

Education

Strategies and Policies to Support and Advance Education in e-Science

Malcolm Atkinson, Elizabeth Vander Meer, David Fergusson, Clive Davenhall, and Hamza Mehammed

In previous installments of this series, we've presented tools and resources that university undergraduate and graduate environments must provide to allow for the continued development and success of e-Science education. We've introduced related summer (<http://doi.ieeecomputersociety.org/10.1109/MDSO.2008.20>) and winter (<http://doi.ieeecomputersociety.org/10.1109/MDSO.2008.26>) schools and important issues such as t-Infrastructure provision (<http://doi.ieeecomputersociety.org/10.1109/MDSO.2008.28>), intellectual property rights in the context of digital repositories (<http://doi.ieeecomputersociety.org/10.1109/MDSO.2008.34>), and curriculum content (<http://www2.computer.org/portal/web/computingnow/0309/education>). We conclude now with an overview of areas in which we must focus effort and strategies and policies that could provide much-needed support in these areas.

We direct these strategy and policy recommendations toward key stakeholders in e-Science education, such as ministries of education, councils in professional societies, and professional teachers and educational strategists. Ministries of education can influence funding councils, thus financially supporting our proposals. Professional societies can assist in curricula revision, and teachers and strategists shape curricula in institutions, which makes them valuable in improving and developing education in e-Science and (perhaps) e-Science in education.

We envision incremental change in curricula, so our proposals aim to evolve existing courses, rather than suggesting drastic upheavals and isolated additions. The long-term goal is to ensure that every graduate obtains the appropriate level of e-Science competency for their field, but we don't presume to define that level for any given discipline or institution. We set out issues and ideas but don't offer rigid prescriptions, which would take control away from important stakeholders.

Background

We define e-Science broadly as the invention and application of computer-enabled methods to achieve new, better, faster, or more efficient research, innovation, decision support, or diagnosis in any discipline. For instance, the UK Arts and Humanities Research Council refers to the Virtual Vellum project it supports—as well as many other projects rooted in the arts and humanities—as e-Science, embracing the term without reservation (see www.rcuk.ac.uk/escience/news/artshum.htm). e-Science draws on advances in computing science, computation, and digital communications (see “Century of Information Research Strategy (CIR): A Strategy to Meet the Research Challenges and Opportunities in the Century of Information” by Malcolm Atkinson and colleagues, http://wikis.nesc.ac.uk/escienvoy/Century_of_Information_Research_Strategy_%28CIR%29:_a_strategy_to_meet_the_research_challenges_and_opportunities_in_the_century_of_information).

We've identified the need for educational programs to prepare graduates to make valuable contributions within the current technologically imbued social context. In the late 20th and early 21st centuries, we've seen the spread of information and communication technology (ICT) in both the personal and work arenas. These technologies allow for greater connectivity to information at home, as well as providing the means to collaborate effectively on research in wide international networks. ICT has also made it possible to amass large amounts of data that then require analysis, often through further development of computer-enabled research methods. Considering this context, it's clear that most individuals and most disciplines would benefit from curricula that include e-Science methods.

Developing Curricula

As our last article pointed out (<http://www2.computer.org/portal/web/computingnow/0309/>

education), educators must devote significant time to the development of e-Science curricula. They must also consider the needs of diverse student groups, which add complexity to the task of curricula design. Figure 1 gives an idea of these diverse disciplinary contexts.

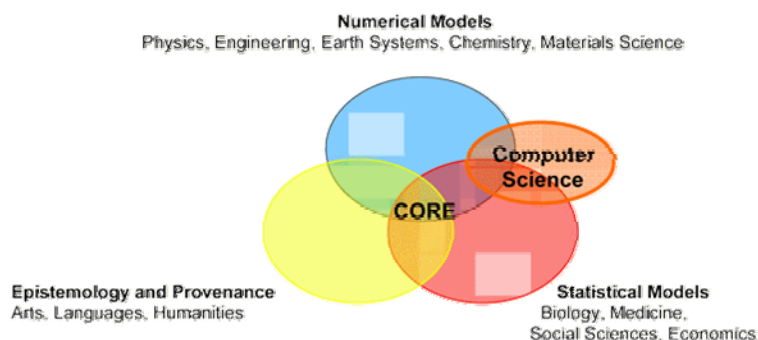


Figure 1. Curricula strands that e-Science educators must take into account.

