

Enhancing and Scaling-up Design-based Research: The Potential of E-Research

Lina Markauskaite, Peter Reimann, Centre for Research on Computer-supported learning and Cognition (CoCo), University of Sydney, Sydney, Australia
Email: l.markauskaite@usyd.edu.au, p.reimann@usyd.edu.au

Abstract: This paper identifies ways in which conceptual, methodological and technical developments in e-research can contribute to solutions of key questions in learning science research and, in particular, design-based research (DBR). The paper focuses on DBR issues in three major areas: methodology, research process, and dissemination. By mapping DBR issues to the conceptual and technological features of e-research, and illustrating those features with concrete examples from a range of research domains, the paper demonstrates how e-research approaches and tools could enhance present DBR practices as well as open avenues for new research questions and new ways of doing research. It concludes by discussing potential challenges and outlining some critical elements for the uptake of e-research in learning science.

Introduction

Design-Based research (DBR) is characterized as an inter-disciplinary “mixed-method” research approach conducted “in the field” and serves applied as well as theory-building purposes. This inquiry approach is very promising and also very challenging for (a) learning sciences as a discipline, because of unresolved methodological issues; (b) researchers working within this epistemic framework, because of practical and technical DBR complexities; and (c) users of DBR outputs, because of numerous dissemination challenges. For these reasons, methodological progress in DBR is relevant across a range of methods practiced in the learning sciences and educational research more generally. Building on the methodological groundwork laid in recent years (e.g., Barab & Squire, 2004; Kelly, 2003; Sandoval & Bell, 2004), our contribution sets out to demonstrate how DBR could profit from adapting and further developing e-research conceptual, methodological and technological approaches, and tools.

The first section of this paper provides a retrospective review of e-research with particular emphasis on those conceptual aspects and technological developments that provide the basis for the enhancement of DBR. The next section reviews methodological and practical developments in DBR and summarizes the key aspects and challenges. It emphasizes three broad areas of issues: (a) methodology, design and data gathering; (b) research process and data analysis; and (c) dissemination. The third section maps the main DBR issues to the conceptual and technological features of e-research, and the paper concludes by outlining some future steps that are critical for enabling a broader application of e-research approaches and technologies in DBR and other areas of learning research.

E-Research: Concepts, Vision, and Technologies

Concepts and Vision

Advanced information and computation technologies and high-speed computer networks have opened the door for new research opportunities. Reflecting on these developments and large scale efforts to promote new research practices, O'Brien (2005) writes, “Whether it’s *e-research* in Australia, *cyberinfrastructure* in the United States, *the grid* in Europe, or *e-science* in the United Kingdom, a transformation is clearly occurring in research practice” (p. 65). While exact definitions of these terms vary across countries, the notion of e-research is typically based on three major developments: (a) sharing of computer power and physical technological resources; (b) distributed access to large federated datasets; and (c) the use of virtual research platforms for collaborative research and communication (Wouters, 2005). The key vision of e-research is that many aspects of scholarly research practice can be improved, enhanced, and transformed by access to distributed data, information processing capabilities, and software tools to operate and manipulate these resources, the opportunities to synthesize, organize and disseminate data and new knowledge effectively, and the possibilities of bringing together and embracing pooled human effort (DEST, 2006; O'Brien, 2005; Wouters, 2005).

E-research is primarily perceived as enhancing existing theoretical research traditions rather than a new theoretically coherent research tradition. E-research capabilities, however, provide more than just a set of powerful research tools and techniques that could improve the productivity and quality of existing research practices (Anderson & Kamuka, 2003). The merits arising from access to shared data, shared technological resources, and collaboration provide opportunities to ask and answer new research questions (DEST, 2006). These include opportunities to work collaboratively on big picture problems and on much larger scales, to use

new data discovery, modeling and visualization methods that primarily rely on large computational capacities, to use new inter-disciplinary and cross-disciplinary organizational models of scientific inquiry, and, thereby, create new knowledge in new ways (Dutton, 2007).

The Technology: Grid Computing

The potential of e-research capacities is essentially embedded in the philosophy, design principles and functions of the underlying network and service architecture – called “community grid” (Berman, Fox & Hey, 2003). As defined by the Grid pioneers, the Grid is a software infrastructure that enables “flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources” (Foster, Kesselman & Tuecke, 2001, p. 200). The core of the Grid architecture is a layered structure formed of various interoperable protocols (see Berman *et al.*, 2003; Foster *et al.*, 2001 and Figure 1).

