

# **Applied eScience: A Research Framework for Mode 2 eScience projects**

**Will Venters  
Avgousta Kyriakidou.**

**London School of Economics, Houghton Street, London, WC2A 2AE.**

**Abstract.** This paper considers the collaborative efforts involved in developing a Grid within a problem-focused and multi-disciplinary project involving commercial, academic and public partnerships. Defining such projects as Applied eScience it argues that these projects present distinctive challenges which have been neglected by existing eScience research. Drawing from the Information Systems literature the paper develops a research framework based on socio-cognitive structures and technology. The framework is then used to describe a case study of an Applied eScience project within Healthcare. This project aimed to deploy a European data-Grid of mammograms to facilitate collaboration and communication between clinicians across the EU. The paper provides practical contributions in the form of a set of recommendations for those involved in Applied eScience collaborations and theoretical contributions in the form of an Applied eScience Research Framework.

**Key words:** Collaboration, Applied eScience, eScience, Mode 2, Grid, Cyberinfrastructure, Health, Technological Frames.

**Pegasus Working Paper  
Information Systems and Innovation Group  
Department of Management, London School of Economics  
Houghton Street, WC2A 2AE.  
[www.pegasus.lse.ac.uk](http://www.pegasus.lse.ac.uk)  
© the authors 2010**

**DRAFT - DO NOT QUOTE WITHOUT PERMISSION OF AUTHORS**

**A VERSION OF THIS PAPER IS UNDER REVISIONS FOR AN INTERNATIONAL JOURNAL**

# 1 INTRODUCTION

This paper addresses the issues associated with constructing eScience infrastructure (cyberinfrastructure) as a collaborative exercise. It does this by focusing upon one form of such collaboration – that to develop eScience Grid infrastructure for a practical purpose, and undertaken with the involvement of academics, industry partners and ‘users’.

The aspiration of Grid technology is to provide a transparent, seamless and dynamic delivery of computing and data resources when and where needed, similar to the electricity power Grid (Chetty and Buyya 2002; Smarr 2004a; Carr 2008). While ill-defined, a Grid is usually understood as a large number of globally distributed processors, storage and other devices linked through networks, and presented to the user as a single computer or computing service and without the need to consider individual resources (Berman, Geoffrey et al. 2003). Berman *et al.* (2003) suggests Grids “will provide the electronic foundation for a global society in business, government, research, science and entertainment” while Carr (2005) boldly argues they will “overturn strategic and operating assumptions, alter industrial economics, upset markets and pose daunting challenges for every user and vendor”.

And yet the reality of Grid technology today is far from such prophecies. While pilot projects are underway to develop Grids for a variety of scientific and industrial domains (Foster, Kesselman et al. 2001; Wladawsky-Berger 2004), it is politics rather than technology which inhibits the success of such initiatives (Orzech 2003; Smarr 2004a). We might aspire to create Grids which, like the power Grid, are (relatively) politically inert, but the creation of such infrastructure is a complex negotiation. Technologies embody their developers’ and users’ social, political, physiological, and professional commitments, skills, prejudices, possibilities and constraints. Grids are a collection of meanings that are contested by different groups involved in their construction and use (Bijker 1995). Developing Grid infrastructure within, and as support for, collaborative practice is inevitably bound up in the processes, interactions, institutional practices and career interests of those involved (Haythornthwaite, Lunsford et al. 2006). If Grids are indeed to support, on a global scale, collaborative working practices (Foster, Kesselman et al. 2001) then

there must be a greater appreciation of the social and political context of such infrastructures' development and use.

Most eScience and Grid research is still firmly located within highly academic collaborations such as in the physical, biological and earth sciences (Olson, Zimmerman et al. 2008). This is unsurprising as one of the primary motivations for eScience funding has been to transform scholarly practice into an increasingly collaborative effort (Atkins 2003; Hine 2006) often founded upon the belief that the lack of inter-disciplinary academic collaboration is due to a lack of ICT infrastructure and that once eScience Grids arrive collaboration can begin to emerge (Atkins 2003). Similarly, the products of e-Research are usually scholarly and contribute to knowledge production (Borgman 2007; Jankowski 2009) rather than directly practical or economic despite the aspirations of e-Research leaders (Hey and Trefethen 2002; Atkins 2003).

Research about Grid based eScience has often focused on collaborative research among academics (Jankowski 2009) – particularly in “big science” (Chompalov, Genuth et al. 2002; Venters and Cornford 2006; Shrum, Genuth et al. 2007; Olson, Zimmerman et al. 2008; Meyer 2009) and drawing upon Social Studies of Science studies (Traweek 1988; Knorr-Cetina 1999). Such research among scientists is not representative of all research practices or of collaborative practices elsewhere in society. These scientists are “often surprisingly egalitarian and broadly international” (Bos, Zimmerman et al. 2008); they work independently and resist corporate control; their knowledge is difficult to codify and disseminate; and they work within universities who institutionally inhibit certain collaboration due to intellectual property rights issues (Bos, Zimmerman et al. 2008) also (Knorr-Cetina 1999), but which also actively promote certain collaborations.

This paper considers another form of eScience research project, under-represented by eScience research but increasingly represented within Grid research practice; the industry-academic-public partnership in which collaboration is between various groups who are not wholly interested in academic research outputs but also in creating direct commercial or public benefit for one of the collaborative parties (similar to action research (Avison, Lau et al. 1999) or design-science (van Aken 2005)). If

scientific collaboration is a social organization oriented toward some objective (Shrum, Genuth et al. 2007) then we consider here collaboration in which the objective is economic and practice-based, targeted at the needs of “users” who maybe participant within the collaboration.

This is a significant area for eScience research as industrial interest in Grid technologies is growing (Goyal and Lawande 2006; Plaszczyk and Wellner 2007), yet research in industrial use of Grids is sparse and fails to address their distributed and collaborative nature (Foster and Kesselman 2004b). The research reported here aims to address this gap. Researching such “interventionist eScience” (Beaulieu and Wouters 2009) partnerships is particularly important as Grid projects increasingly demand wider industry and user involvement<sup>1</sup>, often as a condition of funding<sup>2</sup> or research value<sup>3</sup> and as a focus of national eScience programmes (Newhouse, Schopf et al. 2007; Pothen 2007). Further, as Grid research becomes more commercially relevant and less ad hoc and experimental (Hey and Trefethen 2008) it increasingly involves commercial companies and none-academic user communities.

Such multidisciplinary problem-based research, focused on a specific context-of-application, represents a new form of knowledge production (Gibbons, Limoges et al. 1994) known within science policy as “Mode 2”. Mode 2 knowledge aims at relevance to practice and it is less concerned with disciplinary bases (Starkey and Madan 2001). Mode 2 scientific research operates:

“Within a context of application in that problems are not set within a disciplinary framework [or Mode 1]. It is transdisciplinary rather than mono or multi-disciplinary (...) Research carried out in a context of application might be said to characterise a number of disciplines in the applied sciences and engineering (...) The determinants of a potential solution involve the integration of different skills in a framework of action but the consensus may be only temporary depending on how well it confirms to the requirements set by the specific context of application” (Gibbons, Limoges et al. 1994: vii, 4-6)

While not without criticism (Grey 2001; Mitev and Venters 2009), the Mode 2 collaborative research model is increasingly seen in research practice and indeed many large eScience initiatives explicitly include problem-oriented industry collaboration<sup>4</sup> or are specifically focused on industry problems, for

example in healthcare (Jirotko, Procter et al. 2005), broadcasting<sup>5</sup>, financial services<sup>6</sup>, civil-engineering<sup>7</sup>, construction<sup>8</sup>.

The Mode 2 collaborative research model is intended to aid in bridging a perceived “relevance gap” in research (Starkey and Madan 2001) and to enable research to produce direct public benefit. We argue that there is a need to conceptualise the impact of such Mode 2 research projects within eScience, and in particular whether (and how) the technical, social and political factors influencing eScience are different for such Mode 2 projects. Jirotko et al. (2005) began this exercise by highlighting how “trust” was significant within their (arguably) Mode 2 research project, and that such ‘trust’ was different from that envisaged by Mode 1 (academic, discipline based, investigator initiated) eScience. Lawrence’s (2006) study of multi-disciplinary, geographically distributed large eScience projects similarly highlighted challenges relevant to Mode 2 eScience, particularly the paradox between a focus on research and one on development. Lawrence importantly highlights the need for a balance within large e-Research projects to “satisfy the needs of a multidisciplinary project, managing information and engaging all participants” (Lawrence 2006).

