Computer Supported Cooperative Work (2013) 22:667–715
DOI 10.1007/s10606-012-9184-0

Supporting Scientific Collaboration: Methods, Tools and Concepts

Marina Jirotka[1], Charlotte P. Lee[2] & Gary M. Olson[3]
[1]*Department of Computer Science, University of Oxford, Wolfson Building, Parks Rd, Oxford OX1 3QD, UK (Phone: +44-01865-610613; Fax: +44-1865-273839; E-mail: Marina.Jirotka@cs.ox.ac.uk);* [2]*Department of Human Centered Design & Engineering, College of Engineering, University of Washington, 423 Sieg Hall, Box 352315 Seattle, WA 98195, USA (Phone: +1-206-5430902; Fax: +1-206-5438858; E-mail: cplee@uw.edu);* [3]*Department of Informatics, Bren School of Information and Computer Sciences, University of California, Irvine, DBH 5202, Irvine, CA 92697-3440, USA (Phone: +1-949-8240077; Fax: +1-949-8244056; E-mail: gary.olson@uci.edu)*

Abstract. This paper discusses the interrelationship between e-Science and CSCW in terms of key substantive, methodological and conceptual innovations made in both fields. In so doing, we hope to draw out the existing relationship between CSCW and e-Science research, and to map out some key future challenges where the two areas of research may become more closely aligned. In considering what may be required to draw the two more closely together, the paper focuses primarily on investigations that have been undertaken in two dedicated initiatives into e-Science, along with the key issues emerging from these studies.

Key words: cyber-infrastructure, e-Science, supporting scientific collaboration

1. Introduction

Over the last few decades, the development of advanced technological innovations in high performance computing, storage, communication and software systems have heralded an opportunity for seemingly revolutionary transformations to occur in scientific research settings. This has emerged at a time when conventional research practices are being questioned as to their adequacy to address the grand challenges of the twenty-first century. As a consequence, disciplinary and institutional structures are now increasingly under pressure to change in order to meet what are seen to be the new scientific opportunities.

In response to this, an area of research known in the UK as e-Science, in the US as Cyber-infrastructure (CI) and e-Infrastructure within Europe and elsewhere has emerged that proposes a solution in the form of large scale, collaborative and multidisciplinary research that, in turn, calls for the development of powerful computer-based research infrastructures. John Taylor, former Director General of Research Councils in the UK Office of Science and Technology, encapsulated the vision of e-Science in the following: “e-Science is about global collaboration in key areas of science and the next generation of infrastructure that will enable it” (quoted in Hey and Trefethen 2002: 1017). This vision asserts that the challenging scientific questions faced by researchers in the twenty first century require expertise from a number of different disciplines, resulting in the need to form large, cross-disciplinary collaborations. For example, the drive to understand and model complex biological systems such as the human heart requires collaboration between mathematical modellers, computer scientists and software and medical engineers, as well as scientists drawn from the life sciences disciplines of biology, biochemistry, physiology and medicine.[1] The proposed solution seems to resonate strongly with key issues in CSCW; that in order to support such research between participants disparate in both geographical location and disciplinary background, a technological infrastructure is required that addresses the demands of this new multi-disciplinary, global, large-scale, data intensive, collaborative work.

To date, there have been five special issues of journals, three of them of this journal that have been published dedicated to the socio-technical study of e-Science (Edwards et al. 2009; Jankowski 2007; Jirotka et al. 2006; Lee et al. 2010; Spencer et al. 2011). In addition, six book-length treatments of the issue have also been published (Hine 2006; Jankowski 2009; Nentwich 2003; Olson et al. 2008a; Parker et al. 2010; Dutton and Jeffreys 2010). However, despite such publications that focus on specific issues and cases in e-Science and Cyber-Infrastructure, in general, the discussion of the actual relationship between the programmes that were developing technologies and infrastructures to support scientific collaboration and the work produced in CSCW has been somewhat limited.

Nonetheless, CSCW has considerable experience of, and has produced a large body of work in, the design, development and embedding of distributed technologies to support collaborative work that is of relevance to e-Science and CI. In addition, there is a great deal of work on methods for analysing work practices within CSCW for informing design, along with a wealth of conceptual and analytic findings that may be relevant for the development of the e-Science and CI programmes. In this paper, we will consider some of the substantive, methodological and conceptual contributions from CSCW that may have substantial impact on e-Science and CI.

Similarly, there are distinctive challenges within e-Science in terms of the large-scale distributed nature of the activity, where participants are generating and drawing on ever-increasing amounts of data that is computationally driven, requiring access to large-scale resources. In addition, understanding the specific scientific skills requires significant expertise in terms of being able to draw out the experimental, interpretative and analytic practices that scientists draw upon in

[1] e-Science Pilot Project Integrative Biology (IB), funded by the Engineering and Physical Sciences Research Council (EPSRC)

their research work. This requirement must also be considered within the context of the emphasis being placed on the sharing of data, both institutionally and in practice, amongst scientists who may not either be accustomed to doing so, or may not see the need to do so. CSCW investigations into these areas require greater consideration from e-Science and CI design communities as they raise, if not significant new research challenges, new aspects for rethinking existing concepts and approaches. In this paper, we will also refer to some more recent studies of e-Science that make interesting contributions to CSCW research.

The relationship between e-Science and CSCW is a relatively nascent one. In considering what may be required to draw the two more closely together, this paper will focus primarily on investigations arising from two large, dedicated programmes of work into e-Science supported by funding councils in the UK and the US along with the key issues emerging from these investigations.

2. Background: the challenge of e-science

The UK was one of the first countries to implement a national strategy at government level in which a large, (£250 M) cross-research council programme was initiated in 2002 – 'the UK e-Science Programme'. Central to the vision of the programme was the proposal to create an infrastructure for scientific research; to implement the necessary linking of computational resources, or 'Grid' infrastructure as it later became known (Foster et al. 2001: Foster and Kasselman 2004; de Roure et al. 2003). e-Science was characterised as 'large-scale science' that would increasingly be undertaken through distributed global collaborations enabled by the next generation Internet. A key feature of these types of scientific collaborations is that they are seen to require access to very large data collections, very large scale computing resources and high performance visualisations made available to individual scientists and institutions (Hey and Trefethen 2002, 2003a, 2005).

Within the UK, the e-Science vision has increasingly permeated disciplines outside the natural and physical sciences such as, the social sciences and humanities, to the extent that the term 'e-Research' has now replaced e-Science, and resulted in the emergence of further funded programmes supporting research within Digital Humanities and e-Social Science.[2] e-Research extends the definition to include a set of computing infrastructures and applications designed to be embedded into academic settings for the purposes of supporting distributed, multidisciplinary research collaboration (c.f. Borgman 2007). [3]

[2] For example, in the UK, a National Programme, e-Social Science, funded by the Economic and Social Research Council, was set up in an attempt to harness the technologies developed under the e-Science Programme to enhance social scientific research (Halfpenny and Procter 2010).

[3] For the purposes of clarity, in this paper we will use the term e-Science for general reference to a range of related terms covering similar capabilities.

Similar programmes have emerged in the US. While some efforts at building 'collaboratories' occurred after Bill Wulf's early call for such organisational experiments (Wulf 1989, 1993), these efforts gained considerable momentum following the Atkins report (2003) that led to a programme of work in Cyber-infrastructure (CI) funded by the US National Science Foundation (NSF). The NSF Office of Cyber-infrastructure (OCI) was charged with creating infrastructure comprised of both technology and human expertise necessary to support scientific research processes and collaborations. This included an emphasis on new forms of scientific inquiry that could lead to hybrid or entirely new disciplines and forms of knowledge.

Overall, in the US, significant and substantial investments are being made in this area. The database of NSF's OCI, for example, shows 342 awards that total over $375 million USD since 2001. The OCI is only one area within NSF that funds e-Research projects with further funding coming from other directorates such as Computer & Information Science & Engineering (CISE). The US National Institutes of Health (NIH) is also investing in cyber-infrastructure efforts such as the Biomedical Informatics Research Network (BIRN) and a range of support for research on the 'Science of Team Science' (Stokols et al. 2008; Falk-Krzesinski et al. 2011).[4]

These programmes of research were initially driven largely by the technological vision of creating a Grid-enabled network on a global scale to support academic work (Smarr and Catlett 1992; De Roure et al. 2003) that have also been well documented in a series of books (cf. Berman et al. 2003; Borgman 2007; Dutton and Jeffreys 2010), articles (cf. Foster et al. 2001; Finholt 2003; Young 2008) and government reports (cf. Atkins et al. 2003, 2009; Berman and Brady 2005). However, as with many grand visions, tensions have emerged between the goals of the programme of work to create new forms of knowledge whilst also accelerating scientific knowledge production, and the experience and perceptions of those who are confronted with that vision, or who are trying to translate it into a workable reality (Jirotka et al. 2005; Vann and Bowker 2006).

2.1. Conceptions of collaboration

Collaboration lies at the heart of the e-Science vision and thus, we argue that the concepts and analytic perspectives of CSCW offer potential insights into whether, and in what ways, the vision may be achievable. The vision requires an understanding of how to support large-scale collaboration that is both geographically distributed and co-present. As will be discussed in more detail later in the paper, researchers may be working in groups within their own

[4] A group at NIH produced a report entitled *Collaboration and Team Science: A Field Guide* (Bennett et al. 2010), and a web site rich with resources is at teamscience.nih.gov. Falk-Krzesinski et al. (2011) conducted a concept-mapping study to attempt to give some structure to this emerging field.

institutions and/or across organisations in real time, performing distributed experiments or running complex algorithms over large data sets *in silico*, sharing scientific visualisations or interpreting texts or video materials. Developing the technology that both supports distributed experimentation for example, and the collaborative work that may be necessary to achieve the experimental outcome is of critical importance.

In e-Science, collaboration was perhaps initially conceived of as more a matter of understanding generic research work processes and implementing these in workflows that might be integrated into the technical infrastructure. Various technologies were proposed to support this scientific collaboration such as Virtual Research Environments (VREs) (Dovey 2010; Voss and Procter 2009) that were perceived more as an interface to the digital infrastructure, allowing easy access to data and services in an environment that was focussed on a particular research activity (Carusi and Reimer 2010). These VREs (and other technologies such as AccessGrid (Childers et al. 2000; Stevens 2008)) attempted to support collaboration through a variety of means, mediating communication through use of video, support for collaboration around documents and scientific visualisations, and support for meetings. Nonetheless, save some exceptions, this work did not draw extensively on research findings from CSCW studies. Apart from the engagement of some researchers from a socio-technical background who crossed the fields of CSCW and e-Science (see for example, Star and Ruhleder 1996; Bowker 2000; Finholt 2002; Olson et al. 2008b; Hartswood et al. 2003, 2005) or those who were investigating technologies to support scientific collaboration in different fields such as HCI or participatory design (Yeh et al. 2006; schraefel and Dix 2009; Mackay et al. 2002), the e-Science initiative proceeded initially without engaging deeply with CSCW research.

2.2. Challenges of data-sharing

Perhaps what was not originally anticipated by the programme of research in e-Science and CI was the need to develop understandings of the nature of interdisciplinary and multi-disciplinary work, an appreciation of the contexts in which scientific data is produced and re-used, and how data may be ethically and legally shared across disciplines, institutions and geographical boundaries. Many of these requirements surfaced with the engagement of CSCW researchers in the design, development and evaluation of e-Science prototypes and with social scientists studying the development and impact of the e-Science initiative.

As has often been pointed out (For example, Atkins et al. 2003; Hey and Trefethen 2003) data is the 'life-blood' of the scientific enterprise. A key feature and perceived advantage of the e-Science drive is the potential for developing a 'commons' of information, which may be defined as data that can be easily accessed, reused and shared in collaborations across both geographical and disciplinary boundaries. This very accessibility is seen to offer the conditions for new forms of science to emerge. A core part of the e-Science and CI programme has been concerned with providing researchers with greater access to the large and growing body of heterogeneous data created by and/or drawn from simulations, digital instruments, sensor nets and observatories. This has, in turn, spawned further research into the increased scale and value of data and the demand for semantic federation, active curation and long-term preservation of access. The imminent flood of data predicted from this initiative has come to be known as the 'data deluge' (Hey and Trefethen 2003b) and has led to the notion of 'data intensive scientific discovery' as a fourth paradigm (Hey et al. 2009) – a suggestion that data will drive the scientific theories of the future. However, to what extent and in what ways this particular vision will unfold is yet to be determined.

Researchers investigating the practical work that is involved in sharing or re-using data have highlighted the problematic nature of viewing data as a 'commodity' that may be transferable independent of its vehicle (cf. Berg and Goorman 1999). It has been suggested that what is needed is an understanding of the contextual nature of data with an emphasis on data sharing (Birnholtz and Bietz 2003; Jirotka et al. 2005; Zimmerman 2008; Zimmerman et al. 2009). Taking this perspective requires a re-orientation of enquiry to understanding the ways in which data is entangled with obligations from the different domains and communities of practice (Hartswood et al. 2005).

These tensions between the vision of e-Science and the practice of research have given rise to a series of investigations (cf Wouters and Bealieu 2006; Zimmerman 2008; Zimmerman et al. 2009; Olson et al. 2008a; Faniel and Zimmerman 2011; Dutton and Jeffreys 2010) stressing the problematic nature of data sharing and reuse from different communities of practice – from identifying the institutional arrangements that may mitigate against this vision (David and Spence 2010; David 2006; Welsh et al. 2006), to the ethical issues involved and the conflicting requirements placed upon researchers by different institutional concerns (Carusi and Jirotka 2009) to complex legal challenges such as ownership of images (Piper and Vaver 2010; D'Agostino etc al 2008) or issues of authorship (Pila 2009). Faniel and Jacobsen (2010) note that before data generated by others can be confidently reused, scientists need to assess the data's relevance, such that it may be understood and that the source may be trusted. Coopmans (2006) suggests that sharing data is essentially an exercise in making data mobile. Whilst part of this exercise requires developing standards (Bos et al. 2007), paradoxically, Hartswood et al. (2012) suggest that keeping the link to the originating context is important as that context can be recalled in various critical ways when re-used in different settings. Edwards et al. (2011), describe the problems that emerge when scientists from two or more disciplines work together on related problems, as 'science friction'. These authors suggest that it may be more useful to focus on the role of metadata in the more ephemeral process of scientific communication instead of as a permanent product for scientific work.

A potential implication of such studies suggests that the vision of e-Science tools and infrastructures may often need to be re-calibrated in the light of the organisational and disciplinary practices of researchers (Jirotka et al. 2005). The success of these initiatives relies fundamentally upon collaboration and the tools and technologies developed to support that collaboration within local and across global communities of researchers. Whilst the e-Science vision may reflect potentially unique features such as, the ability to share vast amounts of data through e-infrastructures, or the possibility for real-time monitoring of global networks of streaming data, or the sharing of three dimensional scientific visualisations, many important insights have been gained from research in CSCW regarding the support of both co-present and distributed collaboration in and through various technologies that have implications for the further development of e-Science tools and technologies. (Welsh et al. 2006). However, the implications of studies in CSCW for e-Science development are yet to be debated in detail and we suggest that CSCW as a "pioneering field for experimenting with novel intersections of the social, computational, information and design science and studies" (Ribes and Lee 2010) is the ideal forum for such debates to take place.

To begin such a discussion, this paper will focus on a set of studies undertaken in both e-Science and CSCW that highlight the potential commonalities and differences in tools, methods and concepts. We hope to draw out the existing relationship between CSCW and e-Science research and to map out some key future research challenges where the two areas may become more closely aligned.

3. Understanding collaborative research in scientific practice

Although research in e-Science has focussed on developing innovative tools, technologies and infrastructures to support scientific collaboration, a number of researchers have also attempted to apply systematic approaches to understanding the scientific work and practices that applications are to support, and the context in which technologies are to be embedded. With a concern to inform technological design, common challenges have arisen that resonate strongly with issues in CSCW. For these researchers, the development of e-Science technologies relies upon gaining an in depth understanding of the skills and practices of scientific work, how experiments are undertaken, and how tools, technologies and often quite ordinary artefacts are used routinely in scientific work in scientists' interactions and collaborations.

In the long tradition of workplace studies in CSCW that largely draws on ethnographic fieldwork (Middleton and Engeström 1998; Luff et al. 2000; Randall et al. 2007), researchers have studied scientists as they go about their work, analysing the details of their skills and interactions, and using these analyses to inform the development of systems to support collaboration in the research process. For example, in the domain of e-Health researchers such as (Hartswood et al. 2003; Jirotka et al. 2005) drew on ethnomethodological analyses of the details of radiologists' and radiographers' work practices to inform the design of an early UK e-Science flagship project to support breast screening in the UK (Brady et al. 2004). This research revealed the importance of the sociality of the clinics for the screening process. In addition, the analysis had strong implications for the initial technological vision that stressed the largely unproblematic nature of the sharing of data and collaboration in e-Science domains both within and across institutions.