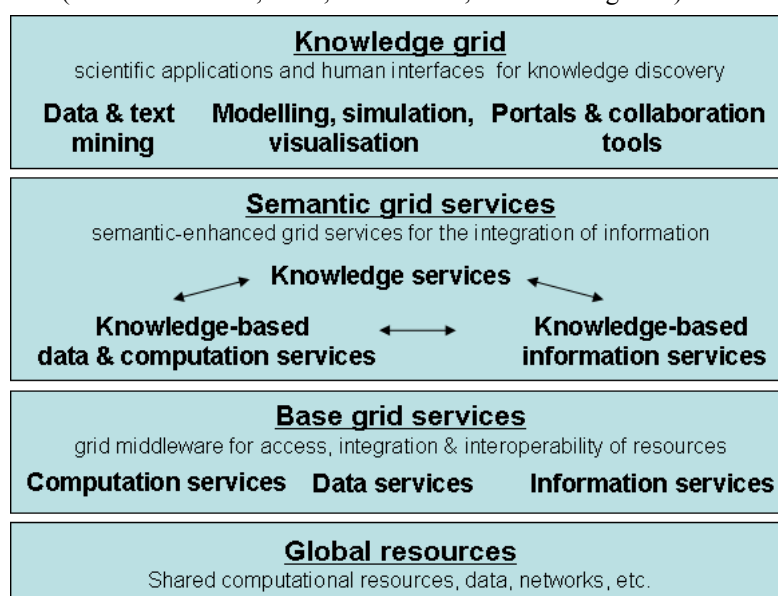


Figure 1. The main layers of the Grid.

The evolution of grids started in the early 1990s (De Roure, Baker, Jennings & Shadbolt, 2003). The first generation systems primarily focused on proprietary technological solutions for sharing high-performance computing resources at the bottom layer of the Grid model. The objective was to provide computational resources to a range of high-performance applications, primary for scientific discoveries in life sciences. The second-generation systems focused on heterogeneity, scalability, and adaptability of computational resources and data. They introduced a layer of base Grid service software tools – called “middleware” – that provides users and applications with homogenous standardized interfaces. This Grid middleware layer has provided the means for interoperability and the integration of heterogeneous resources, and a significantly enhanced uptake of e-research in new research areas, including the humanities and the arts, which have started to compile large data archives. The emerging third generation systems have focused on the reuse of existing components and information resources, and the possibility of assembling these components in a flexible manner. At this stage, the Grid has started to evolve not only technologically, but has also been reconsidered organizationally. The emphasis has shifted from solely powerful and integrated technological tools, to an evolving infrastructure for transient multi-institutional “virtual organizations” (De Roure, Jennings & Shadbolt, 2005). Dynamic and coordinated resource sharing and problem solving are the key features of the third-generation Grid. As De Roure *et al.* (2005) put it, this sharing “is not primarily file exchange, but rather direct access to computers, software, data, and other resources, as is required by a range of collaborative problem-solving and resource brokering strategies emerging in industry, science, and engineering” (p. 669). The key characteristics of the emerging Grid systems are service-oriented architecture and increasing attention to the “knowledge-based layer”. This knowledge-based layer includes such aspects as resource description using metadata and ontologies, annotation using provenance information, and process descriptions, such as workflows. These developments have made the Grid increasingly more relevant for scholarly inquiry and knowledge creation in social sciences, arts, and humanities.

Design-Based Research: Concepts and Challenges

“Design-based” or “design research” was proposed in the early 90s by Brown (1992) and Collins (1992) as an extension of other educational research methods. Since then, DBR has been used in learning science research in various forms, and in recent years it has been the subject of extensive methodological

discussions and reflections in special issues of educational journals, such as “Educational Researcher” (Kelly, 2003), “Journal of the Learning Sciences” (Barab & Squire, 2004) and “Educational Psychologist” (Sandoval & Bell, 2004).