This paper builds upon such work by developing a research framework for evaluating Mode 2 eScience projects – projects we title *applied eScience*. The framework, drawing upon theories from the field of information systems, provides researchers and policy makers with a means of evaluating the technical, political and social challenges of eScience (as called for by Jirotko et al. (2006)). The framework emerges primarily from the literature however an illustrative interpretive case study was used inductively both to hone the framework and in order to reflect upon its relevance for practice. The case study is of a Grid development project within Healthcare. HealthscanGrid was a multi-disciplinary, inter-organisational Grid development research project aimed at improving the clinical practice of Mammography. As one of the first attempts to exploit Grids within Healthcare it investigated the feasibility of developing a European database of Mammogram images for use by clinicians, using a data-grid to support collaboration among the clinicians involved in medical diagnosis and involving, alongside academics, an industry partner and a number of hospitals and clinical staff.

The HealthscanGrid project was set up like a distributed research centre using the Grid to enable collaboration among participants at a distance and reflecting a Mode 2 knowledge production model. Bos et al. (2008) highlight many organizational issues faced by such collaborative research including the challenges of gaining and maintaining participation among the diverse contributors, decision making practices and leadership, intellectual property and the difficult career issues of junior participants. Such problems are also identified within this case, however the focus of our framework is specifically upon the challenges of working between industry, academic and clinical partners within Mode 2 eScience research. As such our research question is: *What technical, political and social factors influence Mode 2 eScience research projects?*

Elements of an answer to this question are well understood; few would be surprised for example by clinicians being interested in the benefits of technology for their patients, corporate partners being driven by profit or academics focused on long-term careers and publications. Our contribution is in providing a framework for empirically grounding and theoretically framing such issues in such a way that future research can learn from them.

The papers' contribution is thus both practical (in providing a set of questions which policy makers, funders, and Applied eScience investigators should explore within their projects) and theoretical (in providing a research framework for evaluating Applied eScience projects).

In the next section we review the relevant literature on eScience and Grid research using this literature to construct our research framework. Using this framework we present a case study which is then analysed to identify salient points for future Applied eScience projects.

## **2 LITERATURE REVIEW**

There are two dimensions which should be addressed in developing a research framework for Applied eScience. Following a structure from Otley (1999) we present these first as questions phrased in

a normative tone. These questions should be obviously relevant across all Applied eScience projects.

They are:

1. *How does Applied eScience collaboration operates?*
2. *How is eScience technology itself implicated in Applied eScience collaborations?*

We take each in turn – presenting a theoretical justification, and then constructing the research framework such that it directly addresses these questions.

## 2.1 How does Applied eScience collaboration operates?

Mode 2 research is characterised by intense interaction between researchers and practitioners which, it is assumed, will increase the relevance of the resultant research product (van Aken 2005). There is a strong relationship between such Mode 2 research and “design science” (Hatchuel 2001; Gregor 2006) in which academics collaborate to develop knowledge that the professionals within a discipline may apply to their practical problems (van Aken 2005). A significant focus of collaboration within Applied eScience research is thus between academic and non-academic participants. It should be noted that many (perhaps most) scientific collaborations involved commercial partners (e.g. IT companies or engineering partners developing facilities for a research collaboration (Shrum, Genuth et al. 2007)). What differs within Applied eScience is the collaboration’s focus on problem-solving for a specific context-of-application.

Within the fields of sciences that have driven Grid development (particle physics, fusion physics, astrophysics, life and earth sciences<sup>9</sup>) collaboration is recognised and takes many different forms. These range from participatory “Athenian democracies” of particle physics (Shrum, Genuth et al. 2007), (and which have been described as “communitarian” (Knorr-Cetina 1999)) to Weberian (Weber 1930) bureaucracies involving legal contracts and project management with significant intellectual property agreements (e.g. materials sciences or space sciences) (Shrum, Genuth et al. 2007). Between these extremes lie levels of bureaucracy representative of Applied eScience collaborations and we thus focus on this third form: the semi-bureaucratic collaboration in which bureaucracy is evident, but in which collaboration rather than control is the norm (Shrum, Genuth et al. 2007).

Such semi-bureaucracies are the mainstay of Mode-2 collaborative research (MacLean, MacIntosh et al. 2001; Macbeth 2002; Tranfield 2002) and represent the most likely form for Applied eScience research. Within such semi-bureaucracies leadership is a key issue (Harvey 2002) and the various stakeholders will need to negotiate among their different traditions and constraints (Etzkowitz 2000) and ultimately must either reach consensus or have consensus imposed by a powerful agent (as Lawrence highlights in her eScience research (Lawrence 2006)). Given that “eScience is interdisciplinary by its nature” and that “interdisciplinary research is notoriously difficult” (POST 2007) reaching such consensus within eScience is often difficult. Applied eScience adds to this the further complication that in addition to different disciplines there are different motivations – the so-called “goal dilemma” (Rapoport 1970) of balancing research output with practical benefit. It is also more likely that the collaboration is not simply constructed of institutional “friends and acquaintances”, but involves participants who are unknown to each other or are even enemies (points of difficulty raised by Shrum et al. (2007)) and are spread across the globe (since global collaboration is the norm within Grid sciences)).

New forms of organisation have emerged to manage such scientific collaborations (Chompalov, Genuth et al. 2002) of which “Virtual Organisations” are one example (and are modelled within the Grid middleware (Foster, Kesselman et al. 2001)). In management science, inter-organizational alliances have been investigated for a long time (Jha and Watson-Manheim 2007) and been termed “Virtual Organizations” since at least 1992 (Davidow and Malone 1992; Jha and Watson-Manheim 2007). Such Virtual Organisations, it is claimed, facilitate a better alignment between research interests and practitioner needs so “enabling complex practical problems to be solved and ferments the creation of networks of experts” (Starkey and Madan 2001). Yet difficulties with inter-disciplinary research within Virtual Organisations are often discounted (Dodgson 1994) and are under researched (Newell and Galliers 2000) but include “mechanistic pooling” (Newell and Galliers 2000) whereby “each member of the pool [is] taking a different ‘slice’ of the project and the work then proceeds with the minimum of communication between members” (Knights and Willmott 1997). Such problems relate to participants lacking an understanding of the wider team’s work practices and epistemic cultures (Knorr-Cetina 1999),

the difficulties of communicating among distributed teams, and the lack of career recognition of such work (Haythornthwaite, Lunsford et al. 2006). Further there are difficulties of instrumentalising the context-of-application for the benefits of each member's key constituencies (Mitev and Venters 2009) rather than for collective aims.

Such collaborative and virtual research endeavours blur the boundaries between researchers, sectors and disciplines (Chompalov, Genuth et al. 2002) and have consequences for the development practices involved in the construction of the Grid infrastructure. Collaborative, multi-organisational relationships are negotiated in “an ongoing communicative process”, emerge from informal, ongoing relationships (Hardy, Phillips et al. 2003) and challenge the traditional emphasis on formal agreements with identified goals, rational partner selection, and performance monitoring (Powell, Koput et al. 1996) aligned with traditional plan-based approaches to systems development practice (Avison and Fitzgerald 2003; Mitev and Venters 2007).

Given these challenges a primary requirement of effective Mode 2 research is to “intensify the researcher-practitioner interaction during this process so that the researcher gets a better understanding of field problems, their possible solutions, the needs of practitioners and the intricacies of effective communication with practitioners” (van Aken 2005). **We therefore suggest the need to focus, within our Applied eScience framework, on the collaborative interaction between the various stakeholder groups and on the nature of the bureaucracy this collaboration enacts.**

## 2.2 2. How is the eScience technology itself implicated in Applied eScience collaborations??

Collaboration within science is inherently “technoscientific” – It is “crucial to ask how the design and construction of [technology] interact with the organisation of the participants” (Shrum, Genuth et al. 2007: 197). eScience has similarly been defined as “the *intersection* of Grid and collaborative research” (David 2004: original emphasis) and Grid technology (in various forms) is thus invariably seen as central to Applied eScience projects. For this reason it is necessary to consider the Grid technology as implicated within the collaboration, and hence an important part of the framework.

A Grid infrastructure is an emerging platform for coordinated resource sharing and problem solving undertaken on a global scale and of particular relevance for data-intensive and compute-intensive applications (Foster, Kesselman et al. 2001). It is a practical solution to the problems of storing, distributing and processing the large amounts of data that are or will be produced by industry and scientific communities over the next decade (Hey and Trefethen 2002). Grids are more and more central to the undertaking of advanced science such as particle physics (Venters and Cornford 2006), geophysics, earth science, social science (Scott and Venters 2007) and increasingly in areas such as financial services, engineering and health (Wouters and Schroder 2003). They are centred around a set of standards (protocols) for the control of distributed resources which are realized as the Grid middleware software (Foster and Kesselman 2004a) which provides technological governance. Just as Internet protocols enable the sharing and integration of information on the Web, so Grid protocols allow the integration not just of information, but of sensors, applications, data-storage, computer processors and most other IT resources (Wladawsky-Berger 2004). It is the development of this application stack (from the middleware which controls Grid resources, through to the applications which achieve users' goals) which is the focus of much Applied eScience development – and is the focus of this paper.