This type of work called for a shift from a focus of technical design and development in e-Science to considering how technologies and infrastructures will be embedded into the everyday working lives of researchers. The shift from laboratory-based to digital science is a significant social development for scientific research communities and institutions, primarily because their working practices and collaboration activities will change as both are augmented with, or entirely transferred into, digital research environments.

Similar exercises drawing upon ethnomethodological analyses to inform design have been undertaken in studies that attempted to reveal the interpretative work of humanities scholars in deciphering ancient manuscripts and the importance of embodied interactions for the scholars' collaboration over these artefacts. These studies, for example, informed the development of a Virtual Research Environment for the Humanities (de la Flor et al. 2010a, b) and a novel approach to the systematic evaluation of interventions (Eden 2011). Further studies involving ethnographic fieldwork and participatory design have informed the development of tools and techniques to support scientists' data gathering activities – particularly in the field. These have focussed primarily upon mobile capture and access, integrating paper notes with digital data captured during field research and supporting the sharing of this data more widely. (Yeh et al. 2006; Warwick et al. 2009).

3.1. Methodological challenges

The involvement of users and domain experts in the e-Science development process has perhaps placed a different set of demands on research than in other participatory design and workplace studies. Whilst many early e-Science approaches attempted to understand the process of research as a workflow, those doing more detailed workplace studies have had to engage with and reveal, diverse sets of skills and expertise as they move across domains and disciplinary practices or indeed study increasingly multi- and interdisciplinary research teams forming new collaborations. This has presented different challenges to fieldwork, not only in the necessity to understand these varied disciplinary contexts and practices, but also of engaging with what may be, perhaps, a minority of scientists in the world, heart modellers or ancient classicists, who have the specific skills and expertise required to bring to bear in the design and embedding of these new collaborative systems.

However, it has been argued that further significant methodological challenges of studying e-Science in action lie in the heterogeneity of the types of researchers, research settings, materials, technologies and institutions involved. Hine (2007a) has argued that certain strategic adaptations need to be effected for ethnography to be able to deal with the spatial complexity of e-Science and the distributed nature of work through advanced ICT. Researching into the biological discipline of systematics and focussing on how the use of a variety of information and communication technologies has become a routine part of disciplinary practice, Hine proposes a connective approach to ethnography where researchers investigate the different ways online activities inform and shape offline work, and how online and offline activities may mutually inform each other. Given the spatial complexity of e-Science domains and the relationship between online and offline activities, Hine (2007b) argues that the relevancies of a discipline may be overlooked if fieldwork is undertaken at single sites. Multi-sited ethnography, she suggests, has the potential to address this concern. However, more specific challenges have been identified with such an approach (Wogan 2004). Whilst there may be dilemmas regarding which locations to observe where not all sites may receive the same level of engagement, Wogan argues that there is also a danger that the most technologically advanced site may set the standard for the types of systems to be developed across all sites. It seems that developing an understanding of the whole system may be compromised through a tendency towards comparative analysis of individual locations. These challenges resonate with work discussing multi-sited ethnography in human-computer interaction (see for example Lindtner et al. 2011; Williams et al. 2008), but perhaps may be more apparent in e-Science settings. These issues are critical to address if the social scientific approaches are to assists in developing the next generation of infrastructure for scientific research. Within CSCW, Randell et al. (2011) have highlighted the potential for multi-site workplace studies to contribute to CSCW, and has called for more investigations into this approach in order to detail commonalities and differences between settings as well as specific analytic contributions (Dourish 2007).

A novel design elicitation method that may seek to address such concerns explores work practices through the use of analogy (schraefel and Dix 2009). Whilst participatory design processes have been successfully employed in scientific settings to create augmented laboratory notebooks (Mackay et al. 2002), the use of analogy to bridge the gap between design team knowledge and domain expertise may complement these techniques in settings where the designers do not share domain or artefact knowledge with design-domain experts, and where the processes may be ‘semi-structured’ in order to help software developers to better understand the laboratory based activities of these scientists (schraefel and Dix 2009). In the synthetic chemistry domain, designers asked chemists to ‘make tea’ as if it were a chemical experiment in order to understand the processes that then informed the design of digital laboratory book prototypes. This work was extended into the domain of bioinformatics, moving from the 'wet' *in virto* lab of synthetic chemistry to the dry *in silico* lab of bioinformatics. In this domain, a jigsaw puzzle analogy was employed to understand the activities and processes and to develop a model of scientists' interactions with data as it moved, in the process of their experiments, from the Web to the desktop, providing scientists with resources to meaningfully interrogate the process and share their results. Whilst this approach seemed to yield recommendations for design of the digital laboratory books in the two cases, it would be most useful for CSCW research to explore more examples of this approach in design elicitation to ascertain how analogies may be developed in different settings, how varied they are and when they do, and do not work in the development process.

The development of e-Science applications has also brought other opportunities for embedding technologies. As users and domain experts participate in the development of bespoke applications, sometimes, contrary to the notions of user's resistance to new technology, scientists may be quite enthusiastic to try out new prototypes, particularly where the value to their research activities and processes is clear (de la Flor et al. 2010a). An obvious advantage in e-Science settings is that such technologies may be evaluated within the context of scholars' work. Rather than setting up experimental or even quasi-naturalistic evaluations, researchers can evaluate technologies in the course of their daily work. (de la Flor et al. 2010b).

However, whilst is seems essential to develop such innovative methods and approaches in order to understand scientific work practices and expertise, there may often be practical problems of coordination when introducing prototypes or interventions to elicit requirements in the development of bespoke systems to support scientific collaboration. Whilst the goal is to develop long-term infrastructures in e-Science, it is often the case that there may be a technology that could be relevant to a scientific activity, but designing and evaluating it in the course of a real scientific experiment may be problematic; it may be a challenge to synchronise the research cycle and the technological development cycle of the project as designers attempt to understand the experiment, how to support some aspect of it technologically and also to get feedback on the prototype itself. Karasti et al. (2010 p380) et al. have well-documented the issue of temporal scales and the need for awareness of "the multiple temporalities and particularly the long-term temporal scales," without which "studies of infrastructure development remain largely influenced by the prevalent, taken-for-granted short-term temporalities" of most technical project development.

But it is not always the case that the design and development of e-Science applications have been undertaken by dedicated software developers. Domain users and experts have also been involved in more fundamental ways in the development of e-Science applications and infrastructures. As the use of computer simulations and the processing of large data sets increasingly became a core feature of many areas of scientific research, end user programming featured as an important activity where some scientists wished to keep a significant element of control over the code used to develop workflows and simulations (de la Flor et al. 2010a). However, the e-Science programmes in the US and UK initially drew upon an industry enterprise model characterised by heavyweight server-side middleware where applications were developed by highly skilled software engineers.

With the emergence of Web 2.0,[5] more lightweight community applications enabled end users to make modifications to applications to fit local purposes. A good example of this is the virtual research environment, myExperiment[6] (de Roure et al. 2007) that supports the social sharing of bioinformatics workflows. These workflows allow the composition of multiple, individual steps in a research process into a single computational entity where research processes can be automated, shared with and re-used by others. This approach does require end users or domain experts to have the necessary programming skills to create or alter applications for use in their local environment. The orientation has a foundation in the open source model (Pierce et al. 2006) where code and applications are shared within a community consisting of developers, users and end-user developers.

A number of experiences have been reported in e-Science drawing upon a more agile orientation to end user development. For example, Pitt-Francis et al. (2008) report that when developing scientific software for computational biology, an agile approach, where close collaboration between domain scientists and software engineers is fostered through such activities as paired programming, is far more useful than plan-driven software development methods. This approach, Pitt-Francis et al. conclude, increases the quality of code through the sharing of different types of expertise, encourages rapid code development through short timeframe iterative cycles based on user stories, and allows for the high turnover in workforce typical of academic projects as it allows for the quick integration of new team members.

Agile development methods have also been investigated within project management activities in e-Science projects (Procter et al. 2010). From an in-depth study of the myExperiment project, Procter et al. suggest that distributed software development may be done in an agile manner, though they do remark that scaling up the user base with the possibility of the escalation for new features, may prove problematic. Thus it still remains a matter of investigation if agile management approaches can successfully be scaled up for use in the larger projects typical of the e-Science programme .

Perhaps a key distinctive feature of e-Science lies in what may be characterised as an almost ‘reflexive turn on the research process’, as some researchers seek to

[5] The infrastructure supporting Web 2.0 tools (cloud computing) seem strikingly similar to grid computing projects (see Pierce et al. 2006), but are quite different in terms of the funding, focus and user engagement

[6] http://www.myexperiment.org

investigate and describe research activities of their own disciplines. A clear example of this may be seen from within the e-social science programme where ethnographic analyses of observations of social scientific work with video, specifically the analytic work in data sessions, determined requirements for a distributed real-time video analysis system MiMeG (Fraser et al. 2006). In a similar manner, other researchers have investigated in what ways digital records might support ethnography most specifically "the work of description and representation that is required to reconcile the fragmented character of interaction in ubiquitous computing environments." (Crabtree et al. 2006). Whilst this is not a new phenomena, it is increasingly the case that interdisciplinary researchers are designing systems and considering the organisational contexts in which in which applications will be embedded for other groups of interdisciplinary scholars. It may be that some reflection on the impact of such scenarios on existing methods and approaches is required whilst also considering how they may be extended or transformed for e-Science.

4. Rethinking technological support for scientific collaboration

Although e-Science and CSCW seem concerned with quite distinct domains of practice, there is a range of interests they share in common, particularly, for example, in various forms of technology. e-Science technologies may seem highly specialised to supporting scientific work, yet there are correspondences between those technologies and those designed within CSCW, principally in trying to support different types of collaboration; whether this is rich co-present collaboration between teams, general asynchronous collaboration between groups to support scientific meetings, longer term asynchronous collaboration to support teams or enhancing artefacts to support everyday scientific work.

4.1. Supporting synchronous collaboration: media spaces and access grid

Long before the programmes in e-Science and Cyber-infrastructure emerged, early CSCW researchers were concerned with developing technologies to support research domains through the support of informal interactions and informal collaboration. Studies of scientific research (Galegher et al. 1990) had provided a basis for identifying the role of communications technology in professional work, in particular when carried out at a distance. What these studies reported upon was the importance of collaborative work. In the late 80s and early 90s this focus suggested particular forms of technologies to support collaborative research such as Media Spaces (Harrison 2009). Examples of early Media Space projects include Xerox PARC (Stults, 1988; Bly et al, 1993; Harrison et al., 1997), EuroPARC (Buxton & Moran, 1990; Gaver et al. 1992; Bellotti & Dourish, 1997; Dourish et al., 1996), BellCore (Fish et al., 1990; Kraut et al., 1994), and US

West Research (Bulick et al., 1989). All these projects were based on costly analog codecs for teleconferencing.

A later wave of projects looked at emerging digital technologies, including the University of Toronto Telepresence Project (Mantei et al., 1991; Buxton et al., 1997), PARC (Tang & Minneman, 1991), NTT in Japan (Ishii, 1990; Ishii & Kobayashi, 1992), and EuroPARC (Dourish & Bly, 1992). While many of the projects focused on always-on video as a way of linking dispersed spaces, Interval Research tried an experiment with always-on audio, with some success (Ackerman et al., 1997). These projects shared in common an interest in creating ways to use media to link together different places in order to facilitate easy, informal interactions, awareness of what was taking place at other locations, as well as planned, more formal interactions. The projects were all undertaken within research laboratories, and may be thought of as 'proof of concept' exercises. Interestingly, this meant that once the projects had finished, or proponents had left the organisation, the technologies did not seem to persist outside the laboratory research space or beyond the length of the projects.

However, as is pointed out in several places in Harrison (2009), much was learned from these projects about how to support a more integrated sense of 'place' even when the spaces involved were geographically dispersed (see also Harrison & Dourish, 1996). For instance, the media space projects were the origin of the major interest in CSCW in awareness, that is, knowing what is going on at a remote site or amongst your colleagues. A perusal of the CSCW and ECSCW research literatures shows a huge number of awareness studies in the 1990s and 2000s. While the always-on video (or audio) that were explored by the Media Space projects has not become the major mode for supporting awareness, a variety of other creative means of doing so have surfaced. Indeed, the emergence of instant messaging (IM), mobile devices such as cell phones and pad computers, and the widespread availability of high speed wireless have changed the landscape for how to support these features (Harrison 2009; Tang, 2009). So whilst the technical landscape has changed, the functionality highlighted by the Media Space work has emerged as key. Curiously a review of e-Science projects shows that this critical function for the support of cross-site collaborations has not been a prominent characteristic of e-Science support environments (Olson et al. 2008a). Media spaces focused on linking people together, but as a result they provided only weak support for the sharing of work objects. Subsequent analyses of media spaces in use revealed how important it is to be able to mutually collaborate over resources like paper or screen based documents, drawings, or views of objects. (Heath and Luff 1992; Dourish and Bellotti 1992; Gaver et al. 1993). In addition, some studies point to the details of how work is coordinated in these spaces, revealing the asymmetries in video mediated presence in these spaces stressing how much work is achieved through gesture and body movement accomplished on the periphery of the visual field (Heath and Luff 1992).

Interestingly, the emergence in recent decades of a wide range of digital technologies, including a rich array of mobile ones, would make the construction of media spaces much more feasible, both technically and economically. However, most attention has focussed on 1-on-1 interactions, with much less effort on creating a sense of virtual place by linking together dispersed sites. This would seem to be an especially valuable exercise for e-Science projects, where some of the advantages of a physical laboratory could be created that would allow for the kinds of informal, serendipitous interactions that are so typical of a collocated space.

Early in the development of e-Science infrastructure, when looking to support scientific teams and scientific meetings, a key technology was developed creating an advanced audio-visual infrastructure. The Access Grid is a technology platform to support scientific distributed collaboration (Childers et al. 2000; see also www.accessgrid.org) and shares many of the same features as Media Spaces. However, rather than supporting 1-1 interactions, Access Grid intends to support team-to-team interactions via high-speed networking over the Web. It also provides high quality real time audio and video connection that enables groups to communicate synchronously at multiple sites potentially from any location in the world. Whilst recognising that previously support for collaboration had most often occurred in small groups with colleagues (Childers et al. 2000), developers began determining the requirements for wide area group collaboration that included: the need to support multiple modes of interaction, from very structured to completely casual; the ability to share with other individuals or the group, scientific visualisations, spreadsheets, presentations, web sites, documents or movies; all requiring a system architecture capable of scalable wide area connectivity.

Early studies in CSCW examining the challenges of group to group remote scientific collaboration suggested that in such collaborations, articulation is a much more challenging activity than within a single group, as to accomplish this, perspectives from entire teams must be reconciled (Mark et al. 2003). In addition these studies suggest that the process through which information is conveyed in the large group-to-group configuration can greatly impact the expertise that is communicated from a site (ibid). Other studies suggest that collocated subgroups may quickly form in groups at the expense of isolated dispersed participants (Bos, et al. 2004).

To date these technologies have aimed to support large-scale collaboration that is increasingly concerned with real-time distributed collaboration over the output of the scientific work. Here the challenge is to enable multiple participants to interact over visualisations and artefacts and perhaps to immersively share three dimensional content, such as scientific and engineering data, in conjunction with their 2D Access Grid content. Indeed the use of three dimensional stereoscopic imaging in conjunction with features of Access Grid take that technology beyond a recording and mapping functionality (Carusi and Reimer 2010). The Collaborative Stereoscopic Access Grid Environment (CSAGE) drew upon semi-immersive stereoscopic facilities to create an enhanced sense of 'presence' in the Access Grid environment.[7] Some very interesting examples of the ways in which this functionality was used within the Arts and Humanities include the facial reconstruction of an archaeological skull conducted in a distributed manner over Access Grid. This work was later extended by Bailey et al. (2009) to support choreography and 'dancing on the grid' – investigating the ways in which performative actions may be undertaken collaboratively with minimum intrusion. Whilst such possibilities are indeed appealing, researchers have suggested that it is important to consider how naturalistic effects are embedded in technology because "in their apparent transparency, they may hide the artifices through which the activity was accomplished – and those artifices sometimes have a hidden formative effect on behaviours, in this case research practices." (Carusi and Jirotka 2010) In addition, researchers from the humanities have great experience and understanding of how such naturalistic effects may be achieved (ibid). This would seem a prime focus for CSCW to investigations – to form a multi-disciplinary focus for the design of technologies to support performative action around three-dimensional images, or indeed support for other types of collaborative performative actions.