DBR was developed to address several key issues central to learning research, including the need: (a) to address theories of learning; (b) to study learning in the real world; (c) to go beyond narrow measures of learning; and (d) to derive research findings from formative evaluations (Collins, 1992). Wang and Hannafin (2005) define DBR “as a systematic but flexible methodology aimed to improve educational practices through iterative analysis, design, development, and implementation, based on collaboration among researchers and practitioners in real-world settings, and leading to contextually-sensitive design principles and theories” (p. 6). They identify five DBR characteristics: (a) pragmatic, i.e. design-oriented and intervention-oriented; (b) grounded in theory and research; (c) interactive, iterative and flexible; (d) integrative; and (e) contextual. These characteristics and aims of DBR have led to a range of methodological and practical challenges, as numerous authors have identified (e.g., Barab & Squire, 2004; Collins, 1992; Dede, 2004; Fishman, Marx, Blumenfeld, Krajcik & Soloway, 2004; Kelly, 2004; Wang & Hannafin, 2005). These issues are mainly related to three broad aspects of DBR:

- *Methodology and research design*, to include such challenges as: (a) researchers’ involvement and potential “Hawthorne effect”; (b) reliability of research techniques, in particular qualitative ones; (c) variable control and replicability of experimental conditions; and (d) comparison of findings across studies and generalization.
- *Data gathering and analysis*, to include such issues as: (a) researchers’ continuous involvement in daily school operations, which is often seen as a distraction for schools; (b) complexities involved in capturing all variables affecting the success of a design; (c) difficulties with integrating different data sources, such as surveys, video observations and evidences of students’ work; and consequently (d) data deluge, which typically leads to the outcome that most of the collected data are not analyzed.
- *Dissemination and scalability of research findings*, to include such aspects as: (a) publishing context-sensitive mixed-method research outcomes, such as communicating classroom context, research design, materials and findings that depend on multimedia or video data; (b) sustainability of change process after the researchers have left the school; and (c) scaling-up DBR beyond the school level.

Additionally, the logic of the control group design in DBR is typically different from canonical experimental design, and DBR studies do not always provide “gold-standard” scientific evidence, which is typically expected by policymakers from educational research. As a result, DBR researchers face a range of challenges related to accountability and trustfulness.

E-Research Potential for Enhancing Design-Based Research

While e-research approaches cannot provide the entire solution to all the DBR challenges outlined above, they have the potential to help to address a number of the issues and limitations. The space does not allow us to consider and scrutinize all of these. We will confine ourselves to indicating the potential of several e-research components: (a) semi-automatic, continuous data acquisition and workflows, which offers some answers to methodological challenges related to continuous researcher involvement, data quality and reliability and replicability of study; (b) cooperative, automated data analysis, because it provides answers to some data handling, data deluge and reliability issues; and (c) new forms of research dissemination, because they address issues of publishing “data rich” research, scaling-up, generalization and accountability.

Acquiring Data and Capturing Context: Trails, Workflows and Provenance

As learning and communication technology becomes increasingly ubiquitous and immersive (viz. the rise of mobile and ambient computing, virtual social and learning environments), a considerable part of learning-relevant performance and interaction takes place online, and can be automatically captured in a digital form to be an essential source of data for performance monitoring and scientific research. Learning science studies typically employ various computer-supported learning methods, and there are rich possibilities for capturing evidence about students’ interactions with ICT and learning progress automatically (e.g., students’ interaction with a learning management system, progress on writing and creating various artifacts online). The nature of student-system interactions and information that can be captured automatically have been analyzed in studies of online learning trails (Schooneboom, Levene, Heller, Keenoy & Turcsanyi-Szabo, 2007). Automatically captured and dynamically analyzed data can also be used to adjust computer-supported instruction (Koedinger & Corbett, 2006).

In DBR, however, the major practical challenge is not only to capture students’ interactions and online data, but to capture a full range of data during the entire study lifecycle and ensure integrity and interoperability of these data. In most empirical learning science research, ICT is not fully exploited in the data acquisition process. E-research offers a number of enhanced strategies to acquire, store, and process performance data. For instance, e-research provides the means to not only deal with static data stored in files and databases, but also

with “streams” of synchronous data, coming from instruments, cameras, computers, and collaboration platforms (De Roure *et al.*, 2005). Technologically it is only a small step to capture and store much of the interaction data as well as other data produced by learners online (e.g., discussion texts) and make them accessible for analysis via federated and integrated databases.

In addition, specialized workflow tools have been developed to plan, and document e-science front-to-end experimental procedures, such as *Taverna* (Oinn *et al.*, 2006). Not only can workflow logic increase the integrity of data and reliability of studies, it can also take advantage of insights from other research fields, and, in general, contribute to shared knowledge of research processes. A nice example of the value of producing and sharing research workflows is *myExperiment* project (<http://myexperiment.org/>), where scientists, mainly from the life sciences, share experimental and analysis procedures in a common workflow language. “Our vision is that scientists should be able to swap workflows and other scientific objects as easily as citizens can share documents, photos and videos on the Web.” (De Roure & Frey, 2007, p. 1).