The construction of a successful Grid is a socio-technical achievement involving the construction of an information infrastructure (Hanseth and Monteiro 1998) rather than simply a technical implementation. Such an infrastructure includes technical elements, but also the practices and routines within which such technology is situated (Ciborra 2000; Cordella 2006). A Grid application is thus a socio-technical network (Hanseth and Monteiro 1998) established through complex socio-technical processes and shaped by events, circumstances, and unpredictable courses of action during development and use (Broadbent and Weill 1999; Cordella 2006). **We therefore suggest the need for a research framework which considers how the technology is socially constructed by those involved in its construction.**

### 3 APPLIED ESCIENCE RESEARCH FRAMEWORK

Having identified above key challenges for Applied eScience projects, this section develops a research framework to enable their exploration and potential improvement. The aim of the framework is to aid the analysis of the interpretive processes of those involved in Applied eScience projects, and through this explore their impact upon the resultant project and Grid technology. The literature review suggests that such a framework provides an interpretive investigation of the Grid implementation, and an institutional analysis of cultural, structural and political issues.

Applied eScience, as this paper defines it, concerns the collaborative construction of a Grid-based technology which achieves a practical benefit for a user-constituency (usually participant within the project). Within the field of Information Systems considerable research has been undertaken on the collaborative development of similar socio-technical systems within similar contexts of application, and drawing upon interpretive practices (Galliers 1997; Avgerou and Cornford 1998; Checkland and Holwell 1998; Avison and Fitzgerald 2003; Gregor 2006). Applied eScience Grids are embedded in social systems and their development and use is shaped by social factors and by the meanings attributed to it by different groups (Kline and Pinch 1996; David 2004). Identifying these groups is thus a key requirement for the research framework.

Rather than focusing on social structures as defining such relevant groups, and so prefacing power and hierarchy, we instead focus on the cognitive processes of Applied eScience participants as the foundation of our research framework. We reflect Weick's (1990: 17) view that "cognitive and micro-level processes are key to understanding the organisational impact of new technologies" and therefore seek to define the interpretations of participants in Applied eScience as the key determinant of the social grouping.

This suggests that within Applied eScience projects social groups can be identified by their shared common beliefs, knowledge and interpretations about the project. These relevant social groups, derived from literature on sociological studies of technological innovation (Bijker, Hughes et al. 1987) are defined

as individuals whose interactions and experiences with, and expectations of, a technology are similar (Davidson 2006: 25).

Orlikowski and Gash (1994) provide the foundational framework for our Applied eScience framework based on this idea of relevant social groups. Their framework focuses on the influence of shared cognitive frames to “trace the often unacknowledged structural influences of shared interpretations” within such relevant social groups during information systems development and implementation (Orlikowski and Gash 1994: 200) and is thus relevant to Applied eScience. Their concept of Technological Frames shares ontological assumptions with Bijker’s use of the term (Bijker 1995). However, as Davidson (2006) highlights, unlike Bijker (where the focus is upon technological innovation over time, with frames as social phenomenon), Orlikowski and Gash (1994) employ Technological Frames as a socio-cognitive phenomena. For Bijker the technological artefact is ‘within’ the frame, yet, for Orlikowski and Gash it is a distinct phenomenon (Davidson 2006: 37). This allows a more direct and immediate focus on the interpretation of technology by the groups, rather than considering the technology within such interpretation as it emerges and evolves over time (e.g. as presented in (Bijker 1995)). This similarly allows for relevant social groups to be considered which are smaller than is usual within social construction of technology studies. Within this study we only employ Technological Frames and Relevant Social Groups as Orlikowski and Gash (1994) define them – our research is thus not a social construction of technology study.

These relevant social groups have a shared frame of reference (Technological Frames) towards the eScience technology being constructed – that is the set of meanings, interpretations, assumptions, expectations and knowledge members use to understand technology. This includes the technologies’ capabilities and functionalities, how it will be used, the consequences of use, the motivation for use and its likely value to the organization (Davidson 2006). Their knowledge about the project and the Grid technology is socially situated within their communities (leading to a close alignment between these groups and communities of practice (Lave and Wenger 1991; Wenger 1998)).

In order to identify issues within an Applied eScience project there is a need to focus on inconsistencies (incongruence) in the Technological Frames between the relevant social groups within a collaboration. Such incongruence can create problems in developing, implementing and using the technology, as well as a breakdown in communication, lack of participation by its users, and failure in engagement by groups (Orlikowski and Gash 1994).

Identifying such relevant social groups, then considering their Technological Frames, aims to address the two challenges presented in the literature review – firstly to consider how the eScience technology is socially constructed by these social groups (since “Technological Frames are built into the artefact that technologists design and construct” (Orlikowski 1992; Orlikowski and Gash 1994)), then secondly to consider how the relevant social groups interact and collaborate in order to construct the technology *as they understand it*. Finally, to consider the nature of bureaucracy and power within this social structure as defined (and understood) by the relevant social groups.

The aspiration for a successful technology is that incongruence between interpretations and conflicting images of the technology may cease as consensus (or domination by one group or by bureaucracy) occurs resulting in so called closure (Klein and Kleinman 2002) leading to the stabilization of the artefact for an extended period (Pinch 1995; Kline and Pinch 1999). Such stabilization is not static or final as further changes may occur over time.

### 3.1 Applied eScience Research Framework:

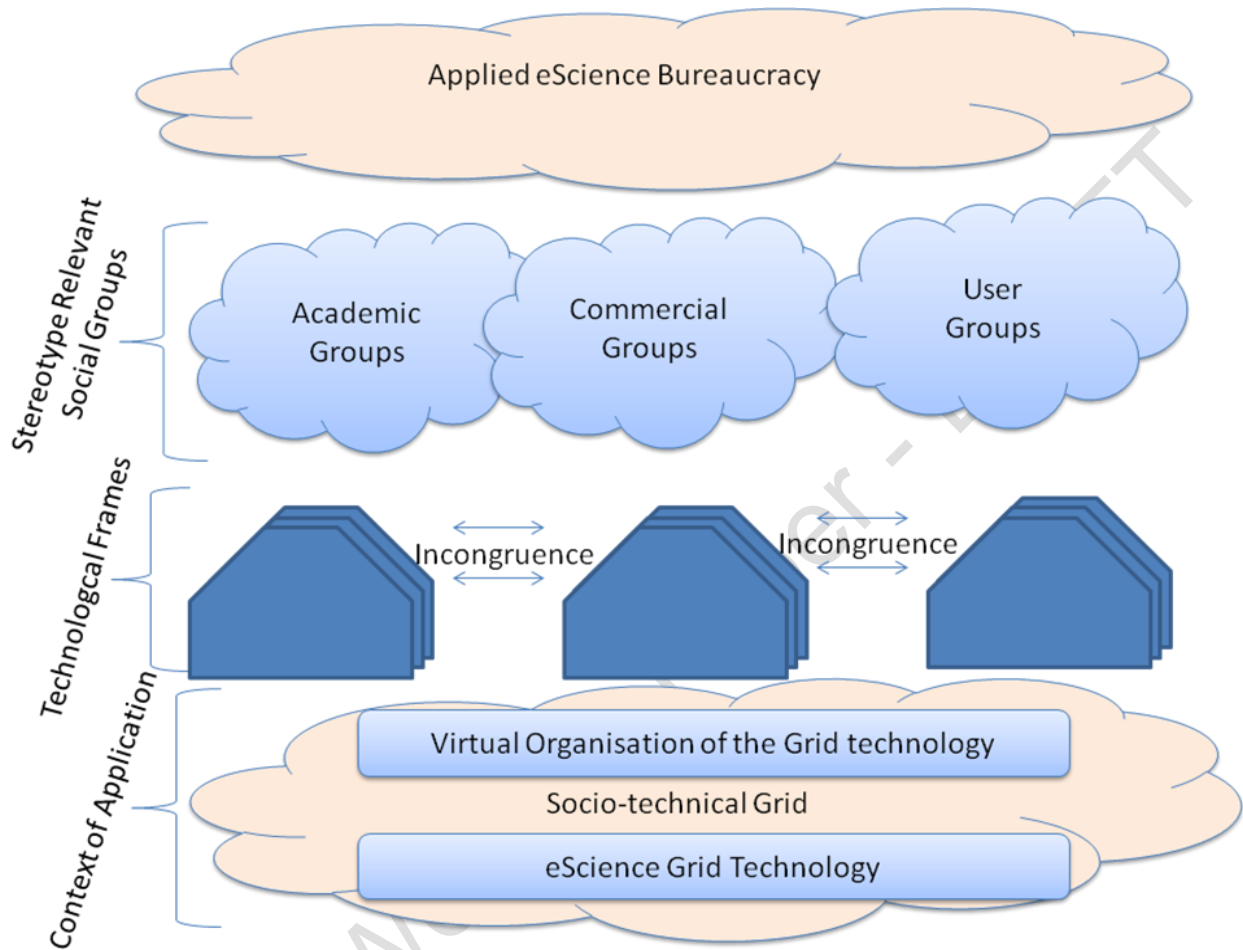
Figure 1 outlines the resulting eScience Research Framework. Orlikowski and Gash (1994) highlight the considerable difference between the Technological Frames of two primary groups within Information Systems in general – users and technologists. Mode 2 research suggests three primary types of relevant-groups which may have significantly differing Technological Frames within Applied eScience – Academics, Commercial-partners and Users (or more likely user-representatives). The Applied eScience research framework thus suggests these three stereotype social groups as the starting point for analysis. It is noted that through analysis of the Technological Frames held by members of the project it is likely that

other relevant social groups will be identified, and that these stereotype social groups may also be subdivided. Considerable effort should also be put into identifying silenced voices, excluded parties and those affected by the eScience technology who may not be represented in the project bureaucracy. Finally these groups may be ill-defined and overlap – hence they are shown as a cloud in the diagram.