Generally, Access Grid seems to have demonstrated successful distance collaborations in terms of usage for video-conferencing and support for meetings, and shows potential for integrating immersive visualisation experiences. Since its inception in 2000 it has had extensive usage within scientific communities globally (Corrie and Zimmerman 2009) and studies such as those described by Corrie and Zimmerman seem to indicate strongly positive findings, whilst also acknowledging, however, the change to social interaction that is required for users to drive these systems (p411). Yet, other studies that have considered the use of Access Grid for supporting analyses, such as those done by social scientists in data sessions around audio-visual materials (Fraser 2004), imply less positive findings. Fraser suggests that developers of e-Science technologies seem to be adopting 'the same old remote misunderstandings such as previous systems supporting remote collaboration. An example given in the paper shows the need for participants to "have access to one another's activities with respect to the data. Participants may have problems reconciling alternative viewpoints and perspectives on action. Simply juxtaposing video of participants and dataset visualisation is not enough (p2)". These misunderstandings, it is argued, arise from a lack of studies of working scientific practice in e-Science leading to simplistic notions of collaboration that are then embedded into distributed systems interfaces and designs. It would seem that what is required next is a deeper, detailed analysis of such technologies in use to determine whether previous issues of remote collaboration, such as those identified in CSCW,

[7] http://www.rcs.manchester.ac.uk/research/collaborativestereoscopicaccessgridenvironment

are of concern to e-Science development, and the extent to which they have been addressed and/or integrated into the design of technologies such as Access Grid.

Overall, research on Media Spaces and the rather widespread use of the Access Grid have helped to define some components of workable collaboration environments. It will be interesting to see how new emerging technologies take on the roles that these pioneering infrastructures have helped to define.

4.2. Sharing information resources: virtual research environments and collaboratories

Virtual research environments (VREs) (see for example Dovey 2010; Carusi and Jirotka 2010; Voss and Procter 2009) and collaboratories (Wulf 1993; Finholt and Olson 1997) have also merged around the capabilities of the new cyber-infrastructure, seeking to support a vision of large-scale collaboration. It is not simple to define VREs, Collaboratories, e-Science and Cyber-infrastructure in ways that may easily distinguish one from the other – (see Borda et al. 2006). The terms are still in flux (Schroeder 2008) which may reflect the novelty of the field. VRE's seek to support the research processes and practices of the institutions in which scientific researchers work, where the challenges of creating these connected distributed individuals along with the broader institutional settings in which they function, are not negligible. In the US, Collaboratories have sought to bring together disparate groups of researchers to tackle hard problems. VREs and collaboratories utilised five classes of functionalities: communication tools which allowed the participants to communicate either in real time or asynchronously; instrument access allowing remote access to instruments that were key to the domain, such as high end microscopes, upper atmospheric ground-based instruments, or instruments required for earthquake engineering; computation providing access to high end computational resources was a key element of the GRID infrastructure; repositories that allowed groups to create and access stores of data, publications, and other materials key to their research; and coordination facilities such as calendars, providing awareness information, and in general facilitating the coordination among members of these projects.

The concept of such remote scientific collaboration is not new however. As early as 1989 Willam Wulf coined the term collaboratory (Wulf 1989, 1993), a blend of laboratory and collaboration, to suggest a 'center without walls' that was enabled by high speed networked computers in which "the nations researchers can perform their research without regard to geographical location – interacting with colleagues, accessing instrumentation, sharing data and computational resources and accessing information in digital libraries" (Wulf 1993).

Whilst Cyber-infrastructure and e-Science refer to all aspects of the digital side of research infrastructure, VREs have perhaps been perceived more as an interface to that infrastructure, allowing easy access to data and services in an environment that is focussed on a particular research activity. (Carusi and Reimer 2010) Collaboration in these programmes is mostly conceived of as large-scale; yet, in the main, the VREs being developed are also intended to support the research process, which of necessity requires understanding the details of collaborative scholarly work that may often occur in smaller groups, and how to support it in a virtual setting. Furthermore, VREs are intended to be used by researchers from any discipline that seek to bring together disparate expertise to make progress on a problem of interest.

Whilst the visionary statements of key leaders like Wulf were critical to creating widespread interest in these models, Cummings and Kiesler (2005) showed that several large-scale efforts by the US National Science Foundation to facilitate such projects resulted in outcomes that were quite mixed. A key feature for making such projects successful was geographic or institutional dispersion. When projects involved multiple locations or institutions, the collaborations were less successful as rated by the participants themselves. So whilst the technologies enabled widespread attempts at aggressive institutional and geographically dispersed collaborations, the overall outcomes were not always successful.

The experiences of the use of VREs in the natural sciences and engineering domains suggest that a number of socio-technical issues are critical to their success (Sonnenwald 2007; Olson et al. 2008a; Cummings and Kiesler 2005). These issues include: the need for a vision-based organisational structure within the VRE (Sonnenwald et al. 2003); meaningful recognition for contributions (Arzberger and Finholt 2002; Welsh et al. 2006); trust among participants (Finholt 2002; Sonnenwald et al. 2003); the need for new information organisation and human-information interaction methods (Arzberger and Finholt 2002); support for situational awareness (Sonnenwald et al. 2004); and, some improvement over the current way of working (Sonnenwald et al. 2003).

Carusi and Jirotka (2010) suggest that it is at the level of creating and sharing knowledge that the important impact of VREs and collaboratories on research and epistemic practices may be found. They suggest that more detailed studies are needed of the practices through which research is conducted. By this they mean both unpacking the activities through which knowledge claims are made and warranted within particular scholarly fields such as measuring, interpreting, segmenting, *and* the epistemic practices through which knowledge goals are achieved. It is often through the challenges of accomplishing specific scientific work in collaboration, where interpretations are produced of the materials, tentative versions of analyses are provided, and the status of these and their relationship to publication and the certainty of data, where issues of sharing data become more problematic. (de la Flor et al. 2010a). What seems critical for all these technologies is to evaluate them, not only for their impact upon social, institutional and organisational collaboration, but also in terms of their epistemic outcomes, which may take longer to achieve.

4.3. Augmenting artefacts in the scientific research process

There are a number of prototype technologies where the relationship between CSCW and e-Science may be immediately closer, where CSCW researchers have investigated technologies to support research activities, and where common themes and interests between researchers in e-Science, CSCW, HCI and Ubiquitous Computing have emerged. One general type of physical artefact of concern is the paper document – whether it is the laboratory notebook, fieldwork notes or everyday notes of the scientist. So, for example, Makay et al. (2002) report on their work developing a prototype augmented laboratory notebook that integrates physical and electronic documents whilst also managing the legal constraints on scientists to keep paper-based notebooks that serve as legal records of their research findings. Yeh et al. (2006), drawing on the advantages that paper affords for mobile work practices, developed a mobile capture and access system for field biologists that integrates paper notes with digital photographs captured during field research. They argue that such next generation tools need to support the capture of heterogeneous data, aid the transformation process, whilst preserving the best aspects of current paper-centric practices. A related tool FlutterbyNet (Kim et al. 2006) is a wall interface for digital notebook sharing, allowing distributed designers to better communicate and collaborate around their ideas. This approach has also been investigated by some researchers in the e-Science programme who were interested in supporting and augmenting artefacts of the research process in specific fields including, biology (Mascord et al. 2006), synthetic chemistry (Hughes et al. 2005), generic e-Lab notebooks (Frey et al. 2005) and archaeology (Warwick et al. 2009).

However, specific discussions of issues arising from these studies across the fields of investigation have yet to occur. It is unclear how such tools might be integrated into the wider infrastructures envisioned by the e-Science and CI programmes - they do not seem to sit comfortably with the larger mission statements of ‘Big Science’ and ‘Big Data’. Yet they share in common the CSCW focus on the scientists’ everyday work practices, with a view to enabling new collaborations. Of related concern is the need to develop deeper and more detailed evaluations of these tools in relation to the researchers’ everyday work practices and the institutional concerns in which their research is embedded. In CSCW, the value of assessing technologies in use has emerged over the last few decades of research into understanding embodied interaction, the nature of teams, space and time, the roles of different people in collaboration, and the details of work practice. Such approaches are yet to form an integral part of the e-Science and CI programmes.

5. Conceptual contributions to understanding scientific practice

Scholars in sociology and science and technology studies (STS) have been attempting to detail the fundamental problem of understanding how complex, large e-Science collaborations form, what participation looks like, and how they are maintained across multiple institutions and diverse cultures over time. More recently, CSCW has also taken up the challenge of studying such collaborations in order to inform the design of collaborative systems and tools to support e-Science and, by extension, the work of the organisations themselves.

Despite the perceived advantages and benefits afforded by large-scale collaborations, the challenges of realising such complex collaborations in e-Science are formidable. Creating a large-scale socio-technical system is a complex matter requiring the collaboration of a great many stakeholders at individual, group, and organisational levels, each of which has established policies, preferred practices, and technological systems. This challenge of understanding such inter-organisational relationships was addressed initially by Anselm Strauss (1988) through the notion of 'social worlds'. In his investigations, instead of defining social organisations by spatial, territorial or formal membership, the boundaries of social worlds were seen to be determined through interactions and communications that intersect and transcend any formal demarcation. With the proliferation of new scientific collaborations, a key question of how best to design and embed e-Science applications and infrastructures has emerged, with concerns relating to how best to create useful, large-scale, distributed, computer-supported, scientific endeavours. In order to address this challenge, Strauss' conceptual innovation has been invoked and elaborated by a number of CSCW and STS researchers with similar theoretical leanings.

Whilst the types of collaborations discussed in this section represent a proportion of activity in e-Science aligned to a particular analytic orientation – that of symbolic interactionism, they are by no means the only types of activities to yield interesting conceptual contributions. Other studies, such as those informed by an ethnomethodological orientation typically focus on the details of how scientific work is produced and achieved in interaction with others. Such investigations are also yielding contributions, often in the form of re-conceptualisations of existing notions such as 'task' 'collaboration' and 'coordination' and yet may often arrive at complementary conclusions. Their differences reflect the analytic distinctions brought to the understanding of collaboration.

5.1. Articulation work and infrastructure

Building on the social worlds and interactionist perspectives, whereby interaction and communication is fundamental to understanding social groups, sociologist Anselm Strauss' work in medical settings (1988) developed the concept of *articulation work* to help understand the interaction of people organising project work. Subsequently, the concept has been taken up by CSCW researchers and expanded upon (Schmidt et al. 1992; Lee 2007; Gerson 2008; Bietz et al., 2012).

Strauss describes articulation work as referring to “the specifics of putting together tasks, task sequences, task clusters—even aligning larger units such as lines of work and subprojects—in the service of work flow” (Strauss 1988, p. 164). The *articulation process* is putting and keeping together elements of work. In discussing articulation work, Gerson (2008) distinguishes between the concepts of ‘metawork’ and ‘local articulation’.

> “Strauss used the notion of articulation work in two different senses (e.g. Strauss 1988). On the one hand, articulation work is about making sure all the various resources needed to accomplish something are in place and functioning where and when they’re needed in the local situation. This means bringing together everything needed to accomplish a task at a particular time and place, including all the administrative and support functions such as janitorial services, food service, equipment maintenance, and covering for staff out sick or on vacation. The concern and emphasis in this sense are on particular situations rather than classes of activity.” (Gerson 2008, p. 196)

Here Gerson calls local articulation the bringing together of local resources for a particular situation. Metawork, on the other hand, refers to classes of activities or formal generalisations (e.g. task lists) of work. For the purposes of this overview, we do not distinguish between the two senses of articulation work and use the concept generically. As we will discuss below, the investigation of articulation work in e-Science has both informed investigations in CSCW and the development of concepts within CSCW.

While interest in infrastructure in CSCW has developed largely apart from research into articulation work, some conceptual work has emerged that attempts to weave the two together (Bietz and Lee 2009). While most cyber-infrastructure studies within CSCW are *not* using the notion of articulation work directly (or may be using related constructs such as *coordination mechanisms* (Schmidt and Simone (1996) or *ordering systems* (Schmidt and Wagner 2005)) it is fair to characterise much of the work in CSCW in the area of cyber-infrastructure studies as investigations of articulation work: It is difficult to explore the development and maintenance of cyber-infrastructure without also grappling with how project work plays out through subprojects and tasks—a fundamental concern of articulation work. The contribution of much of the work on large-scale cyber-infrastructures in CSCW has been a simple, but profound refocusing on how work is organised in larger, more geographically dispersed and not rigidly hierarchal collaborations, and how these are organised and evolve as socio-technical endeavours.

The notion of infrastructure is, in fact, not new to CSCW, but has gained in prominence in recent years. The seminal work by Star and Ruhleder (1996), which may itself be characterised as an early e-Science study, emphasised the importance of infrastructure as a topic of study whilst also highlighting that understanding of infrastructure is necessarily embedded in other relationships requiring ongoing maintenance efforts (Star and Ruhleder 1996). This notion of the relational character of infrastructure has been taken up by other CSCW researchers (see for example Ribes and Finholt 2009; Lee et al. 2006), and the work required to enact and maintain infrastructure and the notion of infrastructure as collaborative accomplishment continues to be influential in studies of infrastructure within CSCW. Infrastructure development often requires the participation and collaboration of diverse work sub-cultures.

Conceptual contributions to CSCW have centred on *creating conceptual frameworks that detail and explain the class of strategies that individuals, groups, and organisations undertake to enact and maintain heterogeneous e-Science organisations* (Ribes and Finholt 2009; Bietz et al. 2010). Within CSCW, researchers have found the notion of infrastructure to be a useful lens for understanding more diffuse types of collaborations. For example, recent work has drawn upon the notion of infrastructure to investigate diverse collaborative practices such as those in governmental information systems, establishing a perspective on organisational Information Technology as work infrastructure that then describes challenges for designing within and for this type of infrastructure. (Pipek and Wulf 2009). It has also been useful in research investigating settings where citizens experienced ongoing disruption in a conflict zone during the Gulf War, to show *how citizens used information and communication technologies continuously to* resolve breakdowns in infrastructure by creating new, reliable technology-mediated social arrangements that enabled people to maintain daily routines (Semaan and Mark 2011); And the concept has been further utilised in distributed online gaming, showing how the experience of playing arcade games changes with different socio-technical infrastructures, highlighting how infrastructure shapes experience and collaboration as well as how new social infrastructures emerge through the interplay between the game, the platform, the community and the media ecology. (Wang et al. 2012). We are still early in the work of creating conceptual frameworks that sufficiently detail the design space for cyber-infrastructure development. For now our conceptual contributions are adding more detail to a, thus far, still rough map of a promising socio-technical design space using the approaches and lenses of infrastructure and articulation work.

5.2. Boundary objects and databases as boundary negotiating artefacts

A concept related to articulation work, also from Susan Leigh Star and collaborators, is the notion of *boundary objects* - objects that may help coordinate activities between different communities of practice (Star and Griesemer 1989; Bowker and Star 1999). The notion of boundary objects has been taken up within CSCW and attempts are being made to refine the concept further in the context of the current socio-technical arrangements in e-Science and beyond. The concept of boundary objects was originally developed as a way to understand how communities with different practices and needs were able to cooperate using the same object. The concept provided a way to look at collaboration in the context of heterogeneity.

Exploring databases drawing on the concept of boundary objects has yielded new understandings of the role of database development in the sciences within CSCW. Boundary objects, to iterate, are objects that inhabit several communities of practice, satisfy the informational requirements of each of them, and maintain a common identity across sites (Bowker and Star 1999). Bietz and Lee (2009) use the extension of boundary objects (Star and Griesemer 1989) - *boundary negotiating artefacts* - to understand the role of database development in developing new research practices and research communities. Boundary negotiating artefacts can be used to cross borders between communities of practice, as do boundary objects, but may also be used to affect the division of labour, or in other words, to push and establish the boundaries between communities of practice (Lee 2007). Boundary negotiating artefacts are used by collaborators to record, organise, explore and share ideas; introduce concepts and techniques; create alliances; create a venue for the exchange of information; augment brokering activities; and create shared understanding about specific problems (Lee 2007). While we privileged the artefacts rather than the practices to keep the construct parallel to the notion of boundary objects, boundary negotiating artefacts must be understood as organising and formalising sets of very different collaborative practices. Each artefact sits at the nexus of certain practices. For example "proposing artefacts" organises design discussions and debates over what will or will not be including in a finished design product and supports micro practices such as 'suggesting', 'debating', and 'declining' or 'accepting'. The key contribution of boundary negotiating artefacts is to understand that unstandardised artefacts can be used to organise work in nascent, unestablished collaborations and that in the course of engaging in work through the artefact, the collaboration itself is organised.