In general, traces and recordings are not “data”. What constitutes a “datum” depends on the purpose of the analysis and the methodological orientation of the person performing the analysis. Not only can data have very different formats (e.g., a test score, a snippet from a video), data can also be distinguished by how *context-bound* they are. One of the key characteristics of DBR is that context matters; applied design work involves situating this work in “naturalistic contexts”. However, as Barab and Squire (2004) argue, “the boundaries of context and what constitutes naturalistic may prove elusive” (p. 11). In order to clarify what context means in design studies, Collins *et al.* (2004) suggest a number of levels on which to perform observations should be distinguished: cognitive, interpersonal, group or classroom, resource, and institutional or school level. Also, they suggest that at least three types of dependent variables be considered: (a) climate (e.g., engagement); (b) learning (e.g., knowledge, skill, strategies); (c) systemic (e.g., sustainability, spread, costs). They also suggest the following types of contextual variables, from which the researchers need to select their independent variables: setting, nature of learners, required resources and support for implementation, professional development, financial requirements and implementation path. An additional important dimension of context, often ignored in learning research, including DBR, is time (Abbott, 1990). Learning and change processes are path-dependent: how learners, groups or organizations behave at any point in time does not only depend on the current context, but also on the history of the system under study. In the learning sciences, we are currently far from making adequate use of the fact that we can keep track of interaction and communication processes automatically (Reimann, 2007).

In order to capitalize on e-research methods and make data interoperable, it is not only necessary to decide which learning and context variables to capture, but also to store and describe these data using common formats and notions (i.e., standard metadata and ontologies). To make research process transparent, data acquisition process, transformation and analysis should be documented and described as well. The research process documentation – called “provenance” in e-research – is of particular importance in DBR.

Groth *et al.* (2006) define the provenance of a piece of data as “...the process that led to that piece of data”, i.e. data lineage, pedigree or history. Provenance relates to the use of data both within a study (e.g., How was the data created? How did it enter into the conclusions?) and across studies (e.g., Which other studies refer to this data?). In this way, provenance can be used to assess data quality and reliability, to establish ownership of data, to establish an interpretation context, or to replicate a study (Chorley, Edwards, Preece & Farrington, 2007). A number of provenance architectures have been developed in science e-research projects, for instance, *CombeChem* (Taylor, Essex, Frey, Mills, Hughes & Zaluska, 2006) and *MyGrid* (Stevens, Robinson & Goble, 2003). These tools track the experiment and analysis process and automatically create provenance record as well as allow researchers to annotate key research steps using controlled vocabulary or narrative descriptions. Similar architectures have been developed for social research, for instance *PolicyGrid* (Philip, Chorley, Farrington & Edwards, 2007).

In learning science research the technological architectures must enable the capture of not only how data were created, but also the context in which data were originated. For DBR, two key aspects of the provenance are important: (a) pedagogical or learning design provenance, related to the design of the pedagogical intervention; and (b) methodological or study design provenance, related to the accompanying study, data acquisition (observations, tests, etc.) and analysis. While structures and ontologies for representing science experiments can be adapted for methodology provenance (e.g., Taylor *et al.*, 2006), new structures and ontologies need to be created to represent pedagogical designs. Design patterns (Goodyear, 2005), learning designs (Koper, 2005) and collaboration scripts (Harrer, Hever & Ziebarth, 2007) could provide a foundation for this work.

Co-operative, Distributed and Semi-automated Data Analysis

ICT-enhanced data analysis has already received some attention and made a substantial progress in social science data and more specifically in learning science research. For example, standard social science software tools for statistical analysis (e.g., *SPSS*, *R*) and qualitative data analysis (e.g., *Nvivo*, *Squanto*) are used

in learning science research. The analysis of video data is frequently conducted with such tools as *Transana* or *Digital Replay System* (Goldman, Pea, Barron & Sharon, 2007) and also special software, such as *(Web)Diver*, that have been developed for the particular types of analysis required for learning research, (Pea, 2005).