Of course, this diagram simplifies contexts. We find specialists in all disciplines that use e-Science techniques across these divisions, turning to any e-Science methods that enable their research. But for the typical student, certain priorities exist for courses that roughly fall into these categories. Where does your research fit into Figure 1? What core elements for e-Science has your discipline left out, or not adequately addressed, in the curriculum? Elements such as collaborative working and project management might be missing or in need of development. But we must also consider how appropriate such elements are to the discipline in question and how much technical detail the curriculum should include. It's also important to consider how much computational modeling is already used in a particular discipline. Some institutions, such as Oxford University, teach computational modeling in economics; they offer an MSc in Mathematical and Computational Finance, for example. But other institutions don't offer computational modeling as part of their courses. For instance, the Carbon Management MSc at the University of Edinburgh, which blends business, economics, and climate-change science, could benefit from this approach.

Use of such computational models is long established, certainly from the early uses of electronic computers—for example, to compute shell trajectories in the 1940s (using ENIAC, the Electronic Numerical Integrator and Computer). Meteorology has advanced as a result of computer methods in the 1950s, which led to development of weather forecasting derived from barotropic models that relied on collected and analyzed numerical data. In medical research, we see the first understanding of the mechanisms that cause the heart to beat, depending on computational models that Denis Noble developed at the end of the 1950s (see chapter 5 of Noble's book *The Music of Life*, Oxford Univ. Press, 2008). Today, the working of computational models of the heart is often considered systems biology, and it begins to explain why the shock from a defibrillator can stimulate the resumption of the heart rhythm (see www.integrativebiology.ox.ac.uk/publications/Capability_Computing_article.pdf). Even in computationally driven disciplines, getting enough fluency and understanding of computational methods into the curriculum is a challenge. There's also the new challenge of using data and metadata in these disciplines—for example, in bioinformatics and chemoinformatics.

Computational models allow exploration of phenomena that we can't reproduce in the laboratory, such as the next 100 years of climate change, the collapse of a star, the spiral of matter around a black hole, the aging of a material subject to radiation decay over 1,000 years, or the collapse of a large building. The challenge is not just to build and run such models, but to understand how to use them, when to trust them, and how to interpret their results. Graduates should do more than treat models as "black boxes." They should think and reason about their domain of competence.

These issues of ethics and reliability are at the core of teaching e-Science and are therefore relevant for all disciplines. Critical thinking about models and the ethics involved must be at the forefront of inquiry in every field. Researchers in engineering have thoroughly explored this issue, confronting the problem of making sure that practitioners who are modeling take proper responsibility for the models they use. Consider, for example, risk and the banking crisis. Those responsible for the crisis took risks individually on the basis of narrow modeling. Analysts failed to view or model risks in aggregate, and this aggregate vision could have predicted the crisis and avoided it.

We must support the continued development of computational models in the 21st century without becoming complacent. Thus, students need courses that will critically introduce them to e-Science tools. Again, we stress the importance of designing curricula that not only focus on issues and concepts central to e-Science but also cater to a diversity of student contexts.

Modern research frequently depends on collecting and then interpreting data. This is often called “data-intensive research” (see “Beyond the Data Deluge” by Gordon Bell, Tony Hey, and Alex Szalay, *Science*, vol. 323, 6 March 2009). There’s a long tradition of collecting time-series data, such as meteorological observations for atmospheric modeling; some meteorological time-series in the UK stretch back a couple of hundred years. Researchers then extend this data by reference to other data (such as tree rings, ice cores, and lake sediments) to build information over a large time series. E-Science contributes greatly to managing, collecting, and integrating such multisourced data. Other forms of data collection, such as systematic sky surveys in astronomy and the collection of DNA and protein sequences, bring other challenges, not just in the management and collation of such large data volumes, but also in its description and calibration so that researchers can combine information from multiple collections.