Some CSCW researchers have investigated how practices must be aligned in the process of sharing scientific data in particular (Birnholtz and Bietz 2003; Faniel and Jacobsen 2010; Zimmerman 2008). The study of data sharing in e-Science has historically focused on scientific and distributed collaborations. Before data may be shared or re-used, work must be done on the data and work practices must be aligned. Conceptual interrogations of data practice have investigated how negotiations around and through database development are integral to the instantiation of new scientific practices. Understanding databases as boundary negotiating artefacts helps us to understand how databases require alignment of actors (Latour 1986) and instantiate processes and negotiations while also being dynamic. For example, research has shown how metadata and metadata standards become contested artefacts and sites of negotiation within the metagenomics (a relatively new scientific approach) and wider genomics communities; a database becomes a contested site around which particular research questions are supported or disenfranchised through the inclusion or exclusion of necessary metadata (Bietz and Lee 2009). Scientific collaboration between biologists and computational disciplines, for example, has been noted as an endeavour that requires interpretive frames to be brought together (O'Day et al. 2001). In their study investigating environmental genomics (metagenomics), researchers developing a database in order to share data, (Bietz and Lee 2009) found that diverse practices, interpretive frames, and different scientific concerns were brought together by databases which are seen to function as boundary negotiating artefacts.

Metagenomic science requires complex coordination around a database. Rather than looking at databases as static, databases are seen as being created and maintained across scientific communities where researchers have different practices and priorities. In dynamic environments, the number of true boundary objects that satisfy the information requirements of multiple communities of practice may be relatively few compared to any number of boundary objects or boundary negotiating artefacts. By studying this metagenomic e-Science environment and the conceptual reframing of the database as another type of artefact that coordinates multiple perspectives, we may begin to see how multiple databases might sometimes be necessary and useful for complex, innovative work like research. Such a conceptual reframing then points to a particular problem space ripe for CSCW investigation; namely how to support large-scale collaborations that are reliant on multiple databases that support a multiplicity of knowledge-building priorities and practices.

Ultimately the line between boundary objects and boundary negotiating artefacts is blurred. The power of the distinction is *not* in the demarcation of the two, but rather in understanding the relationship and interaction between the two and what those mean for the practice and support of articulation work. In other words, by understanding how boundary negotiating artefacts and boundary objects change over time and under what circumstances, we can better understand how to design to support stability or change. Articulation work may be highly diverse depending on the nature of the collaboration being supported. Therefore, by extension, the types of objects and artefacts that 'live' between communities of practice must therefore also be highly diverse. The work of exploring databases as boundary negotiating artefacts provides a link between well-established concepts such as boundary objects and their elaborations that are useful for understanding how temporary or emergent collaborations unfold. For the time being, these concepts may be viewed as sensitising concepts to remind the CSCW community that some practices and artefacts are necessarily emergent, and that encoding particular practices and artefacts into a database or other system is to also privilege those particular practices to the exclusion of others. The concepts of boundary objects, boundary negotiating artefacts, and databases as boundary negotiating artefacts may further assist in understanding what can and cannot be readily codified into a tool or system and where, in a computerised system, flexibility and openness is desirable and, at times, essential.

5.3. Human infrastructure and synergizing

As mentioned in the previous section, the field of CSCW has both applied and extended the notion of *infrastructure*. Both the study (Star 1999) and the design of infrastructures (Ribes & Finholt 2007) present numerous conceptual opportunities for CSCW. Infrastructures are composed of heterogeneous entities and relationships; emerge and evolve over long periods of time and across great physical distances; simultaneously they have embedded in them, and are embedded in, other infrastructures; and are the result of interactions among a variety of entities such as individuals, groups, networks, and organisations (Bietz et al. 2010). Star and Ruhleder explain infrastructure as a fundamentally relational concept marked by ambiguity and multiple meanings: "An infrastructure occurs when the tension between local and global is resolved. That is, an infrastructure occurs when local practices are afforded by a larger-scale technology which can then be used in a natural, ready-to-hand fashion (p. 114)."

Various scholars have built upon Star and Ruhleder's (1996) conceptualisation of infrastructure. For example, Lee, et al. (2006) focused on the human infrastructure of a large distributed cyber-infrastructure project, the Function Biomedical Informatics Research Network, (FBIRN) which itself is one part of the larger BIRN project (Olson et al. 2008b). The goal of FBIRN was to pool functional magnetic resonance images (fMRI) of patients with schizophrenia and comprised a consortium of scientists from thirteen different institutions across the US. Characterising the features of the human infrastructure, as "multimorphous," "dynamic," "processual" and "relational", Lee et al. (2006) point out that the normal concepts of teams, personal networks, and organisations do not fit easily with this human infrastructure. Thus, even though it is the emerging cyber-infrastructure that makes the project possible, sites and face-to-face interactions play a key role in getting the work done. While working groups and task forces are fundamental to the organisation of the project, there is often some confusion about what these are, who is on them, and how they operate. Participants may not know who the members of the project are, including in a number of cases, whether they themselves are part of it. As the participants point out, "there is no defined outer periphery of membership [p. 486]." However, they also note that "not having a clear view of FBIRN membership may actually be advantageous for collaboration [p. 486]." In the host organisations that comprise FBIRN, people come and go, have their job assignments changed, and are all involved in other projects as well. Indeed, they point out that other projects that have a lot of overlap with FBIRN can result in facilitating leverage rather than wasteful duplication. Similarly, there was a lot of fluidity between FBIRN and the other parts of the larger BIRN project. At a broader level FBIRN is similar to a number of other larger, distributed e-Science projects (see for example Brady et al. 2004) thus, the features described, may be common across these types of projects Bietz et al. (2010) investigated how human infrastructure develops e-Science by enacting productive infrastructural relationship between individuals, groups, organisations, and technologies. Based on their qualitative research of the development of a metagenomics database for scientists, they call the work of enacting productive relationships "synergizing" work. If articulation work is the work of making work go well, then synergizing can be crudely summarised as work to make sure that the work happens at all. Strauss (1988) himself noted that additional models are probably needed to analyze articulation work that is between or that encompasses more than one organisation. Inter-organisational relationships do indeed seem to be different from intra-organisational ones (Bietz et al. 2010). Synergizing is useful for theorising inter-organisational relationships. While retaining similarities, Synergizing differs from articulation work in that it is concerned not with modifying and coordinating an existing common field of work, but with creating the field of work itself (e.g. a project).

Consider Strauss' (1988) discussion of the complex analytical problem of dealing with projects within an organisation that have sub-projects. Components may begin to function long before the overall infrastructure is complete, or the same component may be part of multiple infrastructures. Given this piecemeal development, the field of work is constantly in flux. Additionally, whole groups, software tools, communities, organisations, and other elements of human infrastructure may come and go as the project matures. A challenge for synergizing is not only to restrain complexity within existing fields of work, but to also *bring together elements in order to create and maintain a common field of work itself*. The common field of work may be conceived of as the desired socio-technical collaboration, but this goal is itself comprised of sub-projects that serve as their own field of work. Synergizing is a strategy for creating, managing, and utilising complex interdependences in an embedded infrastructure that brings together multiple organisations, projects, people, and technologies.

The exploration of concepts such as articulation work and infrastructuring in conjunction with empirical studies of e-Science has resulted in the refinement of previous concepts such as boundary objects, articulation work, and cyber-infrastructure-as-hardware with refinements such as human infrastructure and socio-technical development features such as synergizing. The elaboration of concepts from science and technology studies within the field of CSCW may be viewed as a reaction to an immediate need within the e-Science and cyber-infrastructure communities for techniques to assist in the design and management of computer supported cooperative infrastructures and the applications, tools, human infrastructures, and virtual organisations that comprise them. Further conceptual development within CSCW may help to map new design spaces in e-Science and further refinement is needed to delineate actionable design problems and to identify desirable design solutions.

6. Future research challenges and opportunities

With the continued advancement and proliferation of technologies and scientific practices, the challenge of supporting the growth of 'computer supported cooperative science' may be simultaneously daunting and exciting. Yet, despite its history and current growth, computer supported cooperative science has remained relatively understudied in the field of computer supported cooperative work. CSCW investigations of e-Science are facing challenges that often surface in an emerging research area: that of drawing on other disciplines for useful theory (Latour 1987; Keptelinin and Nardi 2006; Hutchins 1995) and methodology (Garfinkel 1967; Corbin and Strauss 2007), whilst also investigating scientific areas that have not perhaps been previously explored.

The need for more CSCW research in the e-Science area is, in part, fuelled by the requirement to support the proliferation of scientific practice and different configurations of activities. For example, disciplines may have very different notions of how to collect and document data, different conceptions of the role of software tools in various parts of the scientific process, and different relationships to larger collectivities (Fry 2006). Even within a given scientific discipline, multiple studies may be necessary in order to address the full range of practices from tool development and use, to data collection and analysis, to group, community, organisational, and infrastructural formation and reformation.

The study of large-scale collaborations are already entering the landscape of CSCW and the configurations of these large collaborations play a role in shaping and constraining the work practices of teams and groups. Large-scale collaborations are also integral to the successful accomplishment of a great deal of e-Science activities. Despite the rhetoric of the vision, however, not all e-Science is moving towards larger, data-intensive collaborations. A recent international review of the UK e-Science programme concluded:

> "The technologies and practices of e-Science, together with the e infrastructure on which it rests, must be both a topic as well as an enabler of research and development; and this duality needs to be made synergistic. E-Science as a topic of research includes both technological and social (behavioural, economic, legal, ethical) dimensions." (RCUK Review of e-Science 2009 p.51)

An opportunity, thus, exists for CSCW research to study work in all areas of e-Science in order to understand and design end-user technologies for teams, groups and organisations within a framework of the challenges of embedding e-Science applications. And whilst, most of the studies described in this article have been based or centred within the UK or US, it is increasingly important to develop perspectives that reflect research in e-Science as an international endeavour and that discuss differing perspectives on the critical issues.

In the section below, we report on a few key research challenges that represent general opportunities for e-Science and CSCW in terms of technological (or more precisely socio-technical), methodological, and conceptual (or theoretical) innovation.

6.1. Socio-technical configurations and technologies

Previously we have discussed technologies and socio-technical arrangements as they pertain to e-Science and CSCW looking specifically at the similarities and differences in their use and objectives. Looking towards future research challenges and opportunities, we shift our focus slightly to technologically driven research areas. Particular technologies may change frequently, however, classes of technologies and research areas may endure for much longer. Specific technologies and the details of their development, use, and appropriation continue to be relevant, but for the purposes of this article, we wish to generalise to broader classes of technologies and the socio-technical arrangements to which they point, in the hopes of having more enduring relevance.

Research on the particular collaborative practices of specific scientific sub-domains is a particularly obvious fit for CSCW given the field's history of conducting workplace studies and the focus on the study of smaller groups (Grochow et al. 2010; Poon et al. 2008; Jirotka et al. 2005). As science fields begin to proliferate due to the creation of hybrids like astrobiology, geochemistry, bioinformatics, the number of specialised collaborative work practices seems also to be increasing. Each scientific domain, such as physics or biology, has it's own body of basic knowledge, but as fields become increasingly specialised, so too do work practices and attendant technologies. The challenge here lies in understanding the particular skills and practices of such hybrid domains and scientific sub-cultures and how novel collaboration may be supported by technology. This is often more complicated in domains where practices have not yet been fully formulated and are still in flux.

A second area providing prime research opportunities may be found in the technologies and socio-technical configurations that comprise collaborative activities around the collection, sharing, and analysis of research data (Birnholtz and Bietz 2003; Faniel and Jacobsen 2010; Ure et al. 2009; Zimmerman 2008). The imperatives of complex research questions, such as sharing brain imaging and research subject data may not only require collaboration between researchers but also often necessitate the sharing of research data (Lee et al. 2006). Due to the particular practices that develop in each project group or each sub-domain, data can very rarely be shared without a series of discussions and negotiations (Zimmerman 2008). Establishing metadata standards is a common technical solution, but in fact necessitates in turn, meta-level discussions (Zimmerman 2007; Bietz and Lee 2009). A future challenge for CSCW is to address the complexities the study of data practices as socio-technical practices. Fields such as Computer Science focus on technology, and Information Science has typically focused on the organisation and accessibility of information. CSCW can draw from these fields, but may also integrate an important focus on work-practice informed design. While the CSCW work in this area to date focuses on data sharing and re-use, all related data practices such as, analysis, harmonisation and curation, are also critical areas for study (Cragin and Shankar 2006; Karasti et al. 2006; Oleksik et al. 2012).

A third area concerns supporting work practices around artefacts, or material practices. As noted earlier there are currently various research opportunities investigating humanities, social science, and the arts, with work already being undertaken in CSCW (de la Flor et al. 2010a; Eden and Jirotka 2012; Eden et al. 2012). Research here has focused on understanding how artefacts, instruments and manuscripts are used as resources for collaboration. For example, using image processing to simulate the tilting of an ancient tablet by classicists interpreting ancient markings (de la Flor et al. 2010b). These studies have highlighted the notion of materiality – how scientists make inferences from the material and physical properties of the artefacts that form part of their ongoing interpretations of data. What may be needed in the future is to understand how different qualities of materiality are used and interpreted in a range of settings as features to be included in design, and what types of technologies may best support such detailed, collaborative work.

The fourth area for investigation is the study of large-scale e-Science as virtual organisations (Bietz et al. 2010; Kee and Browning 2010; Lee et al. 2006). We do not mean here that such organisations are purely virtual; rather, that they are supported by distributed ways of working. If we take seriously the vision that e-Science infrastructures and applications support scientific communities, that scientific communities are in the business of rapid knowledge production, and that rapid knowledge production entails not only producing knowledge quickly, but also rapidly producing new ways of 'knowing' that in turn require new technologies and new ways of working, we begin to understand that the grand challenge of these virtual organisations is simultaneously being stable enough to support work while also changing quickly enough to be relevant (Lee et al. 2012; Bietz et al. 2013).

To further complicate matters, e-Science as virtual organisations may take different forms and privilege different features, for example, particular software architectures, sub-disciplines, approaches (e.g. citizen science), collaborators (individual or organisational), instrumentation, collaborative forms (e.g. distributed teams) or some or any number of these. These differing organisational forms and priorities can, in turn, trigger a set of institutional, ethical and legal concerns which, when considered in the global context much proclaimed by e-Science, create a raft of challenges for embedding these virtual technologies (Karasti et al. 2010; Kee and Browning 2010). In addition, recent work on citizen science (Kim et al. 2011; Luther et al. 2009; Wiggins and Crowston 2010), which often uses the terms "crowdsourcing" or "crowdworking", points to additional venues where the coordination of data sharing practices and technologies is also paramount. Further investigation of these areas is needed to support the enactment of e-Science through collaborative data sharing.

Outside the field of CSCW, various aspects of e-Science are attracting a great deal of effort and energy on the part of practitioners in science, engineering, and industry. The technologies that do not yet have an established base within the e-Science-oriented CSCW literature, also represent challenges that could potentially yield larger opportunities for discovery. Such technologies, which are not at all mutually exclusive, include: the visualisation of scientific data (Aragon et al.

2008; Kandel et al. 2011), middleware systems including workflow systems (Salayandia et al. 2006; Vigder et al. 2008; Woollard et al. 2008), the collaborative development of cyber-infrastructure software (Segal and Morris 2008; Segal 2009), and cloud computing (Hoffa et al. 2008; Vecchiola et al. 2009). These aspects are already the subject of research in fields that are more oriented towards the development of applications and production systems. Rapid developments in these areas are embedded in existing socio-technical arrangements, but at the same time are also in the constructive process of creating new socio-technical arrangements. Investigations of these “sites” of research may provide new opportunities for designing collaborative software and work practices and also new opportunities for conceptually mapping, or theorising, the design space.

The four areas of e-Science activities described above are already attracting the interest of CSCW scholars and represent what may be a “sweet spot” for researchers interested in addressing e-Science challenges, as they are available for exploration, whilst having the benefit of existing literature and expertise within the field (Birnholtz and Bietz 2003; Jirotka et al. 2005; Karasti et al. 2006; Olson et al. 2008b; Poon et al. 2008; Luo et al. 2008; Luther et al. 2009; Faniel and Jacobsen 2010; Ure et al. 2009; Zimmerman 2008; Grochow et al. 2010; Bietz et al. 2010; Kee and Browning 2010; Lee et al. 2006; Wiggins and Crowston 2010; Faniel and Jacobsen 2010; Faniel and Zimmerman 2011; Kim et al. 2011; Tabard et al. 2012; Huang et al. 2013).