A good example of the state of the art of co-operative analysis tools of learning science data is the evolving platform for managing and analyzing CSCL interaction data and learning outcomes described by Law et al. (2007). Starting from the observation that while CSCL researchers often perform similar kinds of analysis (e.g., content analysis by applying a coding scheme) on similar kinds of data, the authors propose an integrated toolkit composed from a set of components: (a) data preparation component; (b) discourse segmentation component; (c) coding scheme and coding rules' editor; (d) participation and interaction analysis component, comprising tools for social network analysis; (e) text analysis component, comprising text mining algorithms; (f) coding support component; and (g) learning component that would apply machine learning algorithms to continuously improve the match between coding categories, discourse elements, etc. The main advantages of such platforms are twofold. First, tools that are needed for the entire data analysis process – from initial access to online data to various, partially interdependent, stages of data analysis – are interconnected, and researchers can use all of them seamlessly. Secondly, the environment supports remote access and co-operation around shared datasets and data analysis methods (e.g., content coding). Both of these aspects lie at the heart of e-research.

It is the collaboration area where contributions of the Grid computing are increasingly adopted and can be adapted for social research. For instance the *Access Grid* (www.accessgrid.org) technology enables large format audio and video based collaboration between groups of people in different locations by offering a free and open video conferencing solution. The main additional contribution of the Grid computing community is the development of the middleware services and tools for managing remote resources – including files and programs, and also hardware, such as disk space, memory, and computational power – in fully distributed research environments. This is of particular importance for learning science, as such means open possibilities for developing and using new computation intensive data analysis and visualization algorithms for data-intensive cross-study analysis of large federated datasets.

It is not only data analysis, but also learning theory and model building that could be enhanced by e-research. Despite that fact that modeling is a widely used strategy to deal with complexity in sciences (Jacobson & Wilensky, 2006), modeling has neither been a research practice in the learning sciences nor in educational research and psychology (with the notable exception of work on cognitive architectures (Holland, Holyoak, Nisbett & Thagard, 1986; Newell, 1990)). Modeling, however, holds a huge potential for both research on individual (Reimann & Spada, 1995) and collaborative (Sun, 2006) learning. Computational modeling, as a research method, is hard to learn and hard to practice (Jacobson & Wilensky, 2006). This is likely one of the main reasons for the low acceptance of this method. E-research tools can support those aspects of the modeling cycle that are particularly time consuming and cumbersome. One example of such support is the *FEARLUS* system that takes over the burden involved in the sensitivity analysis of computational models (Pignotti, Edwards, Preece, Polhill & Gotts, 2005).

Research Dissemination

In the past decade the research dissemination landscape has been noticeably changed by three important technology-enabled developments. First, the emergence of domain-specific gateways and repositories of research publications improved significantly public access to traditional research outcomes. For instance, such publication databases as *Education Resources Information Center - ERIC* (<http://www.eric.ed.gov/>), and case study repositories as *What Works Clearinghouse - WWC* (<http://ies.ed.gov/ncee/wwc/>). Second, the appearance of integrated electronic journals and other full text online scholarly collections combined with the opportunity to access this scholarly content via powerful integrated search interfaces have significantly improved access to traditional scholarly outputs. See, for example, information portals such as *ProQuest* (<http://www.proquest.com/>). Third, public gateways for access to social research data, typically coming from large national surveys and censuses, has enabled further analysis of some valuable social databases. A typical example of such public data repositories is the *International Archive of Educational Data - IAED* (<http://www.icpsr.umich.edu/IAED/>). While these three developments have improved access to some aggregated research knowledge and data, this has been far from sufficient to solve major dissemination issues in the area of learning sciences.

Traditional text-based publishing remains a dominant form and practice of scholarly research dissemination. In education, as in other domains (Seringhaus & Gerstein, 2007), this canonical form and process of scholarly publications imposes sharp constraints upon the type and quantity of information that can be disseminated. Firstly, standard closed peer review processes are typically lengthy and do not always succeed in assuring high quality of scholarly output (Poschl, 2004). Secondly, traditional scholarly publishing does not meet the educational community's needs for access to more practically orientated research products, such as teaching materials. This is particularly unfavorable for applied learning science, where research typically

produces pedagogically and/or scientifically valuable research outputs (such as software designs and tools, supplementary learning materials, research instruments and data and evidence about students' performance) throughout the entire research cycle and in variety of media formats (e.g., text, graphics, video, audio).