The Technological Frames of an Applied eScience project will relate to the Grid technology being interpreted by these identified relevant social groups. One of the central ideas within Grid computing however is the “Virtual Organisation” established in order to enable such sharing of geographically distributed resources. A Grid’s “Virtual Organisations” define what is shared, who is allowed to share, and the conditions under which such sharing occurs (Foster, Kesselman et al. 2001) and are encoded as rules within a Grid’s technical infrastructure. Unlike standardised infrastructure, Grids can present different capabilities, services and resources to users as defined by their membership to a Virtual Organisation (Foster and Kesselman 2004b). Virtual Organisations thus “enable disparate groups of organisations and/or individuals to share resources in a controlled fashion, so that members may collaborate to achieve a shared goal” (Foster, Kesselman et al. 2001). Such Virtual Organisations define who uses the Grid and how. The Technological Frames held by Applied eScience project participants will thus be an interpretation of the Virtual Organisation of the Grid presented to that group, rather than just of the Grid itself. This suggests the need to consider within the framework a context of application which includes the Virtual Organisation of the Grid as well as the Grid itself and its related socio-technical infrastructures (e.g. support structures, maintenance, training).

Finally, the literature review identifies a bureaucracy (social structures and practices which exhibit power) that influences the behaviour of the relevant social groups as they collaborate to achieve the projects’ objectives. This bureaucracy may emerge from within the relevant social groups (as leadership is established) or from outside (for example from funders, national eScience programmes, project-steering groups, project plans, reviewers). Further, such bureaucracy represents the formal collaborative arrangements which inhibit or allow collaboration among the relevant social groups. As a representation

of power with Applied eScience projects such a bureaucracy may limit the impact of the differing interpretive frames, and may influence the form of collaboration or motivation of participants.



**Figure 1: Applied eScience Research Framework**

#### **4 CASE STUDY EXPLORED USING THE APPLIED ESCIENCE RESEARCH FRAMEWORK .**

In this section we use of the framework in analysing an Applied eScience project within Healthcare. Evidence was collected by extensive reviewing of HealthscanGrid’s documentation and by undertaking semi-structured interviews with key people from each partner involved. In addition the findings were discussed in detail with the chair of the HealthscanGrid project. The people interviewed were key

representatives of the institutions involved in the project. The following table provides a summary of the interviews conducted.

<b>Position in the HealthscanGrid</b>	<b>Institution</b>
Project Leader	PAL
Middleware developer	PAL
Research Dissemination Coordinator	BigUniversity
Chief Scientist	MedicalXYZ
Promote Standardization technology (SMF)	MedicalXYZ
Research Associate (Working closely with clinicians)	HospitalA
Research Associate (PhD Student)	HospitalA

**Table 1: Summary of interviews**

The relevance of this case for studying Applied eScience is the explicit focus on practical as well as academic results, and the Mode 2 nature of the project. Indeed one of those interviewed described HealthscanGrid as “... one of the first cases of development of a Grid system which was aiming to be used by society”.

#### 4.1.1 *HealthscanGrid*

The medical community has for a long time explored collaborative approaches for managing, storing and analyzing image data and for exchanging knowledge. The Grid was seen to be a promising technology as it enables new collaborative approaches for image analysis without the necessity for clinicians to co-locate (McClatchey, Manset et al. 2003). The HealthscanGrid project was one of the first attempts to exploit Grid infrastructure within Healthcare and aimed to investigate the feasibility of developing a European database of mammogram images, using a data grid to support collaboration among clinicians across the EU in medical diagnosis. A Grid infrastructure was required to make available the large amounts of data as ‘digital libraries’ (Baker, Buyya et al. 2002; Rajasekar, Wan et al. 2002) to Healthcare institutions around Europe, as well as allowing for its efficient management and analysis. The project was funded by the European Union and ran for three years ending successfully with

the delivery of a prototype system to a Spanish private company who are currently trying to develop it into a large-scale product of commercial value.

The project was a collaboration of a number of institutions across three countries (which we anonymise as Narnia, Neverland and Atlantis), with five heavily involved: HospitalA (based in Narnia), HospitalB (Atlantis), BigUniversity (Narnia), MedicalXYZ (Narnia) a small spin-off SME company which was closely collaborating with OldUniversity (Narnia) and finally PAL (Neverland) a particle-physics accelerator laboratory. When these institutions came together for the first time, they realized that there was another group, who used to closely work with the project's leader from PAL on various imaging processing projects, that formed a distant group based in Atlantis: "we haven't been aware of them from the beginning... they were interested in automatic annotation... a quite a distant group".

The Virtual Organisation of HealthscanGrid's users (HSGVO) consisted of three mammography centres in two countries: the HospitalA (Narnia), HospitalB (Atlantis) and the OldUniversity (Narnia). The HSGVO central node at PAL was coordinating the access with PAL's role being to provide the Grid expertise, develop middleware and also leading the project as project-coordinator. MedicalXYZ's role was to adapt their already existing mammogram software to meet the HealthscanGrid's requirements. The BigUniversity was responsible for gathering comprehensive requirements from users and for the development of the database, thus bringing database and software engineering skills. Finally, the hospitals involved represented the end-users of the project, providing the information and medical data to go into the HealthscanGrid, as well as the requirements and feedback to the developers. The deliverable of the project was a software prototype, capable of allowing complex epidemiological studies, statistical and computer aided detection analyses and the deployment of versions of the image standardization software. The prototype was considered a success by its collaborators, despite some small failures particularly its lack of user-friendliness, and the fact that some features had not been developed: "The prototype is a success. It was very ambitious...Of course they were technical problems...But these could be seen as the normal glitches and the small failures that happen when you develop a prototype system".

That there were five different institutions, involving people from different disciplines and with conflicting priorities and expectations, made the collaboration difficult; something that held back the project: “There was the commercial company, talking to research people and each group was not appreciating the other ones work. They all had different goals and priorities. Commercial people were concerned with profit and researchers wanted their job being done”. Although they were involved for different reasons in the project, only one thing linked them together: “Each institution applies for funding... People are meeting together for a common purpose: the funding...It is just business...” Even though the project can be seen as a success in the end, most of the collaborators felt disappointed in terms of poor teamwork and leadership as well as that they had given more input and effort in comparison with what they had received in return. As one respondent said in an interview: “The HealthscanGrid could have great results, it could have been better if people behaved and talked to each other. Two of the teams involved could not work together, could not collaborate.”

## 5 ANALYSIS

Employing the Applied eScience Framework above (itself based on Orlikowski and Gash (1994)) we firstly identify three key relevant social groups within the case study; Academic-developers, Commercial-developers and Healthcare-users. We describe each in turn before considering how these groups’ Technological Frames towards HealthscanGrid (as a context-of-application that is a socio-technical infrastructure consisting of the Grid and Virtual Organisations) were influenced by their actions. In particular we consider how these relevant social groups’ incongruence influenced HealthscanGrid. Within this we discuss the nature of the bureaucracy surrounding the HealthscanGrid project and its role in shaping the project.

### 5.1 Relevant Social Groups.

**Academic-developers’ group** comprised people from three institutions; the particle physics group from PAL, the distant Atlantis Group of image processing experts (automatic-annotators) and the team of

academics and PhD students from BigUniversity. The team from BigUniversity were based at PAL during the deployment of the project which led them to share similar expectations, understandings and assumptions concerning the HealthscanGrid with the particle physicists at PAL. Particle physicists have always employed cutting-edge computing in their research (Jones 2004), are very pragmatic in the way they develop technology (Kyriakidou and Venters 2009) and have a reputation for getting things done (Venters and Cornford 2006) and this attitude (perhaps arrogance (Hermanowicz 2009)) influenced the group. As one interviewee stated: “PAL succeeds; there could be delays and small problems in developing technologies, but they always manage to solve them”. They pursued the opportunity to be involved in the project because they wanted to provide the knowledge they had acquired from fundamental physics experiments to something that could benefit society at large, and also to respond to criticism that physicists kept their knowledge to themselves and only cared about solving their own problems. The academics had backgrounds in computer science and information systems with an interest focused on research. For them the project was an opportunity to engage in the development of a Grid and therefore get the Grid expertise; to get involved in a European project, and to use the knowledge acquired from previous projects based at PAL.