Once data has been collected, annotated with metadata describing its provenance, and made accessible to researchers, the challenges of extracting new information from the data become apparent. Many strategies come into play, such as data mining to discover clusters, recurrent relationships, and outliers. In other cases, researchers analyze the data to show properties predicted by theories, or compare them with the output of models to test whether the understanding of phenomena encoded in the models is consistent with available observations. In many disciplines, specialized methods for collecting, analyzing, and interpreting data are emerging. For educational programs to serve students well, they must include the basics of data-intensive science and might also develop knowledge of the special data-intensive methods appropriate to the discipline.

Resources for Education in e-Science

Delivery of e-Science curricula depends on a number of resources. These resources fall into two key areas: providing educators with the necessary skills to teach, and developing practical skills in students. Some experts already exist, but as we pointed out in our article on curricula development, we need to increase their numbers—for example, through summer schools, winter schools, and refresher courses. At the same time, countries and institutions should set up educational programs in e-Science. To develop practical skills in students, we must think about providing examples of educational materials such as demonstrations and written material that suit the discipline. Repositories to house educational materials and t-Infrastructure to run exercises are also crucial for developing these skills.

Strategies and Policies to Support and Grow e-Science Education

Strategies and policies, then, must support curricula development, considering the complexities we’ve introduced. They should also include means to skill educators to successfully deliver curricula while developing students’ practical skills through adequate access to educational materials and training infrastructure. Crucial in all cases is the promotion of collaborative effort.

Strategy recommendations

In terms of curricula development, the relevant and key stakeholders (as introduced previously) should

- establish a committee of leading educators across disciplines to expedite the creation of curricula goals and principal topics, launched and supported by major conferences highlighting educational priorities and opportunities in the field.

- continue meetings in international contexts, such as the Curricula Development Workshop in Brussels and community group meetings held at OGF 22 and 23, to develop an understanding of educational goals and curricula.
- support e-Science centers in identifying role models to interact with students, and encourage universities to nurture grassroots groups to advance e-Science education.
- continue to build federated repositories of shared experiences and practice in e-Science education.

To support development of practical skills for e-Science, we recommend emphasizing the production of educational materials and then sharing these materials through repositories and sharing use of t-infrastructure. The e-IRG (www.e-irg.eu), OGF (www.ogf.org/index.php), and the EU FP6 ICEAGE Project (www.iceage-eu.org) identified a lack of suitable e-Science textbooks, so it's important for the key stakeholders in e-Science education to support the establishment of specific Web sites and other forums to pool, share, and debate textbook content. For example, see the online SURA Grid Technology Cookbook (<http://hv3.phys.lsu.edu:8000/cookbook/gtcb>) and the forthcoming JISC "Research in a Connected World" brochure. It would also be important to develop incentives such as competitions, in conjunction with editors and publishers, to produce textbooks that follow agreed-upon educational goals and curricula.

Educators and other key stakeholders should then share these materials using repositories and share use of t-infrastructure. We can look to the specific recommendations from the OGF and in previous articles for best practice in these areas. Members of the e-Science educational community should support the development of digital libraries and training infrastructures at local and national levels and provide means to link local and national resources. These initiatives would provide a starting point and standard resources for university programs. Many universities will build on this to tailor education for their student cohorts and discipline specialties.

Policy suggestions

This article is based predominantly on work that has taken place in Europe but is globally applicable. We believe it's preferable for educators who know their own cultural context to translate these strategic ideas. Following on from the discussion of strategies, we can identify policies that would facilitate the advancement of e-Science education in the EU, as initially formulated for the e-IRG Education and Training Task Force (ETTF) Report:

Investment in e-infrastructure education. Investment in relevant education and training, which aims primarily to equip graduates to use e-infrastructure well, should be comparable with the investment that is going into e-infrastructure provision. Universities should adapt their curricula to prepare graduates. Investment, such as incentives to modify curricula and help with the introduction of new curricula, is necessary to trigger the required rapid and extensive change.

Aligning the development of distributed-computation knowledge and skills. Academic institutions, particularly universities, should build on existing "seed" courses and curricula to revise curricula in the majority of disciplines. Relevant professional bodies and individuals proposing to teach this material should undertake further work on curricula. The goal of this alignment should be cross-fertilization, mutual recognition, and increased understanding, not uniformity.