E-research approaches and technologies provide further opportunities to make research dissemination more relevant to public and scholarly needs. The recent developments in this field have been progressing in several directions, including such aspects of dissemination as: (a) open access to public research knowledge and integration of scholarly publishing with data publishing; (b) open and interactive forms of peer-review process; (c) access to research knowledge for non-scholarly communities; (d) emergence of tools and interfaces for access and reuse research knowledge and data; (e) "publication at source" of research during complete chain of knowledge production; and (f) resulting shift from knowledge dissemination and use to knowledge co-construction. We provide only few illuminating examples of these developments.

A group of scholarly dissemination projects has been aiming to reduce the gap between scholarly publishing and deposition of associated data and other research outputs in public data repositories (Seringhaus & Gerstein, 2007). For example, it is established practice in computational biology that before the paper is published in the highly rated *Plos Computational Biology* journal (<http://compbiol.plosjournals.org/>), the associated data set on which findings are based is deposited to an open biology repository, such as *Worldwide Protein Data Bank – PTB* (<http://www.wwpdb.org/>). These projects primarily rely on interoperability of different data and e-print repositories enabled by technological and content-related standards.

Several projects aim to develop necessary infrastructures for capturing and sharing raw data produced during complete chain of knowledge creation - from experiments in the scientific laboratory through to scholarly published research outputs (Paterson, Lindsay, Monotti & Chin, 2007). For instance, *CombeChem* (<http://www.combechem.org/>), *eSciDoc* (<http://www.escidoc-project.de>) and *DART* (<http://dart.edu.au/>) projects. The latter project, for example, aims to enable researchers and reviewers to access original and analyzed data, collaborate around the creation of research outputs, stored publications, and add content, annotations and notes (Paterson *et al.*, 2007). It also aims to provide support for collection of large datasets, including remote control and automated data collection. These projects typically focus on the seamless integration and re-use of scientific knowledge and workflows to support collaborative knowledge creation. At the technological level, these environments are typically grounded in semantic grid approaches and service-oriented architectures. Some of these projects (e.g., *CombeChem*) also focus on the acquisition of "data at source" and automatic or semiautomatic annotation of data with rich contextual provenance. The main emphasis of "data at source", as the authors state, "is on not only the publication of scholarly output which is interlinked to the original data, but on capturing and digital record in all its forms and making this available for re-use" (De Roure & Frey, 2007, p. 5).

Taken together, the above ideas and technologies provide the means for a paradigm shift of scientific dissemination. Knowledge discovery and dissemination are no longer seen as separated from each other, but rather as two interlinked aspects of knowledge creation. Knowledge creators and knowledge users essentially become co-constructors of scholarly knowledge.

As an inquiry approach, DBR in the learning sciences has many inherited features typical for qualitative and quantitative educational research, as well as features of technology enabled scientific experimentation. Well documented, semi-automatically captured DBR processes and workflows could allow for better access to all DBR outputs and, consequently, provides for accountability, adaptability and scaling-up of DBR in new educational settings. The notions of collaboration between researchers and practitioners and iterative work on the learning design are at the heart of the DBR paradigm. Thus, collaborative research environments supporting seamless integration of various research processes and ongoing workflow from research design to outputs and then back to the design are of particular relevance for the educational research communities working in the DBR epistemic framework.

Discussion and Conclusions

It is a challenge to introduce the main ideas behind e-research in several pages without leaving the reader with the impression that e-research is just a set of sophisticated software and hardware technologies connected by the Internet and mainly relevant to sciences. The message we wanted to convey instead is that the e-research paradigm, although at an early stage of conceptual development, offers opportunities for the enhancement of current research practices in all domains and, most importantly, opens up avenues to ask new research questions, to collaborate on much larger scales and to create, use and reuse scientific knowledge in new ways.

Over the last decade we have observed that the field of e-research has been developing a range of new concepts, methodologies, protocols and technologies that could significantly enhance our opportunities to create new knowledge in new ways. Progressing through the first two stages of development, the Grid has created a solid basis of global resource and middleware services for further adoption of e-research in various domains. Many research communities in various disciplines have been rapidly embracing e-research into their epistemic

cultures and are actively working on those aspects of the Grid services and applications that need to be adapted to their disciplinary needs. The biggest success has been achieved by those research communities that worked collaboratively on a global scale developing shared notions, processes and resources (e.g., astronomy, computational biology).