6.2. Methodological challenges and opportunities

As with other design-oriented research disciplines, the challenges CSCW researchers face reflect the challenges presented by the research sites. The development of e-Science technologies and infrastructure requires methods of studying collaborations that are often diffuse, geographically distributed, asynchronous, and sometimes without clear delineations between users and developers (Lee et al. 2010). Such collaborations pose challenges for purely ethnographic methods and whilst early efforts are making good headway (Olson et al. 2008b), we still know too little about how to model these collaborations. More sophisticated models may help us to measure the relative importance of a more comprehensive set of factors and to better understand the interactions between them.

As the creation and maintenance of e-Science requires the collaboration of people from diverse backgrounds and interests, so too does research on e-Science. Researchers in this area often find themselves undertaking interdisciplinary collaborative CSCW research on interdisciplinary collaborative scientific research. Such reflexive research challenges are not unique to e-Science, they exist throughout most areas of CSCW. Yet whilst the field of CSCW has long been interdisciplinary, there now seems to be a need in the e-Science domain to be even more so. Within CSCW, researchers increasingly find it necessary to collaborate with domain scientists, often taking an approach closer to participatory design even if that is not their primary methodological orientation. In these investigations of e-Science, researchers may draw upon complimentary disciplines such as, sociology, psychology or organisational studies (Olson et al. 2008a), depending on the phenomena in question, in order to acquire necessary or helpful knowledge to enable better understanding of the particulars of the accomplishment of work and the larger mores that define success and due process.

Here we suggest envisioning a triangle of interests where the points of the triangle represent the interests of CSCW researchers who study e-Science and with the points of the triangle being scientific discovery, technology development and maintenance, and the socio-technical organisation of work. Researchers may not be equally concerned with all points of the "science—technology—sociality triangle." We can take as a given that CSCW scholars are interested in the relationship between technology and sociality, therefore we draw attention to the other two dimensions. Those researchers who are most interested in the science—technology side of the triangle may need to delve into some of the minutiae of work practice and articulation work in order to appropriately and artfully support and potentially transform current practices. At the same time, researchers who are most interested in the science—sociality side of the triangle may need to understand in more detail some of the minutiae of scientific instruments, software, and other information technologies. As our experience, literature, and human resources expand, so too might CSCW's potential impact on science practice, science policy, and on scientific discovery. The potential and appropriate role for CSCW in making science happen is an open question. As individual researchers, we may need not attend to the entirety of the triangle, but as a field as a whole, such a balance may be desirable.

A related significant ongoing challenge for executing methods is the substantial need for researchers to understand sometimes highly complex scientific information and also to report on highly technical and specialised collaborative artefacts and organisational practices. Learning and reporting on science practice, particularly in the novel cross-disciplinary areas, and translating this information across different domains, is not the same as reporting on every day life, public web services, or familiar office settings. Researchers, and those that oversee them, must be cognizant of the extra amount of time that it will take to become familiar with the skills and practices of these complex interdisciplinary domains and plan projects accordingly.

The opportunities afforded by interdisciplinary collaboration among scientists have been acknowledged for some time (Palmer 2001) and are currently undergoing renewed interest (Klein 2010). Yet such collaboration frequently entails particular challenges including, a potential clash of disciplinary cultures (Karasti et al. 2010), and coordination and management issues, especially in the case of multi-institutional projects (Cummings and Kiesler 2008). These challenges, which we both study and face as researchers, seem to broaden out into more meta-methodological concerns as researchers grapple with competing epistemologies and coordination issues.

A further challenge for future research lies in *finding ways to make the complexity of socio-technical e-Science configurations more tractable without being reductionist*. For example, identifying what or which units of analysis to use within studies of e-Science settings. The most common unit of analysis for much research so far has been the unit of the project (Star and Ruhleder 1996; Ribes and Finholt 2009). A variety of acronyms denoting various e-Science projects have made their appearance in the CSCW literature (e.g WCS, LEADGEON, eDiaMoND LTERUARC/SPARC, FBIRN, NEES, CAMERA). These projects are very much embedded in a complex web of relationships that often put pressure on existing disciplinary, institutional and personal practices (Haythornthwaite et al. 2006). e-Science organisations, as constellations of technical and human infrastructures (Lee et al. 2006), are themselves extremely heterogeneous. Issues of scope and scale within the study of e-Science provide additional methodological challenges. Just as e-Science projects need to be scoped according to whether they will support a team, distributed group, or organisation, researchers too must similarly scope and scale their own research and identify or determine appropriate methods and techniques to undertake the investigations. Thus, for example, we may require new methods and approaches or novel combinations of existing approaches to investigate such highly distributed, cross-cultural, multi-disciplinary, multi-institutional and highly specialised configurations of work practice – approaches that may also be highly agile and reflect longer term considerations of disciplinary concerns and the impact upon the scientific record. These issues of scope become a serious methodological concern for the study and design of e-Science systems. Some recent efforts for example, have attempted to develop new methods for studying distributed socio-technical systems such as “trace ethnography” that allow researchers to generate empirical accounts of network-level phenomena without having to be present at every node (Geiger and Ribes 2010).

Many e-Science systems are bespoke technologies where users may be actually quite keen to try out new prototypes. Yet the lack of uptake of specific technologies may often result from the larger context. The broader socio-technical context that influences the uptake of specific technologies may include considerations pertaining to the sustainability of particular technologies including institutional arrangements and the sustainability practices of individuals and teams (Lee et al. 2012). Sustainability focuses on reconciling the short-term nature of projects and their funding with the long-term view required to develop the infrastructure (Karasti et al. 2010). This has become a key concern for many of the large e-Science programmes as research funding draws to an end (Carusi and Reimer 2010), and has been identified as a key challenge to uptake (Procter et al. 2010). Many researchers simply will not invest the time to learn new ways of working or adopt new technologies if it is not certain these technologies will be sustainable in the long term. In addition, institutional arrangements may encourage or impede particular types of work practices such as, issues of credit and attribution (David and Spence 2010; Welsh et al. 2006). Additional concerns include stakeholder groups (institutions, disciplines, and individuals) having different and conflicting ethical concerns about data and expected levels of privacy and different methodological practices. (Carusi and Jirotka 2009). Challenges such as these for design and research methodologies may not be unique to e-Science, however they rise quickly to the surface in the highly interconnected arrangements of e-Science collaboration. They provide an opportunity to rethink and reframe how we think about collaborative design and the uptake of new technologies.

6.3. Conceptual challenges and opportunities

Considerations about conceptual frameworks and theories in e-Science parallel the methodological issues pertaining to scoping. Four broad challenges for conceptualising e-Science as a space for CSCW include: reframing what have been perceived by other fields as technical problems with technical solutions, as socio-technical problems with opportunities for socio-technical intervention; finding ways to make the complexity of socio-technical e-Science configurations more tractable without being reductionist; unpacking e-Science practices and what it means to support continuous discovery; and supporting and learning from the particularities of specific scientific disciplines and sub-disciplines while also generating conceptual frameworks and theories that transcend individual disciplines and s ub-disciplines.

The first conceptual challenge for CSCW in e-Science is *reframing what have customarily been considered technical problems with technical solutions by other fields as socio-technical, CSCW concerns*. This is a not a radical idea in our field; Ackerman (2000) wrote about a "social-technical gap" almost fifteen years ago. Rather, the conceptual challenge lies in more assertively reframing what may appear to be technical or seemingly mundane aspects of scientific work practice as lying within the scope of CSCW design challenges. CSCW has long studied the very technical or seemingly mundane aspects of work while science has received much less attention. The field of STS has long been studying technical or seemingly mundane aspects of scientific work practice particularly from the perspective of the humanities and social sciences, yet without CSCW's remit to design, and inform the design, of collaborative systems. The design orientation of CSCW, positions the field well to make a positive impact on science collaborations in numerous ways including a conceptual reframing of the design space.

Within CSCW, investigations of database and middleware development are beginning to take on the challenge of reframing technical problems as socio-technical. For example, Brubaker and Hayes (2011) focus our understanding on the fact that data is representational. Before data can be shared or re-used, work must be done on the data and work practices must be aligned (Hartswood et al. 2012). Related research has been done investigating how practices must be aligned in the process of sharing data (Birnholtz and Bietz 2003; Faniel and Jacobsen 2010; Zimmerman 2008). Other work has investigated databases as not just a technology to store data, but also as boundary negotiating artefacts (different from clean-crossing, standardised boundary objects) that are used to negotiate work and establish new practices (Bietz and Lee 2009). Recent research also highlights that digital artefacts are embedded in a broader data economy and points to the importance of accounting for the ways in which data are produced and acquired (Vertesi and Dourish 2011). In a related vein we are also now seeing CSCW research that investigates scientific software as socio-technical phenomena that encode, embody, and influence research practice (Howison and Herbsleb 2011).

These relationships are currently being investigated but there is further work yet to do in developing theories and concepts of collaboration in e-Science. For example, some theories have been developed that seek to explain how multiple units of analysis necessarily function simultaneously (Lee et al. 2006), and how the activities of aligning relationships within and across units of analysis serve to accomplish the enactment of infrastructure generally (Bietz et al. 2010; Jackson et al. 2011). Whilst some research has highlighted the importance of understanding infrastructural work as relying on a variety of dynamic collaborative social forms (e.g. networks, teams, groups, virtual and traditional organisations) in order to accomplish work (Lee et al. 2006), other studies have taken the larger endeavour of e-Science development itself as a unit of analysis (Kee and Browning 2010). As e-Science systems proliferate, theorising how infrastructures are created and how existing local work practices align with larger computer systems and the larger social structures necessary to create and maintain them will become increasingly important.

A third challenge and opportunity is to *unpack e-Science practices and what it means to support continuous discovery.* For this we may draw upon work in the social studies of science and technology, specifically the notion of epistemic culture (Knorr-Cetina 1999). Rather than viewing science simply as a heterogeneous set of practices, this approach takes as its starting point the ways in which scientific knowledge is produced, the types of experiments conducted, and the relations between units in a field. Whilst initially developed to investigate experimental science, researchers are now exploring the applicability of the notion of epistemic culture to knowledge production in areas across the sciences, social sciences and humanities (Wouters and Beaulieu 2006). Furthermore, in an area where continuous discovery is the norm, socio-technical practices are also continuously evolving. The advent of ever-larger datasets and the increasing availability and accessibility of computational power is set to transform how science is done. It has been suggested that in the future quantitative analyses of large data sets will be used to generate theory (Hey et al. 2009). The proliferation of specialised scripts and software tools needed to process and analyse data is also making reproducibility of scientific results more challenging. New practices around the collection, analysis, and sharing of data claim to be changing the scientific method itself. If in research, a state of controlled change is the preferred situation, how does that influence how we think about socio-technical systems and how we design them? There is an increasing need to understand the ways in which these developments may or may not be supporting discovery and scientific advancement by supporting or not supporting changes in practice that are frequently unpredictable. The challenge here lies in being able to construct conceptual models of change that can inform the design and development of information systems that are stable enough to support current research needs yet flexible enough to accommodate expanding scientific frontiers. The conceptual challenges of expanding frontiers manifest themselves not only in the opportunities afforded by new technologies and techniques, but also, more centrally from the perspective of research, the changing objects of research themselves (Ribes and Polk 2012).

The final conceptual challenge mentioned above lies in *supporting and learning from the particularities of specific scientific disciplines and sub-disciplines while also generating conceptual frameworks and theories that transcend individual disciplines and sub-disciplines*. For example, theories of collaborative socio-technical innovation and theories of system and tool design that we might generate that are applicable across research areas, are likely to have broader applicability to other dynamic collaborative situations. Whilst we have made much of the multiplicity of disciplines, sub-disciplines, and hybrid disciplines that have been under-studied, we also wish to look for conceptual frameworks that might be generalisable across e-Science and other types of collaborative knowledge work.

In science, there are frequent debates concerning scientists working in what have been described as disciplinary silos, making the same discoveries in different fields in parallel or rapid succession, and subsequently missing opportunities to gain perspectives and methodologies that have the potential to be transformative. Similarly, CSCW researchers might consider the ramifications of succumbing to the temptation to work within sub-disciplinary silos. It would be a missed opportunity if CSCW e-Science investigators attended only to other researchers working in their sub-discipline (e.g. medical informatics, astrophysics, humanities). Some analytic purchase may be gained from researching in these diverse, specialised areas. It is necessary to push theoretical understandings such that conceptual gains made in one area can be usefully applied to others. We may also wish to consider how to work not only within, but across, scientific disciplines and sub-disciplines in order to create new conceptual frameworks that could transform not only how we think about and approach the design of collaborative systems for e-Science, but also how we conceptualise other discovery-oriented, dynamic collaborative systems. Such conceptual innovations could then also inform other related fields such as HCI, information systems, organisation science, and STS.

There is sufficient existing research to convince inform system developers about approaching e-Science as a socio-technical endeavour with different types of software as being part of a set of practices that are highly dependent on the particularities of the research problem undertaken by scientists and developers. Yet these theories do not yet sufficiently explain or provide a template or model of infrastructural practices. Future models could potentially go farther towards driving specific design and policy guidelines.

7. Conclusion

In this paper we have reported on research initiatives from the US and UK that are driving forward programmes of work to accelerate and transform the ways in which science is undertaken, with a strong remit to create new forms of science. At the heart of the vision lies collaborative scientific production supported by advanced forms of technology. Initially these programmes of work were almost entirely focussed on technical solutions to difficult complex science problems. In the last few years increasingly, socio-technical researchers have begun to investigate the types of activities that need to occur in order to turn the vision into a reality. This work often suggests ways in which the vision needs to be recalibrated in the light of a more detailed understanding of the sociality of work.

An outstanding issue for the study of e-Science within CSCW is one of self-definition. Many more established fields, CSCW being no exception, have gone through a process of self reflection whereby members of the community argue over what they are and what they are not, and which principles and practices are fundamental and which replaceable. Today, we may not agree on where we are going, but we have a collective sense of where we have been. With nascent e-Science, we barely know where we are at present. Perhaps a more immediate and pressing point is that we currently only have a very general sense of what we mean when we talk about e-Science or a more recent term of interest within CSCW, scientific collaboration. Currently the majority of socio-technical researchers within CSCW gravitate towards larger projects in high profile domains; however there are many types of scientific collaboration that have only barely been studied such as, geophysics, paleobotany (Ribes and Finholt 2009) and metagenomics (Bietz et al. 2010) - and completely unstudied within CSCW (as of October 2012), such as bioengineering, entomology and physical chemistry to just name a few. Future challenges for our research area would be to achieve greater coverage by CSCW research across the dimensional space within the science—technology—sociality triangle, across different placements on the spectrum from development to theory and also across different orientations towards scoping the unit of analysis.

Unlike fields such as science and technology studies, CSCW counts fewer practicing domain researchers (e.g. natural sciences, historians, etc.) within its ranks. However, there has been a gradual influx of researchers interested in science informatics, and research systems and practices into CSCW. Within the loosely aggregated community, research interests range not only from the social to the technical, but also between the social, to the technical, to the scientific. Research interests of individual researchers may map anywhere in the space between the three points of this triangle. And even beyond, researchers will differ greatly in terms of the level of analysis that they are scoping: ranging from micro to macro, from small teams, to local organisations, to distributed organisations, to loosely organised communities.

Concentrated around the sociality corner of the science—technology—sociality triangle, there is a long history and body of knowledge of science and technology studies that "offer commentaries on scientific developments, spanning the full range from top-level critique to detailed analysis and making contributions both to understanding the process of design to designing itself" (pg viii, Hine 2006). There may be some tension, however, in the relationship between STS norms and CSCW norms as design-oriented fields. Many STS researchers may not see it as part of their remit to engage with design; or in some cases are perceived to convey an approach that is anti-science and/or neglectful of the concerns of scientists themselves (Hine 2006). CSCW can learn from and participate in STS while also being committed to critiquing, informing, and engaging with design from the level of developing tools to support cooperation within groups (Grochow et al. 2008) to the level of addressing policy for international e-Science collaborations (Voss et al. 2010).