Harmonization of education in the use of e-infrastructure. Professional bodies, such as the Royal Society of Chemists and the Institute for Engineering and Technology in the UK, should identify target attainments in the exploitation of e-Infrastructure for their professions and harmonize these across the European Research Area in accord with the Bologna framework (http://ec.europa.eu/education/policies/educ/bologna/bologna_en.html). Again, the goal of this harmonization isn't uniformity of skills and knowledge. Rather, it's a common framework to support student and worker mobility and mutual recognition of qualifications, particularly where they influence the appointment of staff to positions in which the use or operation of e-infrastructure is life or mission critical.

Standards for identification to enable access to and management of educational t-infrastructure facilities. A task force, set up by the e-IRG, European Grid Initiative (EGI,

<http://web.eu-egi.eu>), and GÉANT (www.geant.net), should extend the eduroam protocols (www.eduroam.org) to cover student and teacher use of collaboration facilities and multisite t-infrastructure.

Standards for sharing training material and t-Infrastructure between institutions. Those in the EU developing e-infrastructure courses should build on the Creative Commons for policies governing the sharing of educational material. The EGI, when it becomes operational, should mediate agreement between National Grid Initiatives (NGIs) on sharing t-Infrastructure.

These policy recommendations apply to the EU, but the core of each point can serve as a springboard for endeavors outside of this context. We see an international need for development in these areas.

Examples of e-Science in Practice across Disciplines

We come full circle and conclude with some exemplary e-Science tools and projects that show the kinds of technologies being used in the areas we identified in Figure 1. These examples are important because they show

- the broad relevance of e-Science methods—that is, that virtually every discipline must consider how to change its curricula to prepare its graduates for work and personal time in which e-Science methods and effects are pervasive;
- the breadth of issues that instructors should consider in designing curricula and for which graduates should be prepared; and
- that there's a rich and publicly available source of compelling examples that teachers in almost every discipline can use. Many of these have clearly stood the test of time, and others raise new challenges.

The Core

The core curricula would include widely applicable tools, services, and data resources. Educators would use examples specific to their discipline to illustrate the curricular strand that generally, but not exclusively, guides research in that field. The following examples provide a taste of some of the resources available. Individuals in disciplines where numerical models are prevalent can use the Matlab tool to develop algorithms, analyze data, and carry out numeric computations (www.mathworks.com/products/matlab). Researchers can turn to large digital-storage facilities for access to a vast array of resources and, often, analytical tools to help them find patterns in data. For instance, the National Center for Atmospheric Research (www.ncar.ucar.edu) provides climate and weather data collections while also offering supercomputing facilities, models, and observing laboratories to facilitate research in earth systems science. The International Virtual Observatory Alliance (www.ivoa.net) provides astronomers with a vast array of integrated resources, which together create a global virtual observatory.

The National Institutes of Health (NIH, www.nih.gov), a key resource for biomedical researchers, serves a similar purpose for the US health-care community. NIH brings together a wealth of information gathered from a network of institutions across the country. For example, it offers information on clinical trials, health care topics, and the NIH Stem Cell Information Page, supporting a wide range of research. The National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/About/glance/programs.html) supports the NIH mission by tackling the data problem (the amount of data and its complexity) through development of databases and software tools for molecular biologists, biochemists, and geneticists. In Europe, the European Bioinformatics Institute (www.ebi.ac.uk/Information/About_EBI/about_ebi.html) provides systems biologists with databases and tools for their research.

In the UK humanities, archaeologists can look to the Archaeology Data Service database (<http://ads.ahds.ac.uk>) for digital resources and Archaeotools (<http://wit.shef.ac.uk/archaeotools>) to assist in analysis of that data.

Numerical models

Numerical models are integral to physics, engineering, earth systems, chemistry, and materials

science. Numerous projects using computer-enabled methods show the promise of e-Science in these fields. In engineering, civil engineers use these methods to create 3D simulations and analyses of building structures—for instance, to assess a building’s performance during an earthquake (see NEESGrid, <http://neesgrid.ncsa.uiuc.edu>) or a fire (FireGrid, www.firegrid.org; see Figure 2). In many cases, simulation allows exploration of phenomena that would be infeasible to explore directly. For example, there’s a limit to the variety of buildings we can burn down while measuring their internal conditions and structural response.



