</td>
</tr>
<tr>
<td>15.20 - 15.45 tea</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 3: Panel discussion, featuring funders, providers and consumers of e-Science education and training</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.15 JORUM (Moira Massey)</td>
</tr>
<tr>
<td>17.00 Close of Day One</td>
</tr>
</tbody>
</table>

Tuesday, November 2nd:

<table>
<thead>
<tr>
<th>Session 4: Collaboration for education &amp; training in e-Science (cont’d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>09.00 Legal issues: IPR and DRM (Charlotte Waelde)</td>
</tr>
<tr>
<td>09.35 EGEE training and repository plans (David Fergusson)</td>
</tr>
<tr>
<td>10.10 National Grid Service (Peter Berrisford)</td>
</tr>
<tr>
<td>10.30-11.00 coffee</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session 4: Collaboration for education &amp; training in e-Science (cont’d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.00 eLearning for eScience (Tristram Wyatt)</td>
</tr>
<tr>
<td>11.35 Educational Programmes in Australian E-Science (Kevin Burrage)</td>
</tr>
<tr>
<td>12.10-16.15 Session 5: Discussion</td>
</tr>
<tr>
<td>16.15-17.00 Session 7: Next steps</td>
</tr>
<tr>
<td>17.00 Close of meeting</td>
</tr>
</tbody>
</table>