More research is needed to study emerging e-Science practices, technologies in use, and the impact upon the scientific record. Researchers in the field of STS have been studying aspects of science and scientific collaboration for many years, but have arguably lacked the engagement with design and implications for design of CSCW. Similarly, technical fields have been deeply involved in e-Science, but lack the theory and methods for understanding work practices and for designing and informing the design of, not only relatively simple tools and scripts, but also larger and more complex socio-technical systems. Thus, while there exist a multitude of willing and appropriate partners, the particular perspectives and skills of CSCW researchers are most especially required. In order to meet our future research challenges, it is necessary to educate and prepare new members of our community to undertake investigations in the complexities of e-Science collaboration and the design of collaborative systems for e-Science. We would argue that CSCW, with its history as a forum for discussing research from various disciplines including computer science, informatics, organisational sciences and social sciences - and as a leading field in the presentation of work in understanding the social in computing - is ideal for providing the space for the necessary translational work to occur.

Acknowledgments

We would like to thank Drew Paine, Paul Luff and the anonymous reviewers for their constructive and insightful comments on previous drafts of this paper. Marina Jirotka has been supported by the UK Engineering and Physical Sciences Research Council (EPSRC) and by the UK Economic and Social Research Council (ESRC). Charlotte Lee has been supported in part by US National Science Foundation Grant IIS-0954088. Gary Olson has been supported by the Army Research Institute grant W91WAW-07-C-0060, the National Science Foundation grants OCI 1025769 and IIS 0085951, Google, and the Donald Bren Foundation.

References

Ackerman, Mark S., Brian Starr, Debby Hindus and Scott D. Mainwaring (1997). Hanging on the 'wire: a field study of an audio-only media space. ACM Transactions on Computer-Human Interaction (TOCHI) 4, no. 1, pp. 39–66.

Ackerman, Mark. (2000). The Intellectual Challenge of CSCW: The Gap between Social Requirements and Technical Feasibility. *Human-Computer Interaction*. vol. 15, nos. 2-3. September 2000, pp. 179–203.

Aragon, Cecilia. R., Sarah Poon, Gregory S. Aldering, Rollin C. Thomas, Robert Quimby (2009). Using Visual Analytics to Develop Situation Awareness in Astrophysics. *Information Visualization,* vol. 8, no 1, March 2008, pp. 30–41

Arzberger, Peter and Thomas A. Finholt (2002). *Data and Collaboratories in the Biomedical Community.* Final report of an NIH panel of experts meeting held September 16–18, 2002 in Ballston, Va. http://nbcr.sdsc.edu/Collaboratories/CollaboratoryFinal2.doc.

Atkins, Daniel. E, Christine Borgman, Nathan Bindoff, Mark Ellisman, Stuart Feldman, Ian Foster, Albert Heck, Dieter Heermann, Julia Lane, Luciano Milanesi, Jayanth Paraki, Alexander Szalay, Pauk Tackley, Wolfgang von Ruden, Han Wensink and Andres Ynnerman,. *Building a UK Foundation for the Transformative Enhancement of Research and Innovation*. Report of the International Panel for the 2009 Review of the UK Research Councils e-Science Programme. Technical report, Research Councils UK, March 2010. http://www.epsrc.ac.uk/research/intrevs/escience/Pages/default.aspx

Atkins, Daniel. E., Kelvin K. Droegemeier, Stuart I. Feldman, Hector Garcia-Molina, Michael L. Klein, David G. Messerschmitt, Paul Messina, Jeremiah P. Ostriker, Margaret H. Wright (2003). *Revolutionizing Science and Engineering Through Cyber-infrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyber-infrastructure*. Washington, DC: National Science Foundation. http://www.nsf.gov/od/oci/reports/toc.jsp

Bailey, Helen, Michelle Bachler, Simon Buckingham Shum, Anja Le Blanc, Sita Popat, Andrew Rowley and Martin Turner (2009). Dancing on the Grid: Using e-Science Tools to Extend Choreographic Research. *Philosophical Transactions of the Royal Society A,* vol. 367, no. 1898, July 2009, pp. 2793–2806.

Bellotti, Victoria, and Paul Dourish (1997). Rant and RAVE: experimental and experiential accounts of a media space. In K. Finn, A. Sellen and S. Wilbur (Eds) Video-mediated communication (1997) Lawrence Erlbaum, pp. 245–272.

Bennett, Michelle, Howard Gadlin, and Samantha Levine-Findley (2010). *Team Science and Collaboration: A Field Guide.* National Institutes of Health. August 2010 http://teamscience.nih.gov

Berg, Marc and Els Goorman (1999). The Contextual Nature of Medical Information. *International Journal of Medical Informatics*, vol. 56, no. 1–3, December 1999, pp. 51–60.

Berman, Francine and Henry Brady (2005). *Final report NSF SBE-CISE workshop on cyber-infrastructure and the social sciences*. National Science Foundation. http://vis.sdsc.edu/sbe/reports/SBE-CISE-FINAL.pdf

Berman, Francine, Geoffrey Fox, and Hey, J.G. Anthony (2003). The Grid: past, present, future. In F. Berman, G. Fox, and A. J.G. Hey (eds): *Grid Computing: Making The Global Infrastructure a Reality*. Chichester: Wiley, pp.9–50.

Bietz, Matthew. J. and Charlotte. P. Lee (2009). Collaboration in Metagenomics: Sequence Databases and the Organization of Scientific Work. In E. Balka, L. Ciolfi, C. Simone, H. Tellioğlu & I. Wagner (eds): ECSCW 2009: *Proceedings of the 11th European Conference on Computer Supported Cooperative Work, 7–11 September 2009*, Vienna, Austria. London: Springer-Verlag. pp. 243–262.

Bietz, Matthew J., Eric P.S. Baumer, and Charlotte P. Lee (2010). Synergizing in Cyberinfrastructure Development. *Computer Supported Cooperative Work (CSCW)*, *An International Journal,* vol 19, no 3–4, September 2010, pp 245–81.

Bietz, Matthew J., Toni Ferro, and Charlotte P. Lee (2012) Sustaining the development of cyberinfrastructure: an organization adapting to change. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, ACM, 2012, pp. 901–910.

Bietz, Matthew J., Drew Paine and Charlotte P. Lee (2013). The Work of Developing Cyberinfrastructure Middleware Projects. In *proceedings of ACM Conference on CSCW '13*, San Antonio, TX.

Birnholtz Jeremy, P. and Matthew, J. Bietz (2003) Data at Work: Supporting Sharing in Science and Engineering. In: M. Pendergast (ed): *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work (GROUP'03) 2003, November 9–12,*: ACM Press; 2003. p. 339–48.

Bly, Sara A., Steve R. Harrison, and Susan Irwin (1993). Media spaces: bringing people together in a video, audio, and computing environment. Communications of the Association of Computing Machinery NY. Vol. 36, No. 1, January 1993, pp. 28–46.

Borda, Ann, Jason Careless, Maia Dimitrova, Michael Fraser, Jeremy Frey, Paul Hubbard, Stéphane Goldstein, Caroline Pung, Michele Shoebridge and Norman Wiseman (2006). *Report of the Working Group on Virtual Research Communities for the OST e-Infrastructure Steering Group*. London, UK, Office of Science and Technology, pp.46.

Borgman, Christine (2007). *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. MIT Press.

Bos, Nathan, N. Sadat Shami, Judith S. Olson, Arik Cheshin, and Ning Nan (2004). In-group/out-group effects in distributed teams: an experimental simulation. In Proceedings of the 2004 ACM conference on Computer supported cooperative work, ACM, 2004, pp. 429–436.

Bos, Nathan, Ann Zimmerman, Judith Olson, Jude Yew, Jason Yerkie, Erik Dahl and Gary Olson (2007). From Shared Databases to Communities of Practice: A Taxonomy of Collaboratories. *Journal of Computer-Mediated Communication,* vol 1, no. 2, pp. 652–672.

Bowker, Geoffrey C. (2000). Biodiversity datadiversity. *Social Studies of Science*, vol 30. no 5, October 2000, pp 643–683.

Bowker, Geoffrey. C. and Susan Leigh Star (1999). *Sorting Things Out: Classification and Its Consequences*. MIT Press: Cambridge, MA.

Brady, J. Michael, Fiona Gilbert, Sharon Lloyd, Marina Jirotka, David J. Gavaghan, Andrew Simpson, Ralph P. Highnam, Tom Bowles, David Schottlander, David McCabe, David Watson, Brian Collins, John Williams, Alan Knox, Manfred Oevers, and Paul Taylor (2004). e-DiaMoND: the UK's national digital mammography database. *International Workshop on Digital Mammography (IWDM)* 18–21 June 2004, Durham, North Carolina, USA) 2004.

Brubaker, Jed. R. and Gillian.R. Hayes (2011). SELECT ∗ FROM USER: Infrastructure and Socio-technical Representation. In the P. J. Hinds, J. C. Tang, J. Wang, J. E. Bardram, N. Ducheneaut

(Eds.): *Proceedings of the 2011 ACM Conference on Computer Supported Cooperative Work, CSCW 2011, Hangzhou, China, March 19–23, 2011*. ACM Press, pp. 369–378.

Bulick, Stephen, Mark Abel, Doug Corey, Jill Schmidt, and Stephen Coffin (1989). The US West Advanced Technologies prototype multimedia communication system. In Global Telecommunications Conference and Exhibition. Communications Technology for the 1990s and Beyond. GLOBECOM '89., IEEE, pp. 1221–1226.

Buxton, William, and Tom Moran (1990). EuroPARC's Integrated interactive intermedia facility (IIIF): early experiences. In Proceedings of the IFIP WG, vol. 8, pp. 11–34.

Buxton, William AS, Abigail J. Sellen, and Michael C. Sheasby (1997) Interfaces for multiparty video conferences. In K. Finn, A. Sellen and S. Wilbur (Eds) Video-mediated communication (1997) Lawrence Erlbaum, pp. 385–400.

Carusi, Annamaria and Marina Jirotka (2010). Reshaping Research Collaboration: the case of Virtual Research Environments. In W.H. Dutton, and P.W. Jeffreys (eds): *World Wide Research: Reconfiguring Access to Information, Expertise and Experience in the Sciences and Humanities*. Cambridge, MA: MIT Press, pp 277–294.

Carusi, Annamaria and Marina Jirotka (2009). From data archive to ethical labyrinth. *Qualitative Research*, vol 9, no. 3, July 2009, pp. 285–298.

Carusi, Annamaria and Torsten Reimer (2010). *Virtual research environment collaborative landscape study.* 31st December 2010. JISC Report. http://www.jisc.ac.uk/publications/reports/2010/vrelandscapestudy.aspx

Childers, Lisa, Terry Disz, Robert Olson, Michael Papka, Rick Stevens, and Tushar Udeshi (2000). Access Grid: Immersive Group-to-Group Collaborative Visualization. *Proceedings of the 4th International Immersive Projection Technology Workshop, Ames IA, June 19–20, 2000.*

Coopmans, Catelijne (2006). Making mammograms mobile: Suggestions for a sociology of data mobility. *Information, Communication & Society*, vol 9, no 1, February 2006, pp. 1–19.

Corbin, Juliet and Anselm. Strauss (2007). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, CA: Sage Publications.

Corrie, Brian and Todd Zimmerman (2009). Build It: Will They Come? Media Spaces in support of Computational Science. In S. Harrison, (Ed): *Media space: 20+ years of mediated life*. London: Springer-Verlag, 2009, pp. 393–413.

Crabtree, Andy, Andrew French, Chris Greenhalgh, Steve Benford, Keith Cheverst, Dan Fitton, Mark Rouncefield and Connor Graham (2006). Developing digital records: Early experiences of record and replay. Special Issue of *Computer Supported Cooperative Work (CSCW): An International Journal*, (eds): M Jirotka, R. Procter, T. Rodden and G. Bowker. Collaboration in e-Research, vol. 15, no. 4, August 2006, pp. 281–319.

Cragin, Melissa and Kalpana Shankar (2006). Scientific Data Collections and Distributed Collective Practice. *Computer Supported Cooperative Work (CSCW): An International Journal*, vol.15, no. 2, June 2006, pp. 185–204.

Cummings, Jonathon. N., & Sara Kiesler (2008). Who collaborates successfully? Prior experience reduces collaboration barriers in distributed interdisciplinary research. In *Proceedings of the 2008 ACM conference on computer supported cooperative work* (*CSCW '08)* ACM New York, NY, USA, pp. 437–446.

Cummings, Jonathon, N., and Sara Kiesler (2005). Collaborative Research Across Disciplinary and Institutional Boundaries. *Social Studies of Science,* vol. 35 no. 5, October 2005, pp. 703–722.

D'Agostino, Giuseppina., Hinds Christopher, Marina Jirotka, Charles Meyer, Tina Piper, Mustafizur Rahman and David Vaver (2008). On the Importance of Intellectual Property Rights for e-Science and the Integrated Health Record. *Health Informatics Journal,* vol. 14 no 2, June 2008, pp 95–111.

David, Paul A. (2006) Towards a cyberinfrastructure for enhanced scientific collaboration: "Proving its soft foundations may be the hardest part" In B. Kahin and D. Foray (Eds): *Advancing Knowledge and the Knowledge Economy*. MIT Press, pp. 431–454.

David, Paul. A. and Michael Spence (2010). Institutional Infrastructures for Global Research Networks in The Public Sector. In (eds): W.H. Dutton and P.W. Jeffreys *World Wide Research: Reconfiguring Access to Information, Expertise and Experience in the Sciences and Humanities*. Cambridge, MA: MIT Press, 2010, pp. 191–213.

de la Flor, Grace, Marina Jirotka, Paul Luff, John Pybus and Ruth Kirkham (2010a). Transforming scholarly practice: Embedding technological interventions to support the collaborative analysis of ancient texts. *Computer Supported Cooperative Work (CSCW)*, *An International Journal,* vol. 19, nos. 3–4, August 2010, pp. 309–334.

de la Flor, Grace, Paul Luff, Marina Jirotka, John Pybus, Ruth Kirkham and Annamaria Carusi (2010b). The Case of the Disappearing Ox: Seeing Through Digital Images to an Analysis of Ancient Texts. In *Proceedings of 28th ACM Conference on Human Factors in Computing Systems, CHI 2010,* Atlanta GA, USA, pp. 473–482.

de Roure, David, Carol Goble and Richard Stevens (2007) Designing the myExperiment Virtual Research Environment for the Social Sharing of Workflows In *e-Science 2007: Proceedings of the Third IEEE International Conference on e-Science and Grid Computing,* December 2007, Bangalore, India. IEEE Computer Society, pp. 603–610.

de Roure, David, Mark Baker, Nick Jennings and Nigel Shadbolt (2003). The evolution of the Grid. In, F. Berman, G. Fox, and A.J.G. Hey (eds): Grid Computing - Making the Global Infrastructure a Reality. John Wiley and Sons Ltd, pp. 65–100.

Dourish, Paul (2007) Responsibilities and Implications: further thoughts on ethnography and design. In *Proceedings of the 2007 conference on Designing for User eXperiences.* November 05–07, 2007, Chicago Illinois ACM.

Dourish, Paul, and Sara Bly (1992). Portholes: Supporting awareness in a distributed work group. In Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, 1992, pp. 541–547.

Dourish, Paul, and Victoria Bellotti (1992). Awareness and coordination in shared workspaces. In Proceedings of the 1992 ACM conference on Computer-supported cooperative work, ACM, 1992, pp. 107–114.

Dourish, Paul, Annette Adler, Victoria Bellotti, and Austin Henderson (1996). Your place or mine? Learning from long-term use of audio-video communication. Computer Supported Cooperative Work (CSCW), An International Journal Vol 5, no. 1, pp. 33–62.

Dovey, Matthew (2010) The Future of Virtual Research Environments. In W.H. Dutton, and P.W. Jeffreys, (eds): *World Wide Research: Reconfiguring Access to Information, Expertise and Experience in the Sciences and Humanities*. Cambridge, MA: MIT, 2010, Press, pp. 295–298.

Dutton, William. H. and Paul Jeffreys (eds): (2010) *World Wide Research: Reshaping the Sciences and Humanities.* Cambridge, MA: The MIT Press.

Eden, Grace. (2011). *The Context Evaluation Framework: A prototype evaluation technique for e-Research.* Doctoral Thesis. University of Oxford, UK. http://ora.ox.ac.uk/objects/uuid:e816cf1a-3514-400c-9f2f-e022b56710d9

Eden, Grace and Marina Jirotka (2012) Digital images of medieval music documents: transforming research processes and knowledge production in musicology. In 45th Hawaii International Conference on System Sciences (HICSS), January 4–7, 2012, Maui HI. 2012.