Figure 2. The BRE Center for Fire Safety Engineering at the University of Edinburgh and other partners in the FireGrid project conducted a series of large-scale fire tests in a high-rise building in Dalarnock, Glasgow. (figure courtesy of Rochan Upadhyay, FireGrid Project Team)

In earth systems science, the GENIE (Grid Enabled Integrated Earth System Model) Project (www.genie.ac.uk) develops climate simulations to make predictions and help determine whether the climate change we’re witnessing now is more dramatic than what has occurred on Earth before. The eMinerals project (www.eminerals.org) involves research into mineralogical processes at an atomistic level and addresses questions regarding the transport and immobilization of contaminants. For example, the project has focused on nuclear waste and its disposal: can nuclear waste be contained in ceramic materials to prevent radioactive decay products leaking into water systems for enough centuries? It seeks to determine the best materials to immobilize high-radiation waste. The NanoCMOS Project (www.nanocmos.ac.uk) simulates the behavior of integrated circuits. As individual transistors in CMOS (complementary metal-oxide semiconductor) devices become smaller, it becomes increasingly difficult to control the variability between transistors because the nature of matter and charge impose limitations. Simulations are performed at the level of individual transistors, circuits, and whole systems.

Statistical models

Biology, medicine, the social sciences, and economics generally use statistical models. Some projects in these areas include KidneyGrid (www.gridbus.org/reports/KidneyGridPortal.pdf), eDiaMoND (the Diagnostic Mammography National Database Project, www.ediamond.ox.ac.uk), EMBRACE (European Model for Bioinformatics Research and Community Education, www.embracegrid.org/page.php?page=home), Avian Alert (www.avianalert.eu), and MoSeS (Modeling and Simulation for e-Social Science, www.comp.leeds.ac.uk/moses).

KidneyGrid supports the development of kidney models to better understand the structure and function of kidneys to tackle various medical problems. Another medical example is eDiaMoND, which provides medical professionals with easy access to an extensive database of mammogram images from women throughout the UK. Pooling and sharing this information will help in the early detection of breast cancer and its treatment.

To solve large biological dataset analysis problems using grid computing, the EMBRACE Network of Excellence (EU FP6 NoE) simplifies and standardizes the presentation of biological information to researchers. The project integrates gene and protein analysis data (for example, genomics, proteomics, and phylogeny) with the major databases and software tools used in bioinformatics in the Grid using the EGEE infrastructure and its middleware.

Avian Alert is a European Space Agency Integrated Application Program Initiative that follows bird movement by integrating a variety of tracking systems (satellites, radar, and so on). This integration assists a broad range of users in the flight safety, human health, migration ecology, conservation, and education communities. It can answer a number of research questions, such as how to reduce the likelihood of aircraft accidents caused by collisions with birds (see Figure 3) and how to map and understand bird migrations.



Figure 3. Hudson River landing, 15 January 2009 (see Wikipedia at [http://en.wikipedia.org/wiki/File:Plane_crash_into_Hudson_River_\(crop\).jpg](http://en.wikipedia.org/wiki/File:Plane_crash_into_Hudson_River_(crop).jpg)).

MoSeS provides an example from within the social sciences. This University of Leeds grid computing project has used anonymized 2001 census data to create a population model. This model represents the entire UK population, but can be considered at various levels—individual, household, or city. Researchers can use the model to explore future demography and test possible consequences of policy decisions such as transport and hospital planning on the predicted populations.

Epistemology and provenance

Epistemology and provenance predominate in the arts, languages, and humanities. We see exciting e-Research possibilities in these disciplines as well. For instance, the AMUC Project (Associated Motion Capture User Categories, at www.arts-humanities.net/image/amuc_project) has designed a motion-capture database deployed across the grid. It offers a data-mining opportunity to performing artists, such as musicians and conductors, dancers and choreographers, the martial arts community, jugglers, and magicians. These performers can map their movements and refine routines to improve their choreography and body movements (see Figure 4).