Although we have seen that many e-research developments are very relevant for the research questions and challenges in learning science and in education more generally, e-research approaches have been embraced in educational research marginally and fragmentally only. As we have shown in this paper, some progress has been achieved in such areas as data analysis and research dissemination, but research practices have not changed fundamentally. If the full potential of e-research is to be realized, the international learning science research community needs to work collaboratively on shared notations for pedagogical and methodological designs, and for other aspects of context, in addition to developing shareable analysis tools. Among the pivotal issues for embracing e-research approaches in learning science and DBR are the development of appropriate data structures for storing and integrating different information types supplemented with meaningful metadata and provenance based on generic and domain-specific ontologies. This will provide a necessary foundation for further development of interoperable domain-specific data and workflow repositories, applications and interfaces for knowledge discovery and collaborative research and dissemination environments at the top knowledge grid level.

While there are many hurdles to be overcome to realize the vision of e-research work, some of those essential for learning science are not technological. The real challenges lie in questions such as: How does one deal with ethical and privacy issues, in particular when using video data? What are acceptable practices for sharing data and results of intermediate analysis? Each research community needs to work on standards, procedures, norms and other issues specific to it in order to make the e-research vision work for it.

References

- Abbott, A. (1990). Conceptions of time and events in social science methods: Causal and narrative approaches. *Historical Methods*, 23(140-150).
- Anderson, T., & Kamuka, H. (2003). *E-research: Methods, strategies and issues*. Boston: Pearson Education.
- Barab, S. A., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, 13(1), 1-14.
- Berman, F., Fox, G., & Hey, A. (2003). The Grid: Past, present, future. In F. Berman, G. Fox & A. Hey (Eds.), *Grid computing: Making the global infrastructure a reality* (pp. 9-50). West Sussex: Wiley.
- Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions. *The Journal of the Learning Sciences*, 2, 141-178.
- Chorley, A., Edwards, P., Preece, A., & Farrington, J. (2007, October). *Tools for tracing evidence in social science*. Paper presented at the Third international conference on e-social science, Ann Arbor, MI, US.
- Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology*. Berlin: Springer.
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: theoretical and methodological issues. *The Journal of the Learning Sciences*, 13(1), 15-42.
- De Roure, D., & Frey, J. (2007). *Three perspectives on collaborative knowledge acquisition in e-science*. Paper presented at the Workshop on semantic web for collaborative knowledge acquisition, Hyderabad, India.
- De Roure, D., Jennings, N. R., & Shadbolt, N. R. (2005). The semantic grid: Past, present, and future. *Proceedings of the IEEE*, 93(3), 669-681.
- Dede, C. (2004). If design-based research is the answer, what is the question? *The Journal of the Learning Sciences*, 13(1), 105-114.
- DEST. (2006). *An Australian e-research strategy and implementation framework: Final report of the e-research coordinating committee*. Commonwealth of Australia: Australian Government, DEST.
- Dutton, W. H. (2007, October). *Reconfiguring access to information and expertise in the social sciences: The social shaping and implications of cyberinfrastructure*. Paper presented at the Third international conference on e-social science, Ann Arbor, MI.
- Fishman, B., Marx, R. W., Blumenfeld, P., Krajcik, J., & Soloway, E. (2004). Creating a framework for research on systemic technology innovations. *The Journal of the Learning Sciences*, 13(1), 43-76.
- Foster, I., Kesselman, C., & Tuecke, S. (2001). The anatomy of the grid. Enabling scalable virtual organizations. *International Journal of Supercomputer Applications*, 15(3), 200-222.
- Goldman, R., Pea, R., Barron, B., & Sharon, S. D. (2007). *Video research in the learning sciences*. Mahwah, NJ.: Erlbaum.
- Goodyear, P. (2005). Environments for life-long learning: Ergonomics, architecture and the practice of educational technology. In J. M. Spector & T. M. Anderson (Eds.), *Integrated and holistic perspectives on learning, instruction, and technology: Understanding complexity* (pp. 1-18). Dordrecht: Kluwer.