Eden, Grace, Marina Jirotka and Eric Meyer (2012) Interpreting Digital Images Beyond Just the Visual: Crossmodal Practices in Medieval Musicology. Interdisciplinary Science Reviews, vol. 37, no.1, March 2012, pp. 69–85.

Edwards, Paul. N., Geoffrey C., Bowker, Steven J. Jackson, and Robin Williams (eds) (2009). Special Issue on eInfrastructure. Journal of the Association for Information *Systems*, May 2009, vol. 10, no.5.

Edwards, Paul, N., Matthew S. Mayernick, Archer Batcheller, Geoffrey Bowker and Christine Borgman (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science,* vol. 41, no. 5, October 2011, pp. 667–690.

Falk-Krzesinski Holly, J., Noshir Contractor, Stephen, M. Fiore, Kara L. Hall, Cathleen Kane, Joann Keyton, Julie, T. Klein, Bonnie Spring, Daniel Stokols and William Trochim. (2011). Mapping a research agenda for the science of team science. *Research Evaluation,* vol. 20, no. 2, 2011, pp145-158.

Faniel, Ixchel M., and Ann Zimmerman (2011). Beyond the data deluge: A research agenda for large-scale data sharing and reuse. *The International Journal of Digital Curation*, vol. 6, no. 1, 2011, pp 58–69.

Faniel, Ixchel M., and Trond E. Jacobson (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work (CSCW)*, *An International Journal*, vol. 19, nos. 3–4, August 2010, pp 355–375.

Finholt, Thomas, A. (2002) Collaboratories. *Annual Review of Information Science and Technology*, vol. 36, no.1, 2005, pp. 73–107.

Finholt, Thomas, A. (2003). Collaboratories as a new form of scientific organization. *Economics of Innovation and New Technology*, vol. 12, no 1, 2003, pp. 5–25.

Finholt, Thomas. A. and Gary M. Olson (1997). From Laboratories to Collaboratories: A New Organizational Form for Scientific Collaboration. *Psychological Science,* vol. 8, no. 1, January 1997, pp. 28–36.

Fish, Robert S., Robert E. Kraut, and Barbara L. Chalfonte (1990). The VideoWindow system in informal communication. In Proceedings of the ACM conference on Computer-supported cooperative work. pp. 1–11. ACM, 1990.

Foster, Ian and Carl Kasselman (2004). *The Grid 2: Blueprint for a New Computing Infrastructure*. Morgan Kaufman.

Foster, Ian, Carl Kesselman and Steven Tueke (2001). The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *International Journal of Supercomputer Applications,* vol. 15, no. 3, Fall 2001, pp. 200–222.

Fraser, Michael (2004). The same old remote misunderstandings: Object-Focused Interaction in e-Social Science, workshop paper for Giving Help at a Distance' Workshop at the *Sixth International Conference on Ubiquitous Computing*, Nottingham, September, 2004.

Fraser, Michael, Jon Hindmarsh, Katie Best, Christian Heath, Greg Biegel, Chris Greenhalgh and Stuart Reeves (2006). Remote collaboration over video data: Towards real-time e-Social Science. *Computer Supported Cooperative Work (CSCW)*, *An International Journal*, vol. 15, no. 4, August 2006, pp. 257–279.

Frey, Jeremy, Hugo Mills, Gareth Hughes, Jamie Robinson, David de Roure, mc schraefel and Luc Moreau (2005). Semantic Support for Smart Laboratories. *229th American Chemical Society (ACS) National Meeting,* San Diego, USA, 13–17 Mar 2005, USA, American Chemical Society.

Fry, Jenny (2006). Coordination and Control of Research Practice across Scientific Fields: Implications for a Differentiated e-Science. In C Hine (ed): *New infrastructure for knowledge production: understanding e-Science* (2006). Information Science Publishing, pp. 167–188.

Galegher, Jolene, Robert E. Kraut, and Carmen Egido (Eds) (1990) Intellectual teamwork: Social and technological foundations of cooperative work. Lawrence Erlbaum Associates, Inc, 1990.

Garfinkel, Harold (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall, 1967.

Gaver, William, Thomas Moran, Allan MacLean, Lennart Lövstrand, Paul Dourish, Kathleen Carter, and William Buxton (1992). Realizing a video environment: EuroPARC's RAVE system. In Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, 1992, pp. 27–35.

Gaver, William, Abigail Sellen, Christian Heath and Paul Luff (1993). One is not enough: Multiple Views in a Media Space, in *Proceedings of INTERCHI '93* April 1993, pp. 335–341.

Geiger, R. Stuart and David Ribes (2010). The Work of Sustaining Order in Wikipedia: The Banning of a Vandal. In *CSCW '10 Proceedings of the 2010 ACM conference on Computer supported cooperative work.* Savannah, Georgia, USA: ACM, pp. 117–126.

Gerson, Elihu M. (2008). Reach, bracket, and the limits of rationalized coordination: Some challenges for CSCW. In M.S. Ackerman, C, Halverson, A. T Erickson,. and. W. A Kellogg, (Eds): *Resources, Co-Evolution and Artifacts: Theory in CSCW.* London: Springer. pp. 193–220.

Grochow, Keith, Mark Stoermer, Deb Kelley, John Delaney, and Ed Lazowska (2008). Cove: A visual environment for ocean observatory design. In Journal of Physics: Conference Series, vol. 125, no. 1. IOP Publishing, 2008.

Grochow, Keith, Mark Stoermer, James Fogarty, Charlotte Lee, Bill Howe, Ed Lazowska (2010). COVE: A Visual Environment for Multidisciplinary Ocean Science Collaboration. In *IEEE Sixth International Conference on eScience* 2010, Brisbane, Australia, pp. 269–76.

Halfpenny Peter and Rob Procter (2010). The e-Social Science research agenda. *Philosophical Transactions of the Royal Society A,* vol. 368, 2010, pp. 3761–3778.

Harrison, Steve (ed) (2009). *Media space: 20+ years of mediated life*. London: Springer-Verlag.

Harrison, Steve, and Paul Dourish (1996). Re-place-ing space: the roles of place and space in collaborative systems. In Proceedings of the 1996 ACM conference on Computer supported cooperative work, ACM, 1996, pp. 67–76.

Harrison, Steve, Sara Bly, Susan Anderson, and Scott Minneman (1997). The media space. In K.Finn, A. Sellen and S. Wilbur (Eds) Video-mediated communication. Lawrence Erlbaum, 1997 pp. 273–300.

Hartswood, Mark, Rob Procter, Mark Rouncefield and Roger Slack (2003). Making a Case in Medical Work: Implications for the Electronic Medical Record. *Computer Supported Cooperative Work (CSCW), An International Journal*, vol. 12, no. 3, 2003, pp. 241–266.

Hartswood, Mark, Marina Jirotka, Rob Procter, Roger Slack, Alex Voss and Sharon Lloyd (2005). Working IT out in e-Science: Experiences of requirements capture in a HealthGrid project. In *Proceedings of the HealthGrid Conference* 2005, IOS Press. pp 234.238.

Hartswood, Mark, Rob Procter, Paul Taylor, Lilian Blot, Stuart Anderson, Mark Rouncefield and Roger Slack (2012). Problems of Data Mobility and Reuse in the Provision of Computer-based Training for Screening Mammography. In *CHI '12 Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems* ACM New York, NY, USA, Pp. 909–918.

Haythornthwaite, Caroline, Karen J., Lunsford, Geoffrey Bowker and Bertram Brude (2006). Challenges for Research and Practice in Distributed, Interdisciplinary Collaborations. In Hine, C (ed): *New infrastructure for knowledge production: understanding e-Science*. 2006, Information Science Publishing, pp 143–166.

Heath, C., and Paul Luff (1992). Media space and communicative asymmetries: Preliminary observations of video mediated interaction. *Human Computer Interaction*, vol. 7, no 3, pp. 99–103.

Hey, Anthony, and Anne Trefethen (2003a). e-Science and its Implications. *Philosophical Transactions of the Royal Society A,* vol. 361, pp. 1809–1825.

Hey, Anthony and Anne Trefethen (2005). Cyberinfrastructure for e-Science. *Science* May 2005, vol. 308, no. 5723, pp. 817–821.

Hey, Anthony and Anne Trefethen (2002). The UK e-Science core programme and the Grid. *Future Generation Computing Systems*, vol. 18, no. 8, pp. 1017–1031.

Hey, Anthony and Anne Trefethen (2003b). The data deluge. In Berman, F. Fox G. and Hey, A. (eds): *Grid computing: Making the global infrastructure a reality.* Chichester, England: Wiley, pp. 809–824.

Hey, Anthony, Stewart Tansley, and Kristin Tolle (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research.

Hine, Christine (ed) (2006) *New infrastructure for knowledge production: understanding e-Science*. Information Science Publishing.

Hine, Christine (2007a). Connective ethnography for the exploration of e-science. *Journal of Computer-Mediated Communication*, vol. 12, no. 2, article 14.

Hine, Christine (2007b). Multi-sited ethnography as a middle range methodology for contemporary STS. *Science, Technology and Human Values,* vol. 32, no. 6, pp. 652–671.

Hoffa, Christina, Guarang Mehta, Timothy Freeman, Ewa Deelman, Kate Keahey, Bruce Berriman, and John Good (2008). On the Use of Cloud Computing for Scientific Workflows. In *Proceedings of IEEE Fourth International Conference on e-Science* 2008 Indianapolis, USA, pp. 640–645.

Howison, James and James Herbsleb (2011). Scientific software production: incentives and collaboration. *In Proceedings of the ACM Conference on CSCW.* Hangzhou, China, ACM: pp. 513–522.

Huang, Xing, Xianghua (Sharon) Ding, Charlotte P. Lee, Tun Lu and Ning Gu (forthcoming 2013). Meanings and Boundaries of Scientific Software Sharing. *In the Proceedings of the Conference on CSCW '13*, San Antonio, TX, ACM.

Hughes, Gareth, V., Hugo R. Mills, Graham Smith, Terry R. Payne, and Jeremy Frey (2005). Breaking the book: translating the chemistry lab book into a pervasive computing lab environment. *In Proceedings of the SIGCHI conference on Human factors in computing systems,* ACM, 2004, pp. 25–32.

Hutchins, Edwin (1995). *Cognition in the Wild*. MIT Press: Cambridge, MA.

Ishii, Hiroshi (1990) TeamWorkStation: towards a seamless shared workspace. In Proceedings of the 1990 ACM conference on Computer-supported cooperative work, ACM, 1990, pp. 13–26.

Ishii, Hiroshi, and Minoru Kobayashi. (1992). ClearBoard: a seamless medium for shared drawing and conversation with eye contact. In Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, 1992, pp. 525–532.

Jackson, Steven, David Ribes, Ayse Buyuktur and Geoffrey Bowker (2011). Collaborative Rhythms: Temporal Dissonance and Alignment in Distributed Scientific Work. *In Proceedings of the 2011 Computer-Supported Cooperative Work (CSCW) Conference*, Hangzhou, China, March 20—23, 2011.

Jankowski, Nicholas W. (2007). Exploring e-Science: an introduction. *Journal of Computer-Mediated Communication,* vol. 12, no 2, January 2007, pp 549–62.

Jankowski, Nicholas W. (Ed) (2009): *e-Research: Transformation in Scholarly Practice*. Routledge.

Jirotka, Marina, Rob Proctor, Tom Rodden, and Geoffrey Bowker (eds) (2006). Special Issue: Collaboration in e-Research. *Computer Supported Cooperative Work (CSCW)*, *An International Journal*, vol. 15, no 4, August 2006, pp. 251–255.

Jirotka, Marina, Rob Procter, Mark Hartswood, Roger Slack, Catelijne Coopmans, Christopher Hinds and Alex Voss (2005). Collaboration and Trust in Healthcare Innovation: the eDiaMoND Case Study. *Computer Supported Cooperative Work (CSCW)*, *An International Journal*, vol. 14, no.4, August 2005, pp. 369–389.

Kandel, Sean, Jeffrey Heer, Catherine Plaisant, Jessie Kennedy, Frank van Ham, Nathalie H. Riche, Chris Weaver, Bongshin Lee, Dominique Brodbeck, and Paolo Buono (2011). Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization,* vol. 10, no 4, October 2011, pp. 271–288.

Karasti, Helena, Karen Baker and Florence Millerand (2010). Infrastructure time: Long-term matters in collaborative development. *Computer Supported Cooperative Work (CSCW)*, *An International Journal*, vol. 19, no. 3–4, pp. 377–415.

Karasti, Helena, Karen Baker, K. and Eija Halkola (2006). Enriching the Notion of Data Curation in E-Science: Data Managing and Information Infrastructuring in the Long Term Ecological

Research (LTER) Network. *Computer Supported Cooperative Work (CSCW)*, *An International Journal*, vol. 15, no 4 August 2006, pp. 321–358.

Kee, Kerk and Larry Browning (2010). The Dialectical Tensions in the Funding Infrastructure of Cyberinfrastructure. *Computer Supported Cooperative Work (CSCW)*, *An International Journal*, vol 19, no 3, August 2010, pp 283–308.

Keptelinin, Victor and Bonnie A. Nardi (2006). *Acting With Technology: Activity Theory and Interaction Design*. Cambridge, MA: MIT Press.

Kim, Isabelle, Lora Oehlberg, and Ashley Rayner (2006). FlutterbyNet: Distributed Logbook Collaboration. Extended Abstract for Poster Session at Ubicomp 2006.

Kim, Sunyoung, Christine Robson, Thomas Zimmerman, Jeffrey Pierce and Eben M. Haber (2011). Creek watch: pairing usefulness and usability for successful citizen science. In *Proceedings of the 2011 annual conference on Human factors in computing systems* Vancouver, BC, Canada: ACM, pp 2125–34.

Klein, Julie Thompson (2010). *Creating interdisciplinary campus cultures: A model for strength and sustainabilit.y.* San Francisco, CA, USA: Jossey Bass/The Association of American Colleges and Universities.

Knorr-Cetina, Karen (1999). *Epistemic Cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.

Kraut, Robert E., Ronald E. Rice, Colleen Cool, and Robert S. Fish (1994). Life and death of new technology: task, utility and social influences on the use of a communication medium. In Proceedings of the 1994 ACM conference on Computer supported cooperative work, ACM, 1994, pp. 13–21.

Latour, Bruno (1986). *Science in Action: How to Follow Scientists and Engineers Through Society.* Harvard University Press.

Latour, Bruno (1987). *Science In Action.* Cambridge, MA: Harvard University Press.

Lee, Charlotte P. (2007). Boundary negotiating artifacts: Unbinding the routine of boundary objects and embracing chaos in collaborative work. *Computer Supported Cooperative Work: An International Journal,* vol. 16, no. 3, pp. 307–339.

Lee, Charlotte, P., Paul Dourish and Gloria Mark (2006). The Human Infrastructure of Cyberinfrastructure. In *CSCW '06. Proceedings of the Conference on CSCW,* Banff, Alberta, Canada: ACM, pp. 483–92.

Lee, Charlotte P., David Ribes, Matthew Bietz, Helena Karasti and Marina Jirotka (eds): (2010) Special Issue: The Human Infrastructure of Cyberinfrastructure *Computer Supported Cooperative Work (CSCW): An International Journal. (CSCW),* vol. 19 no 3.

Lee, Charlotte P., Matthew, J. Bietz, Katie Derthick and Drew Paine (2012). A Sociotechnical Exploration of Infrastructural Middleware Development. In *CSCW '12 Proceedings of the ACM 2012 conference on Computer Supported Cooperative. ACM New York, NY, USA* Austin, TX, USA: ACM. *Work, pp. 1347–1350.*

Lindtner, Silvia, Ken Anderson and Paul Dourish (2007). Multi-Sited Design: An Approach Towards Addressing Design-Use Relations in Transnational Processes Position Paper for the *CHI 2011 Workshop* on Transnational HCI. Humans, Computers and Interactions in Transnational Context.

Luff, Paul, Jon Hindmarsh and Christian Heath (2000). *Workplace studies: recovering work practice and informing system design.* Cambridge University Press.

Luo, Tiejian, Wei Liu, Su Chen, Jinliang Song, Yuli Jin, and Cheng Du (2008). Utilizing grid to build cyberinfrastructure for biosafety laboratories. In *12th International Conference on Computer Supported Cooperative Work in Design, 2008. (CSCWD 2008),* pp. 501–506.

Luther, Kurt, Scott Counts, Kristin B. Stecher, Aaron Hoff, Paul Johns (2009). Pathfinder: an online collaboration environment for citizen scientists. In *CHI '09 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* ACM New York, NY, USA pp. 239–48.