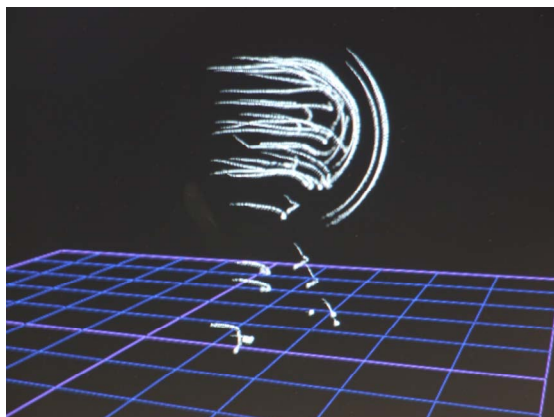


Figure 4. Motion capture trace of dancer Gretchen Schiller, AMUC project. (figure courtesy of Dave Green, Culture Lab, Newcastle University)

Historians can benefit enormously from the widespread availability of digitized online texts and documents. One such example is Virtual Vellum (www.shef.ac.uk/hri/projectpages/virtualvellum.html), the Jean Froissart Project. In the 15th century, Froissart wrote *Chroniques*, a manuscript considered a key historical source on the 100 Years War between England and France. This Engineering and Physical Sciences Research Council project uses Virtual Vellum, an image-viewing tool (particularly targeted at arts and humanities researchers), to create a digital “surrogate” of the original manuscript to allow wider access to it. The actual manuscript is in danger of disintegration if viewed and handled too frequently: vellum, animal skin used as paper, needs a constant temperature and humidity because of its fragility. Too much humidity causes it to mold, and it becomes brittle under dry conditions. Historians have better access to this primary source through digitization. (See also collections such as digitized Egyptian and Greek papyri at the University of Michigan, www.lib.umich.edu/pap, and the epigraphical archive available through the Center for the Study of Ancient Documents at Oxford, www.csad.ox.ac.uk/index.html.) Computer-aided visualization of archaeological sites provides archaeologists with virtual environments that bring the details of their research to life—for instance, allowing 3D reconstruction of ancient cities or caves (see 3DVisA, <http://3dvisa.cch.kcl.ac.uk>). The VERA project (Virtual Environments for Research in Archaeology, www.jisc.ac.uk/whatwedo/programmes/vre2/vera.aspx) provides archaeologists with numerous tools to use onsite during excavations and later in analysis and further research. Digitized records from the Silchester excavation, the site of a Roman settlement near Reading, UK, feed into an expansive database, allowing for better management of and access to detailed information relating to the town’s history from its beginnings in Roman times.



Figure 5. Silchester excavation site. (figure courtesy of Mark Baker, ACET, University of Reading)

The Importance of Collaborative Effort to Meet the Challenges

Appropriate education is required to equip our graduates for the rapid evolution of pervasive computational, data, and digital systems. The rate of change in the contemporary work, social, governmental, and research environments has generated challenging conditions for educators. Pioneering effort is needed to recognize the nature of the new opportunities and the skills and understanding future citizens will need. We’ve illustrated the diversity and scale of this challenge and recommended strategic collaboration at national, regional, institutional, and personal levels to accelerate and improve the educational system’s response. Many educators have deep insights into aspects of the ways in which education must change. We call for collaboration and pooling of ideas to stimulate the development of an integrated conceptual approach to this great challenge and opportunity.

Acknowledgments

Thanks to Donal Fellows (University of Manchester), Geoffrey Fox (University of Indiana), and John Boisseau (University of Texas) for their comments and suggestions regarding this article.

Malcolm Atkinson is director of the e-Science Institute and e-science envoy at the National e-Science Centre, Edinburgh. Contact him at mpa@nesc.ac.uk.

Elizabeth Vander Meer is the education and training policy officer at the National e-Science Centre, Edinburgh. Contact her at evmeer@nesc.ac.uk.

David Fergusson is deputy director of training, outreach, and education at the National e-Science Centre, Edinburgh. Contact him at dfmac@nesc.ac.uk.

Clive Davenhall is senior research engineer at the National e-Science Center, Edinburgh. Contact him at acd@staffmail.ed.ac.uk.

Hamza Mehammed is a senior trainer at the National e-Science Center, Edinburgh. Contact him at hamza.mehammed@nesc.ac.uk.