**Commercial-developers’ group** comprised of staff at MedicalXYZ and associated academics from OldUniversity. MedicalXYZ was a small start-up company interested in business, growing as a company and making money. They were focused on two projects: image fusion and mammographic image processing. At the beginning of HealthscanGrid, the computing competency and technical skills of the Commercial-developers group was low, so the project was an opportunity for them to: (1) get access to the Grid technology and gain the technical expertise and core competencies for more advanced products using the Grid and (2) expand their panel of collaboration partners, so that data and clinical expertise would be available: “The main expectation for us was to demonstrate that we can access remote data through the Grid and perform state-of-the-art visualization for it. Commercially, it was important for us to be visible as technical leaders”. While the project was still running, the company was taken over by a

larger company and consequently their focus changed and mammography stopped being central to their activities. From then on the resources they provided were restricted.

**Healthcare-Users' group:** these clinicians were the end-users of the project; there was little computing competency or technical skills among them and their interests had nothing to do with commercial work. Their goals as a community were three-fold: to improve the techniques for breast cancer diagnosis; to push forward the frontiers for using software in Healthcare; and to promote teamwork among hospitals for clinical research. They were quite optimistic about the opportunity offered “The system will be able to project cancer risk, improve the diagnosis; it will be able to do so many things... ”.

In addition to these dominant social groups above, most participants in HealthscanGrid agreed that a strong bureaucratic influence on the project was presented by the EU as the funding body. “There was clearly a leader and that was the European community... if we didn't deliver, we wouldn't get the money and the project would stop”. This reflects the strong influence funding bodies have on multi-disciplinary research (Newell, Swan et al. 2001) and in particular the need for ‘solutions’ to reflect the pre-specified ‘deliverables’ set out in project funding proposals (Newell, Swan et al. 2001). While we cannot identify this funder as a social group (since its interpretation of the project is not known), it still remains relevant. In addition, we note the automatic-annotation group in Atlantis as a silenced but important sub-group of the Academic-developers. We also see PAL as in some ways a sub-group of the Academic-developers – sharing their broad interpretation but with specific differences. Finally the women whose mammograms were to be put on the Grid exist as a silent but relevant group.

## 5.2 Analysis of HealthscanGrid: incongruence and bureaucracy.

Having identified relevant social groups we now consider how they shaped the project and more particularly the resulting HealthscanGrid technology. We consider the technology's development, its capabilities, its use, and the expectations of the groups concerning HealthscanGrid's future. We highlight

incongruence within their interpretation of the project and the technology, demonstrating how such incongruence influenced the final project outcome.

#### *5.2.1 Divergent motivation for Participation.*

The three relevant social groups were motivated to become involved in the project for different reasons. The expectations of the high impact the Grid could have in medicine was the primary reason the project was initiated by the particle physicists within the Academic-developers' group. They felt that the Grid at PAL, and the advanced techniques they had developed, could be adapted in the medical field and lead to a very important output. They expected that HealthscanGrid could not only help improve breast cancer diagnosis but, if successful, be adapted and used for other applications in medicine. Contrary to this, the Commercial-developers groups' expectations concerned the impact HealthscanGrid could have on their company's strategy, structure, status and way of doing business. HealthscanGrid could provide them with the expertise they badly needed and commercially it would make them visible as technical leaders. The Healthcare-Users group, in contrast, wanted a system that could provide a solution to the problems surrounding breast cancer today. No technical expertise existed among them, so they were introduced to the Grid and what it could do in medicine by the Academic-developers' group. Having understood its importance and the impact it could potentially have in saving lives, they agreed to initiate the project and collaborate with the other institutions.

#### *5.2.2 Divergent Ideas of the nature of the Grid.*

The functionalities and capabilities HealthscanGrid would provide; how it would be used on a daily basis, and the consequences of its usage were not clear to any of the groups, and the groups had different interpretations of its nature. The Academic-developers' group knew from previous experience that HealthscanGrid could provide capabilities that at the time did not exist in Healthcare in particular facilitating communication, sharing and storing of data and information across hospitals linked to the system in a transparent way and without a loss in quality. They expected that the HealthscanGrid would be used for epidemiological studies, as a "second opinion" doctor, as well as for the comparison of similar breast cancer cases. However, their interpretations concerning HealthscanGrid led them to emphasize the

advanced technical capabilities and leave other aspects aside such as “making it user-friendly and automated” and ultimately such features were not implemented. Those with an academic career seemed to be more focused on exploring advanced technology than on the practicalities. This is unsurprising given their expectation that a successful prototype would provide them with a large number of PhDs and publications upon which their personal success is judged. To them it was just a prototype and hence basic capabilities such as being user-friendly and automated were not as important as exploring advanced capabilities of research interest. For this reason this group were comfortable to somewhat ignore the necessity for security and confidentiality (though these were important issues for the Healthcare-Users’ group).

### 5.2.3 *Challenges developing the collaborative Grid system.*

The Academic-developers introduced into HealthscanGrid Grid software called AliEn<sup>10</sup> that was specifically developed for the Large Hadron Collider ALICE experiment at CERN. This was argued to be a “quick and dirty solution...where 95% of the code was taken from [other OpenSource projects] and 5% was what the team developed”. This allowed the project to proceed quickly as desired by all collaborators (as one stated: “there was no time to hang around and wait for the new version of Globus or waiting for EGEE to deliver”<sup>11</sup>). However its messy “academic” nature, which reflected the systems development culture at PAL, proved a significant problem later in the project as the “HealthscanGrid code and AliEn code became extremely mixed up” and this messy code “looked like that might have killed [the HealthscanGrid]’s work” since it was unacceptable to the Healthcare-Users’ group and did not reflect the formal software engineering approaches demanded by the commercial-developers.

There were also problems with communication between the different software components, and between the developers. HospitalB could not connect to its own software within the HSGVO and its doctors had to travel to HospitalA in order to see their own images: “The technical ‘hiccups’ were holding back the project. MedicalXYZ’s software wouldn’t talk to PAL’s software and the only thing they did was just blaming each other.” In stark contrast to the Academic-developers, the Commercial-Developers (in particular MedicalXYZ) saw Intellectual Property Rights issues in all of the software associated with

HealthscanGrid. Adding to the problems of these technical ‘hichups’ MedicalXYZ was unprepared to share their source-code with the project (unsurprising given its commercial demands but in conflict with the Academic-developers’ interpretation of the nature of code), making it hard for the system to develop and support the clinicians directly in their work. “We didn’t have the [MedicalXYZ] source code [because] there was a commercial contradiction. That was a significant delay”. Indeed some systems needed to be re-developed by the Academic-developers because of this.

A similar example of commercial interpretation of the nature of code hindering the project occurred when the Academic-developers attempted to adapt a piece of software incorporating MedicalXYZ’s software components. MedicalXYZ became agitated as they did not wish their software to be used within the academic software. Their reason for this were believed to be because the academics would publish papers about this new software, and that such papers would either, argue MedicalXYZ’s product was problematic, or that their product was more successful than the commercial version. Either way, MedicalXYZ felt they gained nothing but might lose much from such academic publications. Indeed throughout the project’s life-cycle their focus was “making sure that they were in there making a profit”. They were described by an academic as “corporate players”.

### 5.3 Systems Development Practices

The project failed to develop formal software-procedures which made development frustrating. Commercial-developers were not automatically informed of changes to the Grid. One recounted: “We knew about updates of the libraries only by finding out that one day, the workstation stopped working. After spending 2 weeks on it, we realized that the libraries had changed and nobody told us. That was 2 weeks totally wasted.”

The gathering of user requirements also proved difficult as different partners did not appreciate the demands of other partners. Academic-developers and Commercial-developers were trying to make Healthcare-Users understand that once features are defined they need to be developed, tested and reviewed before the next round of features were raised. On numerous occasions sets of features were gathered and agreed but the list kept on increasing during development. Clinicians became frustrated,

since they felt that the commercial-developers, were not doing their job, and were not delivering the software for them. MedicalXYZ however argued that clinicians did not realize that the list of features they were asking necessitated excessive manpower in comparison to what was budgeted for in the project. In the end MedicalXYZ ceased to participate in further development: “at some point they said we spent all the money that we have been allocated in this project so there is nothing else [MedicalXYZ] can do”. At this point they ceased to contribute further.