- Groth, P., *et al.* (2006). *An architecture for provenance systems*. Retrieved 28-Oct-2007, from <http://eprints.ecs.soton.ac.uk/12023/>.
- Harrer, A., Hever, R., & Ziebarth, S. (2007). Empowering researchers to detect interaction patterns in e-collaboration. In R. Luckin, K. R. Koedinger & J. Greer (Eds.), *Artificial intelligence in education - Building technology rich learning contexts that work* (pp. 503-510). Amsterdam: IOS Press.
- Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.
- Jacobson, M. J., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *The Journal of the Learning Sciences*, 15(1), 11-34.
- Kelly, A. E. (2003). Theme issue: The role of design in educational research. *Educational Researcher*, 32(1).
- Kelly, A. E. (2004). Design research in education: Yes, but is it methodological? *The Journal of the Learning Sciences*, 13(1), 115-128.
- Koedinger, K. R., & Corbett, A. (2006). Cognitive tutors. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61-77). New York: Cambridge University Press.
- Koper, R. (2005). An introduction to learning design. In R. Koper & C. Tattersall (Eds.), *Learning design. A handbook on modelling and delivering networked education and training* (pp. 3-20). Berlin: Springer.
- Law, N., Yuen, J., Huan, R., Li, Y., & Pan, N. (2007). A learnable content and participation analysis toolkit for assessing CSCL learning outcomes and processes. In C. Chinn, G. Erkens & S. Puntambekar (Eds.), *Mice, minds, and society. Proceedings of the CSCL 2007 conference* (Vol. 8, pp. 408-417). New Brunswick, NJ: ISLS.
- Newell. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- O'Brien, L. (2005). E-research: An imperative for strengthening institutional partnerships. *Educause Review*, 40(6), 64-77.
- Oinn, T., *et al.* (2006). Taverna: Lessons in creating a workflow environment for the life sciences. *Concurrency and Computation: Practice and Experience*, 18(10), 1067-1100.
- Paterson, M., Lindsay, D., Monotti, A., & Chin, A. (2007). DART: A new missile in Australia's e-research strategy. *Online Information Review*, 31(2), 116-134.
- Pea, R. (2005). Video-as-data and digital video manipulation techniques for transforming learning science research, education, and other cultural practices. In J. Weiss, J. Nolan & P. Trifonas (Eds.), *International handbook of virtual learning environments*. Dordrecht: Kluwer.
- Philip, L., Chorley, A., Farrington, J., & Edwards, P. (2007, October). *Data provenance, evidence-based policy assessment, and e-social science*. Paper presented at the Third international conference on e-social science, Ann Arbor, MI.
- Pignotti, E., Edwards, P., Preece, A., Polhill, G., & Gotts, N. (2005). Semantic support for computational land-use modelling. In *Fifth IEEE international symposium on cluster computing and grid* (Vol. 2, pp. 840-847): IEEE Press.
- Poschl, U. (2004). Interactive journal concept for improved scientific publishing and quality assurance. *Learned Publishing*, 17, 105-113.
- Reimann, P. (2007). Time is precious: Why process analysis is essential for CSCL. In C. Chinn, G. Erkens & S. Puntambekar (Eds.), *Minds, minds, and society. Mice, minds, and society. Proceedings of the CSCL 2007 conference* (Vol. 8, pp. 598-607). New Brunswick, NJ: ISLS.
- Reimann, P., & Spada, H. (Eds.). (1995). *Learning in humans and machines: towards an interdisciplinary learning science*. Oxford: Pergamon.
- Sandoval, W. A., & Bell, P. (2004). Design-based research methods for studying learning in context: introduction. *Educational Psychologist*, 39(4), 199-201.
- Schooneboom, J., Levene, M., Heller, J., Keenoy, K., & Turcsanyi-Szabo, M. (Eds.). (2007). *Trails in education: Technologies that support navigational learning*. Rotterdam, Taipei: Sense Publishers.
- Seringhaus, M., & Gerstein, M. (2007). Publishing perishing? Towards tomorrow's information architecture. *BMC Bioinformatics*, 8(17), 1-5.
- Stevens, R., Robinson, A., & Goble, C. (2003). MyGrid: Personalised bioinformatics on the information grid. *Bioinformatics*, 19(1), 302-304.
- Sun, R. (Ed.). (2006). *Cognition and multi-agent interaction. From cognitive modelling to social simulation*. Cambridge: Cambridge University Press.
- Taylor, K., *et al.* (2006). The Semantic Grid and chemistry: Experiences with CombeChem. *Journal of Web Semantics*, 4(2), 84-101.
- Wang, F., & Hannafin, M. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology, Research and Development*, 53(4), 5-23.
- Wouters, P. (2005, June). *The Virtual knowledge studio for the humanities and social sciences*. Paper presented at the First international conference on e-social science, Manchester.