Mackay, Wendy .E., Guillaume Pothier, Catherine Letonfal and Hans Erik Sørensen (2002). The Missing Link: Augmenting Biology Laboratory Notebooks. In *Proceedings of the. ACM Symposium on User Interface Software and Technology (UIST 2002),* vol. 4, no. 2 of CHI Letters, Paris, France, ACM Press, pp 41–50.

Mantei, Marilyn M., Ronald M. Baecker, Abigail J. Sellen, William A. S. Buxton, Thomas Milligan, and Barry Wellman (1991). Experiences in the use of a media space. In Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology, ACM, 1991, pp. 203–208.

Mark, Gloria, Steve Abrams and Nayla Nassif (2003). Group-to-Group Distance Collaboration: Examining the "Space Between" *ECSCW'03. In Proceedings of the eighth conference on European Conference on Computer Supported Cooperative Work* Kluwer Academic Publishers Norwell, MA, USA 2003 pp 99–118.

Mascord, Matthew, Marina Jirotka and Annamaria Carusi (2006). Putting pen to paper: capturing the integrative biologist's paper trail. Poster, eScience All Hands Meeting 2006, 19 September 2006.

Middleton, David and Engeström Yrjo (eds) (1998): *Cognition and Communication at Work*. Cambridge: Cambridge University Press.

Nentwich, Michael (2003). *Cyberscience: Research in the age of the internet.* Vienna: Austrian Academy of Science.

O'Day, Vicki, Annette Adler, Allan Kuchinsky and Anna Bouch (2001). When worlds collide: Molecular biology as interdisciplinary collaboration, in W. Prinz, M. Jarke, Y. Rogers, K. Schmidt and V. Wulf (eds.): *Proceedings of the Seventh European Conference on Computer-Supported Cooperative Work*, 16–20 September 2001, Bonn, Germany, 2001 Kluwer Academic Publishers, pp. 399–418

Oleksik, Gerard, Natasa Milic-Frayling and Rachel Jones (2012). Beyond data sharing: artifact ecology of a collaborative nanophotonics research centre. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. Seattle, Washington, USA, ACM: 1165–1174.

Olson, Gary. M., Ann Zimmerman and Nathan Bos (eds) (2008a) *Scientific Collaboration on the Internet*. Cambridge, MA: MIT Press.

Olson, Judith, S., Erik, C. Hofer, Nathan Bos, Ann Zimmerman, Gary, M. Olson, Daniel Cooney and Ixchel Faniel (2008b). A Theory of Remote Scientific Collaboration. In G. M. Olson, A. Zimmerman and N. Bos (eds): *Scientific Collaboration on the Internet*. Cambridge, MA: MIT Press, pp. 73–97.

Palmer, Carole L. (2001). *Work at the boundaries of science: Information and the interdisciplinary research process* . Springer Netherlands.

Parker, John, N., Niki Vermeulen, and Bart Penders (2010) (eds): *Collaboration in the New Life Sciences*. Ashgate Publishing Company 2010.

Pierce, Marlon, E., Geoffrey Fox, Huapeng Yuan and Yu Deng (2006). Cyberinfastructure and Web 2.0. In *Proceedings of HPC 2006, Cetaro, Italy.* July 2006.

Pila, Justine (2009). Authorship and e-Science: Balancing Epistemological Trust and Skepticism in the Digital Environment. *Social Epistemology*, vol. 23, no. 1, pp. 1–24.

Pipek, Volkmar and Wulf, Volker (2009). Infrastructuring: Toward an Integrated Perspective on the Design and Use of Information Technology, *Journal of the Association for Information Systems*, vol. 10, May 2009, pp. 447–473.

Piper, Tina and David Vaver (2010). Ownership of Medical Images in e-Science Collaborations: Learning from the Diagnostic Mammography National Database In W.H. Dutton, and P.W. Jeffreys (eds): *World Wide Research: Reconfiguring Access to Information, Expertise and Experience in the Sciences and Humanities*. Cambridge, MA: MIT Press, pp 214–217.

Pitt-Francis, Joe, Miguel, O. Bernabeu, Jonathan Cooper, Alan Garny, Lee Momtahan, James Osborne, Pras Pathmanathan, Blanca Rodriguez, Jonathan Whiteley and David Gavaghan

(2008). Chaste: Using agile programming techniques to develop computational biology software. *Philosophical Transactions Of the Royal Society A*, vol. 366, pp. 3111–3136.

Poon, Sarah S., Rollin C. Thomas, Cecilia R. Aragon, and Brian Lee (2008). Context-linked virtual assistants for distributed teams: an astrophysics case study. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, ACM New York, NY, USA, 2008, pp. 361–370.

Procter, Rob, Robin Williams, James Stewart, Meik Poschen, Helene Snee, Alex Voss, and Marzieh Asgari-Targhi (2010). Adoption and use of Web 2.0 in scholarly communications. *Philosophical Transactions of the Royal Society A*, vol. 368, no. 1926, 2010, pp. 4039–4056.

Randall, David, Richard Harper and Mark Rouncefield (2007). *Fieldwork for Design: Theory and Practice*. Springer.

Randell, Rebecca, Stephanie Wilson and Peter Woodward (2011). Variations and Commonalities in Processes of Collaboration: The Need for Multi-Site Workplace Studies. *Computer Supported Cooperative Work (CSCW), An International Journal*, vol. 20, no. 1, pp. 37–59.

RCUK Review of e-Science 2009: Building a UK Foundation for the Transformative Enhancement of Research and Innovation.

Ribes, David, and Thomas A. Finholt (2007). Tensions across the scales: planning infrastructure for the long-term. In Proceedings of the 2007 international ACM conference on Supporting group work, ACM, 2007, pp. 229–238.

Ribes, Davd and Thomas Finholt (2009). The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems* vol. 10, no. 5, pp. 375–398.

Ribes, David, and Jessica Beth Polk. Historical ontology and infrastructure. In *Proceedings of the 2012 iConference*, ACM, 2012, pp. 254–262.

Ribes, David and Charlotte P. Lee (2010). Sociotechnical studies of cyberinfrastructure and e-Research: Current themes and future trajectories. *Computer Supported Cooperative Work (CSCW)*, *An International Journal*, vol.19, nos. 3–4, pp. 231–244.

Salayandia, Leonardo, Paulo Pinheiro da Silva, Ann Q. Gates, and Flor Salcedo (2006) Workflow-driven ontologies: An earth sciences case study. In *e-Science'06. Second IEEE International Conference on e-Science and Grid Computing,* IEEE, 2006, Texas USA, pp. 17–17.

Schmidt, Kjeld, and Carla Simonee (1996). Coordination mechanisms: Towards a conceptual foundation of CSCW systems design. *Computer Supported Cooperative Work (CSCW) An International Journal,* vol. 5, nos. 2–3, pp.155–200.

Schmidt, Kjeld and Ina Wagner (2005). Ordering Systems: Coordinative Practices and Artifacts in Architectural Design and Planning. *Computer Supported Cooperative Work* (CSCW) *An International Journal,* vol. 13, pp. 349–408.

Schmidt, Kjeld and Liam, J. Bannon (1992). Taking CSCW Seriously: Supporting Articulation Work, Computer Supported Cooperative Work (CSCW) *An International Journal*, vol. 1, nos. 1–2, pp. 7–40.

schraefel, m. c. and Alan Dix (2009). Within Bounds and Between Domains: Reflecting on Making Tea within the Context of Design Elicitation Methods. *International Journal of Human Computer Studies*, vol. 67, no. 4, pp. 313–323.

Schroeder, Ralph (2008). e-Sciences as research technologies: reconfiguring disciplines, globalizing knowledge. *Social Science Information*, vol. 47, no. 2, pp. 131–157.

Segal, Judith (2009). Software development cultures and cooperation problems: A field study of the early stages of development of software for a scientific community. *Computer Supported Cooperative Work (CSCW) An International Journal*, vol. 18, no. 5, pp. 581–606.

Segal, Judith, and Chris Morris (2008). Developing scientific software. *Software*, IEEE, vol. 25, no. 4, 2008, pp. 18–20.

Semaan, Bryan and Gloria Mark (2011). Technology-mediated social arrangements to resolve breakdowns in infrastructure during ongoing disruption. *ACM Transactions on Computer-Human Interaction (TOCHI),* vol. 18, no. 4, Article No. 21. ACM New York, NY, USA .

Smarr, Larry and Charles E. Catlett (1992). Metacomputing. In F. Berman, G. Fox, and A. Hey, (eds): *Grid Computing: Making The Global Infrastructure a Reality*. Chichester: Wiley, 2003, pp.825–835.

Sonnenwald, Dianne. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology,* vol. 41, no. 1, pp. 643–681.

Sonnenwald, Diane H., Mary C. Whitton, and Kelly L. Maglaughlin. (2003). Evaluating a scientific collaboratory: Results of a controlled experiment. *ACM Transactions on Computer-Human Interaction (TOCHI), vol.* 10, no. 2, ACM New York, NY, USA. 2003, pp. 150–176.

Sonnenwald, Diane H., Kelly L. Maglaughlin, and Mary C. Whitton (2004). Designing to support situation awareness across distances: an example from a scientific collaboratory. *Information Processing & Management*, vol. 40, no. 6, pp. 989–1011.

Spencer, Dimitrina, Ann Zimmerman and David Abramson (eds): (2011) Project Management in E-Science: Challenges and Opportunities. *Computer Supported Cooperative Work (CSCW) An International Journal*, vol. 20, no. 3.

Star, Susan Leigh (1999). The ethnography of infrastructure. American behavioral scientist. Vol. 43, no. 3, pp. 377–391.

Star, Susan Leigh, and James R. Griesemer (1989). Institutional ecology, ‘translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social studies of science*,, vol. 19, no. 3, pp. 387–420.

Star, Susan Leigh, and Karen Ruhleder (1996) Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information systems research*, vol. 7, no. 1, pp. 111–134.

Stevens, Rick (2008). Group Oriented Collaboration: The Access Grid Collaboration System in I. Foster and C. Kesselman (eds): *Grid 2: blueprint for a new computing infrastructure.* Morgan Kaufmann, pp. 191–200.

Stokols, Daniel, Kara L. Hall, Brandie K. Taylor, and Richard P. Moser (2008). The science of team science: overview of the field and introduction to the supplement. *American journal of preventive medicine*, vol. 35, no. 2, pp. 77–89.

Strauss, Anselm (1988). The Articulation of project Work: An Organisational Process. *The Sociological Quarterly*, vol. 29, no. 2, pp. 163–178.

Stults, Robert (1988). Experimental Uses of Video to Support Design Activities (No. SSL-89-19): Xerox PARC.

Tabard, Aurélien, Juan David Hincapié Ramos, and Jakob Bardram (2012). The eLabBench in the wild: supporting exploration in a molecular biology lab. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, ACM, 2012, pp. 3051–3060.

Tang, John C., and Scott Minneman (1991). VideoWhiteboard: video shadows to support remote collaboration. In Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology, ACM, 1991, pp. 315–322.

Tang, John C. (2009). Fast Forward: Applying Media Space Experiences to Current technologies in S. Harrison (ed) (2009). Media space: 20+ years of mediated life. London: Springer-Verlag. pp. 417–424.

Ure, Jenny, Rob Procter, Yu-wei Lin, Mark Hartswood, Stuart Anderson, Sharon Lloyd, Joanna Wardlaw, Horacio Gonzalez-Velez, and Kate Ho (2009). The development of data infrastructures for ehealth: a socio-technical perspective. Journal of the Association for Information Systems, vol. 10, no. 5.

Vann, Katie, and Geoffrey C. Bowker (2006). Interest in Production: On the Configuration of technology-bearing labours for epistemic IT. In C. Hine (Ed): *New infrastructures for knowledge production: Understanding science.* Hershey: Information Science Publishing, pp. 71–97.

Vecchiola, Christian, Suraj Pandey, and Rajkumar Buyya (2009). High-performance cloud computing: A view of scientific applications. *In 10th International Symposium on Pervasive Systems, Algorithms, and Networks (ISPAN,)* Australia 2009, pp. 4–16.

Vertesi, Janet, and Paul Dourish (2011). The value of data: considering the context of production in data economies. *In Proceedings of the ACM 2011 conference on Computer supported cooperative work (CSCW '11).* ACM, New York, NY, USA, pp. 533–542.

Vigder, Mark, Norman, G. Vinson, Janice Singer, Darlene Stewart and Kay Mews (2008). Supporting Scientists' Everyday Work: Automating Scientific Workflows. *IEEE Software*, vol. 25, no. 4., pp. 52–58.

Voss, Alexander and Rob Procter (2009). Virtual research environments in scholarly work and communications. *Library Hi Tech Journal*, vol. 27, no 2, pp. 174–190.

Voss, Alex, Jan Kmunicek, Ludek Matyska, Marco Paganoni, and Simon Lin, (2010). The EUAsiaGrid Roadmap: Paths to a Sustainable, Persistent e-Infrastructure in the Asia-Pacific Region. In: Simon C. Lin and Eric Yen (eds): Data-Driven e-Science: Use Cases and Successful Applications of Distributed Computing Infrastructures. *Proceedings of the 2010 International Symposium on Grid Computing.* Springer, 2010.

Wang, Qi, Xianghua Ding, Tun Lu, Huanhuan Xia, and Ning Gu (2012) Infrastructural experiences: An empirical study of an online arcade game platform in China. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12),* ACM, New York, NY, USA, pp. 583–592.

Warwick, Claire, Claire Fisher, Melissa Terras, Mark Baker, Amanda Clarke, Mike Fulford, Matt Grove, Emma O'Riordan, and Mike Rains (2009). iTrench: A study of user reactions to the use of information technology in field archaeology. *Literary and Linguistic Computing*, vol. 24, no. 2, pp. 211–223.

Welsh, Elaine, Marina Jirotka and David Gavaghan (2006). Post-Genomic Science: multidisciplinary and large-scale collaborative research and its organisational and technological implications for the scientific research process. *Philosophical Transactions of the Royal Society* A, vol. 364 no.1843.

Wiggins, Andre and Kevin Crowston (2010). Developing a Conceptual Model of Virtual Organizations for Citizen Science. *International Journal of Organisational Design and Engineering,* vol 1, no. 1. Pp.148–62.

Williams, Amanda, Ken Anderson, and Paul Dourish. (2008). Anchored mobilities: mobile technology and transnational migration. In *Proceedings of the 7th ACM conference on Designing interactive systems,* ACM New York, NY, pp. 323–332.

Wogan, Peter (2004). Deep Hanging Out: Reflections on fieldwork and multi-sited Andean ethnography. *Identities: Global Studies in Power and Culture,* vol. 11, no. 1, pp. 129–139.

Woollard, David, Nenad Medvidovic, Yolanda Gil, and Chris A. Mattmann (2008). Scientific software as workflows: From discovery to distribution. *Software, IEEE,* vol. 25, no. 4, pp. 37–43.

Wouters, Paul and Anne Beaulieu (2006). Imagining eScience beyond Computation in Hine, C (ed): *New infrastructure for knowledge production: understanding e-Science*. Information Science Publishing, pp 48–70.

Wulf, William, A. (1989). The national collaborator- a white paper. In *Towards a national collaboratory.* Unpublished report of a National Science Foundation invitational workshop, Rockefeller University, New York.

Wulf, W. (1993) The collaboratory opportunity. *Science*, vol. 261, pp. 854–855.

Yeh, Ron, Chunyuan Liao, Scott Klemmer, François Guimbretière, Brian Lee, Boyko Kakaradov, Jeannie Stamberger, and Andreas Paepcke (2006). ButterflyNet: A Mobile Capture and Access

System for Field Biology Research. In *CHI '06: Proceedings of the SIGCHI conference on Human factors in computing systems,* New York, NY, USA, 2006. ACM Press, pp. 571–580.

Young, Jeffrey, R. (2008*). New Ways to Connect Data, Computers, and People*. The Chronical of Higher Education, at:[http://chronicle.com/wiredcampus/article/3146/new-ways-to-connect-data-computers-and-people

Zimmerman, Ann (2007). Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse*. International Journal on Digital Libraries,* vol. 7, nos. 1–2, pp, 5–16.

Zimmerman, Ann (2008). New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology, and Human Values*, vol. 33, no. 5, pp. 631–652.

Zimmerman, Ann, Nathan Bos, Judith S., Olson, and Gary M. Olson (2009). The promise of data in e-Research: Many challenges, multiple solutions, diverse outcomes. In N. Jankowski (ed): *e-Research: Transformation in Scholarly Practice*, Routledge, pp. 222–239.