### *5.3.1 Commercialisation of the Applied eScience output.*

For the Academic-developers the projects’ purpose was to demonstrate the potential for HealthscanGrid to become a commercial product at some point in the future rather than as an outcome of the project: as a member of this group stated “We hope to see the son of HealthscanGrid based on a commercial grid used by doctors in an efficient way. We know that it is quite difficult to be adopted, especially at the beginning. But just look at medicine today compared with 20 years ago. It can happen”. On this basis the group made assumptions regarding the use of the technology by clinicians; assuming it did not require a formal implementation and training plan, and considering that security, patient consent and confidentiality could be left for consideration in the long term.

The Commercial-developers group, by contrast, learned about the Grid while searching for a technology to support the two commercial projects they were working on at the time. They felt that the Grid could bring a revolution in communication and sharing of data and give them the opportunity to develop better algorithms for their projects. Their expectations and assumptions concerning how the HealthscanGrid would be used on a daily basis were not so clear. They knew that the project could help them commercially and, in the opinion of those interviewed, that was the main thing they cared about. Regarding the consequences of the actual usage of the HealthscanGrid, they argued that one possible problem would be if hospitals did not wish to share their data: “All institutions work for themselves. You need to give them value in return for what they are giving you”. From the Commercial-developer groups’ perspective their expectations concerning the final output were more aligned to acquiring technical expertise throughout the development, rather than expecting benefit from the actual prototype itself. In

particular they felt it necessary to be visible as technical leaders in the market field, a goal they achieved evidenced by their acquisition by a larger company, and later by a well known international engineering and electronics company.

The Healthcare-Users' group felt they understood that technology would support collaboration between doctors from hospitals connected to the HealthscanGrid, epidemiology studies, teaching, and to be used as a second opinion doctor. They expected the system to improve diagnosis of cancer risks, and that it would directly facilitate real-time usage so that clinicians would be able to collaborate during surgery. They had little experience with the technology and expected a formal implementation and training plan. Their main concern regarding its use related to security and confidentiality. Furthermore, they expected the system to be user-friendly and automated. When the prototype was delivered to the clinicians and put into use they found it extremely hard to cope with, since it was not user-friendly and did not meet their requirements. In addition the final prototype proved to be too slow and inefficient to be used for real time workflows.

During the project MedicalXYZ was acquired, firstly by a specialist company and then by a larger multinational creating many further problems and tensions, since the priorities of the company necessarily shifted dramatically away from the HealthscanGrid project and MedicalXYZ stopped providing any input to the project: "...with a little more effort from MedicalXYZ in the last phase of the project, we could have had a better prototype, with some functionalities better developed... If there is something it doesn't work as it should, it's because at a certain time MedicalXYZ deployed a product which was not the full thing they could build..."

### 5.3.2 *Bureaucratic power of the funder.*

At the end of the project the necessity to comply with the EU funding requirements, rather than common consensus, ensured that the relevant social groups completed their stated objectives on time. This ensured that a form of closure and stabilization was achieved and documented in the final project submission. The groups knew that if they did not deliver what they had promised and what the EU was paying for, they would not get the money and they would have to explain why this happened. As one

interviewee stated: "...What needs to be done, has to be done to deliver; if it wasn't being done, then we would all work as a team to get it done...If one of us failed, all five of us failed. We had to make sure that all five of us would succeed, otherwise we would all be giving back our money." Therefore all groups reached a form of closure because in a way they had to. They put aside their expectations and understandings concerning the HealthscanGrid, and they all worked together in order to manage to deliver.

## 6 DISCUSSION

The HealthscanGrid was one of the first attempts to apply Grid technology within Healthcare. The project was a success insofar as a prototype was delivered and has the potential to become a useful commercial product which continues to play a significant role in the future of using Grids for collaborative clinical diagnosis and treatment. Yet a range of issues, both technical and bureaucratic emerge from our analysis using our Applied eScience framework.

This research highlighted the difficulties of collaborative Applied eScience. The project did not achieve a trans-disciplinarily or truly collaborative form of practice; instead it remained wedded to incongruent interpretive perspectives and practices. Unsurprisingly then the project proceeded through relevant social groups "mechanistically pooling" (Knights and Willmott 1997; Newell, Swan et al. 2001) their knowledge and skills and concentrating on their own agendas rather than the specific collective need – and in particular the needs of the context-of-application (the clinical specialists or the women whose mammograms were to be analysed); "Collaboration existed within the groups but not among them". This reflects other studies of scientific collaboration which highlight the opportunity cost of collaboration (Shrum, Genuth et al. 2007: 206) and that participants thus demand personal-benefit or fail to engage.

Grids may aim to be transparent, seamless and dynamic (Chetty and Buyya 2002; Smarr 2004b), however various technical problems inhibited this within the HealthscanGrid project. The relevant social groups' interpretation of HealthscanGrid was incorporated within its socio-technical fabric and remained separate. As one person stated: "it's remarkable how this virtual collaboration [HSGVO] worked. They

actually did enough for each other, the different teams, to make things possible. But it was also possible because the tasks they undertook were sufficiently separate”. There was a need for stronger control over the software development practices within such Applied eScience projects to enable a more coherent system to be developed. In particular problems occurred when cohesion was needed between socio-technical Grid modules developed by different relevant social groups. Such problems ranged from Intellectual Property Rights issues, software-interface issues and lack of source-code sharing, to lack of expertise and skill. One interviewee stated this “I suppose one lesson learned here is to agree on how to minimize the overlap in responsibility and how the two [modules] are going to communicate, both in the sense of how the two teams will communicate and how the [Grid] artefacts are going to work together: how the software will interface with each other”. It was unsurprising then that the Virtual Organisation and Grid (supposedly reflecting the workflow for the clinical users) failed to reflect their demands for security, control and training; instead they were constructed from the incongruent interpretations of what was required from the different collaborators.

Like the Jirotko et al. (2005) study, trust was an issue within the HealthscanGrid. This manifested itself most visibly in the Commercial-developers unwillingness to share source-code. However, trust is reciprocal and one can imagine that in this case the asymmetric nature of the source-code ownership issue was a significant barrier. While MedicalXYZ naturally cared passionately about their intellectual-property, academic collaborators only desired to publish and share the intellectual spoils of the project. Their different interpretations of “value” and “trust” therefore inhibited collaboration.

Ultimately it was bureaucratic control which led the project to achieve its success. Power within the project was distributed and, like many inter-disciplinary and multi-institutional research projects, HealthscanGrid appeared leaderless (Chompalov, Genuth et al. 2002). This lack of leadership and direction led to problems with coordination – indeed the nominal project leader was the funder’s representative as this was the source of power. Newell et al. have highlighted that the “perceived requirement for industrial relevance is changing the ways that funding bodies are attempting to manage research with growing tendencies towards managerialism” (Newell, Swan et al. 2001: 99). Increasing

managerialism can be seen within inter-disciplinary projects (Powell, Koput et al. 1996), particularly EU projects<sup>12</sup>. It was however power (manifested from the funding) rather than managerialism (in the form of rigorous demands for documentation and progress reports) which provided most control over the collaboration. When participants failed to communicate “[the funding representative] was shouting at them ‘that’s it, pack up the project and give me back the money’... That’s when they got scared and started being nice to each other”.

We summarise these issues in Table 2, and present lessons for future eScience projects from this case study. These lessons are derived from the case but are not canonical though are written in a normative form.

<b>Framework Concept</b>	<b>HealthscanGrid Examples</b>	<b>Potential lessons for future Applied eScience</b>
Bureaucracy	<ul style="list-style-type: none"> <li>• Commercial company vs. academics.</li> <li>• Profit and visibility as technical leaders in Grid development vs. pragmatism with academic excellence through publications.</li> <li>• Funding as a driver to create an “image of success” but not to force collaboration.</li> <li>• Limited managerial control. No clear leader of the development process. Escalating requirements due to lack of control (Chompalov, Genuth et al. 2002).</li> <li>• PAL was at sometimes seen as leading and providing coordination to the project because it was handling the EU money.</li> </ul>	<ul style="list-style-type: none"> <li>• Discuss group’s collective responsibilities throughout the project.</li> <li>• Confront conflicts among the group head-on.</li> <li>• Explore interpretation of the project by each group. Interpretive problem structuring techniques may help (Rosenhead 1989; Checkland 1995).</li> <li>• Identify how funding will shape outcome.</li> <li>• Identify sources of leadership and control.</li> </ul>
Relevant Social Group’s collaboration	<ul style="list-style-type: none"> <li>• Breakdowns in collaboration and mechanistic pooling of actions (Knights and Willmott 1997).</li> <li>• Chaotic systems development practices. Poor effort in making agreements to ensure success.</li> <li>• Source-code ownership and the pooling of source-code.</li> <li>• Data ownership issues (within a clinical setting) and publication strategies for eScience outputs.</li> <li>• Limited social interaction and</li> </ul>	<ul style="list-style-type: none"> <li>• Appreciate participants may mechanistically pool their knowledge and skills.</li> <li>• Consider how to engender distributed Grid development projects with both commercial and non-commercial development traditions.</li> <li>• Discuss Intellectual Property Rights, and publication strategies at the outset and throughout.</li> <li>• Ensure group interaction which is focused on the outputs of the project (as</li> </ul>

	collaboration led to the different interpretations of the final prototype.	tangible eScience systems) rather than on listing requirements. <ul style="list-style-type: none"> <li>• Discuss the context-of-application and silenced beneficiaries or victims of the eScience technology.</li> </ul>
Socio-technical Grid	<ul style="list-style-type: none"> <li>• Generic technological solutions (neutral Grid) aspired to be used by academics vs. proprietary solutions desired by commercial partners.</li> <li>• Contrasting developers' cultures within the project. Battles between agile pragmatism of PAL and formalism of Software Engineering.</li> <li>• Virtual Organisation failed to reflect project structure. Technical issues inhibited collaborative use of Grid.</li> <li>• Adoption of quick and dirty Middleware solution.</li> <li>• Failure to appreciate importance of medical computing issues (e.g. security and confidentiality) and clinical demands (e.g user-friendly and speed of workflow).</li> </ul>	<ul style="list-style-type: none"> <li>• Gain understanding of how the different software will interface and how different teams will communicate to achieve such interfacing.</li> <li>• Consider Grid Virtual Organisation early within the project.</li> <li>• Develop coherent systems development practices.</li> <li>• Identify clear change control practices.</li> <li>• Clearly identify tasks and ensure low cohesion among tasks between different relevant social groups.</li> <li>• Develop modular development strategy with strong discussion of interfaces.</li> </ul>

**Table 2: Lessons from HealthscanGrid for future research projects.**

## 7 CONCLUSIONS

This paper addresses one specific form of collaborative eScience endeavour – those projects which align with the Mode 2 model of Knowledge Production (i.e. multidisciplinary, context-driven, problem-focused) in constructing Grid based systems (Gibbons, Limoges et al. 1994). Defining such projects as Applied eScience, we argue that the involvement of commercial and public collaborators, in addition to academics, adds complexity to eScience development projects and represents a distinct and neglected form of eScience research. By introducing a significantly new means of analyzing such projects (drawn from the Information Systems literature which has thus-far largely ignored eScience research (Scott and Venters 2007)) the paper begins a new stream of eScience research with a distinct focus on bureaucracy and technology.

Constructing a Grid reflects the challenges of infrastructure development and is contingent and uncertain, however within this there are “shared patterns, processes and emergent lessons” (Edwards, Jackson et al. 2007). One particularly significant barrier is the construction of path dependencies which constrain future development and adoption (Voss, Mascord et al. 2007) suggesting a need to consider a Grid’s use in practice throughout the development process. But the development of a Grid (like any information infrastructure) is rarely “build in an entirely top-down, orderly, blueprint-like-way” (Edwards, Jackson et al. 2007) and so demands a means of exploring the emergent web of socio-material relations (Voss, Mascord et al. 2007) involved in its construction. Such exploration requires that those involved in eScience Grid development appreciate the socio-cognitive structures of participants, and their influence on the socio-technical Grid. While we desire the creation of Grids as transparent information infrastructures, during their development they are far from transparent and far from black-boxed and the interpretation of them is thus significant. It is this problem that our Applied eScience framework aims to address.

The framework extends earlier work on socio-cognitive structures towards technology (Orlikowski and Gash 1994) into the area of eScience. Specifically for Applied eScience, the framework introduces two additional components to such socio-cognitive structures – Bureaucracy (as an institutional dimension) and the Socio-Technical Grid. These elements allow a focus on the Technological Frames’ influence on the Grid (as socio-technical), and also acknowledge that power is a significant influence on the action to construct a Grid. While incongruence between different Technology Frames will be significant, it is also bureaucracy which ensures certain relevant social groups have greater agency. It is hoped that the framework will prove useful for both researchers attempting to better understand collaborative eScience, and for those involved in funding, managing or participating in Applied eScience.

The key contributions of the paper are modest. The research framework addresses only Applied eScience projects though it may be relevant to other eScience. The framework’s foundation is socio-cognitive (Orlikowski and Gash 1994) and thus underplays the agency of technology within its analysis (HealthscanGrid’s problems are not attributed to failings in the Grid itself). While alternative frameworks

may have been used to provide such agency (Actor Network Theory for example (Callon 1986)) the complexity they add would detract from the aims of this study – analysis would quickly become lost in the complexities of eScience technology. Similarly, the framework does not consider the infrastructural nature of eScience (David 2004) and how eScience becomes embedded within the context-of-application, with network effects proving dominant (Voss, Mascord et al. 2007) and technology socially constructed in use over time. This seems to be an area ripe for future research. The social construction of Grids within Applied eScience using a SCOT analysis (Bijker, Hughes et al. 1987; Bijker 1995) would be particularly relevant.

## 8 REFERENCES

- Atkins, D. E. (2003). Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Arlington, VA, National Science Foundation.
- Avgerou, C. and T. Cornford (1998). *Developing Information Systems: Concepts, Issues and Practice*. Basingstoke, Palgrave.
- Avison, D. and G. Fitzgerald (2003). *Information Systems Development: Methodologies, Techniques and Tools*. London, McGraw-Hill.
- Avison, D., F. Lau, et al. (1999). "Action Research." *Communications of the ACM* **42**(1): 94-97.
- Baker, M., R. Buyya, et al. (2002). Grids and Grid Technologies for Wide-Area Distributed Computing. *In: Software Practice and Experience*, John Wiley & Sons.
- Beaulieu, A. and P. Wouters (2009). e-Research as Intervention. *e-Research: Transformation in Scholarly Practice*. N. Jankowski. Abingdon, Routledge: 54-69.
- Berman, F., F. Geoffrey, et al. (2003). The Grid: past, present, future. *In: Grid Computing- Making the Global Infrastructure a Reality*. F. Berman, F. Geoffrey and T. Hey, John Wiley & Sons.
- Bijker, W. (1995). *Of Bicycles, Bakerlites and Bulbs; Toward a theory of Socio-Technical Change*. Cambridge, MA, MIT Press.
- Bijker, W., T. Hughes, et al., Eds. (1987). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA, MIT Press.
- Borgman, C. (2007). *Scholarship in the Digital Age*. Cambridge, MA, MIT Press.
- Bos, N., A. Zimmerman, et al. (2008). From Shared Databases to Communities of Practice: A taxonomy of collaboratories. *Scientific Collaboration and the Internet*
- G. Olson, A. Zimmerman and N. Bos. Boston, MA, MIT Press: 1-12.
- Broadbent, M. and P. Weill (1999). "The implication of information technology infrastructure for Business-Process-Redesign." *MIS Quarterly* **23**(2): 159-182.
- Callon, M. (1986). Some Elements For A Sociology of Translation: Domestication of the Scallops and the Fishermen of St-Brieuc Bay. *Power, Action and Belief: a New Sociology of Knowledge?* J. Law. London, Routledge and Kegan Paul: 196-229.
- Carr, N. (2005). "The End of Corporate Computing." *MIT Sloan Management Review* **46**(3): 67-73.
- Carr, N. (2008). *The Big Switch: Rewiring the World, from Edison to Google*, W. W. Norton & Co.

- Checkland, P. (1995). Soft systems methodology and its relevance to the development of information systems. *Information Systems Provision*. F. Stowell. London, McGraw-Hill Book Company: 1-17.
- Checkland, P. and S. Holwell (1998). *Information, Systems and Information Systems*. Chichester, John Wiley & Sons.
- Chetty, M. and R. Buyya (2002). "Weaving computational grids: how analogous are they with electrical grids?" *Computing in Science & Engineering* 4(4): 61-71.
- Chompalov, I., J. Genuth, et al. (2002). "The Organization of Scientific Collaborations." *Research Policy* 31: 749-767.
- Ciborra, C. (2000). *From Control to Drift: The dynamics of Corporate Information Infrastructure*. Oxford, Oxford University Press.
- Cordella, A. (2006). Information Infrastructure in action, London School of Economics and Political Sciences, Department of Information Systems.
- David, P. A. (2004). Towards a cyberinfrastructure for enhanced scientific collaboration: providing its 'soft' foundations may be the hardest part. Oxford, Oxford Internet Institute.
- Davidow, W. H. and M. S. Malone (1992). *The virtual corporation: Structuring and revitalizing the corporation of the 21st century*. New York, Harper-Collins.
- Davidson, E. (2006). "A Technological Frames Perspective on Information Technology and Organizational Change." *The journal of applied behavioral science* 42(1): 23-39.
- Dodgson, M. (1994). Technological Collaboration and Innovation. *Handbook of Industrial Innovation*. M. Dodgson and R. Rothwell. Aldershot, UK, Edward Elgar Publishing Ltd.
- Edwards, P., S. Jackson, et al. (2007). Understanding Infrastructure: Dynamics, Tensions and Design. Report of a Workshop on History and Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures. . Ann Arbor, University of Michigan.
- Etzkowitz, H. (2000). "The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university-industry-government relations." *Research policy* 29(2): 109.
- Foster, I. and C. Kesselman (2004a). Concepts and Architecture. *The Grid 2*. I. Foster and C. Kesselman. San Francisco, Morgan Kaufmann: 37-63.
- Foster, I. and C. Kesselman (2004b). *The Grid 2: Blueprint for a New Computing Infrastructure*, Elsevier.
- Foster, I., C. Kesselman, et al. (2001). "The anatomy of the Grid." *International Journal of Supercomputer Applications* 15(3): 200-222.
- Galliers, B. (1997). Reflections on IS Research. *Information Systems: An Emerging Discipline?* J. Mingers and F. Stowell. London, McGraw-Hill: 141-157.
- Gibbons, M., C. Limoges, et al. (1994). *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London, Sage Publications.
- Goyal, B. and S. Lawandé (2006). *Grid Revolution: An Introduction to Enterprise Grid Computing*. New York, McGraw-Hill Osborne.
- Gregor, S. (2006). "The Nature of Theory in Information Systems." *MIS Quarterly* 30(3): 611-642.
- Grey, C. (2001). "Re-imagining Relevance: A Response to Starkey and Madan " *British Journal of Management* 12(Suppliment 1): S27-S32.
- Hanseth, O. and E. Monteiro. (1998). "Understanding Information Infrastructure." <http://heim.ifi.uio.no/~oleha/Publications/bok.html> Retrieved 15th of June, 2006.
- Hardy, C., N. Phillips, et al. (2003). "Resources, Knowledge and Influence: The Organizational Effects of Interorganizational Collaboration." *Journal of Management Studies* 40(2): 321-347.
- Harvey, J. (2002). "The Determinants of Research Group Performance Towards Mode 2." *Journal of Management Studies* 39(6): 747-774.
- Hatchuel, A. (2001). "The Two Pillars of New Management Research " *British Journal of Management* 12(Special Issue): 533-539.
- Haythornthwaite, C., K. Lunsford, et al. (2006). Research in Practice in Distributed, Interdisciplinary Collaboration. *New Infrastructures for Knowledge Production: Understanding E-Science*. C. Hine. Hershey, PA, Information Science Publishing: 143-166.

- Hermanowicz, J. (2009). *Lives in Science: How institutions Affect Academic Careers*. Chicago, University of Chicago Press.
- Hey, T. and A. Trefethen (2002). "The UK e-Science Core Programme and the Grid." *Future Generation Computer Systems* **18** 1017-1031.
- Hey, T. and A. Trefethen (2008). E-Science, Cyberinfrastructure, and Scholarly Communication. *Scientific Collaboration and the Internet*
- G. Olson, A. Zimmerman and N. Bos. Boston, MA, MIT Press: 15-31.
- Hine, C. (2006). *New infrastructures for knowledge production: Understanding e-science*. London, Information Science Publishing.
- Jankowski, N. (2009). The Contours and Challenges of e-Research. *e-Research: Transformation in Scholarly Practice*. N. Jankowski. Abingdon, Routledge: 3-24.
- Jha, S. and M. Watson-Manheim (2007). Conduct, Performance, & Dilemmas of Inter-Organizational Virtual Organizing: A Literature Review. *Virtuality and Virtualization: The International Federation for Information Processing (IFIP) Working Group 8.2 on Information Systems and Organizations and 9.5 on Virtuality and Society*. S. S. Wanda Orlikowski, Kevin Crowston, Eleanor Wynn, Cynthia K. Pickering.
- Jirotko, M., R. Procter, et al. (2005). "Collaboration and Trust in Healthcare Innovation: The eDiaMoND Case Study." *Computer Supported Cooperative Work* **14**: 369-398.
- Jirotko, M., R. Procter, et al. (2006). "Special Issue: Collaboration in e-Research." *Computer Supported Cooperative Work* **15**: 251-255.
- Jones, C. (2004). "Computing at CERN: the mainframe era." *CERN courier* **44**(7): 32-35.
- Klein, H. and D. Kleinman (2002). "The Social Construction of technology: Structural Considerations." *Science, Technology and Human Values* **27**(1): 28-52.
- Kline, R. and T. Pinch (1996). "Users as agents of technological change: The social construction of the automobile in the rural United States." *Technology and Culture* **37**(4): 763-795.
- Kline, R. and T. Pinch (1999). The Social Construction of technology. In: *The Social shaping of technology*. D. Mackenzie and J. Wajcman. Buckingham Open University Press.
- Knights, D. and H. Willmott (1997). "The Hype and Hope of Interdisciplinary Management Studies." *British Journal of Management* **8**(1): 9-14.
- Knorr-Cetina, K. (1999). *Epistemic Cultures: How the sciences make knowledge*. Cambridge, MA, Harvard University Press.
- Lave, J. and E. Wenger (1991). *Situated Learning: Legitimate Peripheral Participation*. New York, Cambridge University Press.
- Lawrence, K. (2006). "Walking the Tightrope: The Balancing Acts of a Large e-Research Project " *Computer Supported Cooperative Work* **15**(4): 385-411.
- Macbeth, D. (2002). "From Research to Practice via Consultancy and Back Again: A 14 Year Case Study of Applied Research " *European Management Journal* **20**(4): 393-400.
- MacLean, D., R. MacIntosh, et al. (2001). Mode 2 Management Research: probing the boundaries. *Academy of Management Conference*. Washington.
- McClatchey, R., D. Manset, et al. (2003). *The MammoGrid Project Grids Architecture*. Proceedings of the 10th International Conference on Computing for High Energy Physics (CHEP'03), San Diego, USA.
- Meyer, E. (2009). Moving from Small Science to Big Science. *e-Research: Transformation in Scholarly Practice*. N. Jankowski. Abingdon, Routledge: 147-159.
- Mitev, N. and W. Venters (2007). "Self-Reflexive Evaluation of an Academic-Industry Management Research Collaboration: Instrumentalisation in Developing Knowledge Management Technology." *Journal of Management Studies*(Under review).
- Mitev, N. and W. Venters (2009). "Self-Reflexive Evaluation of an Academic-Industry Management Research Collaboration: Instrumentalisation in Developing Knowledge Management Technology." *Journal of Management Studies* **46**(5): 733-754.

- Newell, S. and R. Galliers (2000). *More than a Footnote: The Perils of Multidisciplinary Research Collaboration*. Americas Conference on Information Systems, Long Beach, California.
- Newell, S., J. Swan, et al. (2001). "The Role of Funding Bodies in the Creation and Diffusion of Management fads and Fashions." *Organisation* **8**(1): 97-120.
- Newhouse, S., J. M. Schopf, et al. (2007). Study of User Priorities for e-Infrastructure for e-Research. *UK e-Science Technical Report UKeS-2006-07*.
- Olson, G., A. Zimmerman, et al. (2008). Introduction. *Scientific Collaboration and the Internet*
- G. Olson, A. Zimmerman and N. Bos. Boston, MA, MIT Press: 1-12.
- Orlikowski, W. (1992). "The duality of technology: Rethinking the concept of technology in organisations." *Organisational Science* **3**(3): 398-426.
- Orlikowski, W. J. and D. C. Gash (1994). "Technological Frames: Making Sense of Information Technology in Organizations." *ACM Transactions on Information Systems* **12**(2): 669-702.
- Orzech, D. (2003, March 14, 2003). "Politics, not Technology, is Biggest Barrier to Grid Computing." Retrieved 26 November 2006, 2006, from <http://www.gridcomputingplanet.com/news/article.php/2110181>.
- Otley, D. (1999). "Performance management: a framework for management control systems research." *Management Accounting Research* **10**: 363-382.
- Pinch, T. (1995). The Social Construction of Technology: A review. In: *Technological Change: Methods and Themes in History of technology*. R. Fox, Martin Dunitz.
- Plaszczak, P. and R. Wellner (2007). *Grid Computing: A Savvy Manager's Guide*. Amsterdam, Elsevier.
- POST (2007). Grids and e-Science. P. O. o. S. a. Technology.
- Pothen, P. (2007). Developing the UK's e-infrastructure for science and innovation. O. o. S. a. Innovation, OSI e-Infrastructure Working Group.
- Powell, W., K. Koput, et al. (1996). "Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology." *Administrative Science Quarterly* **41**(1).
- Rajasekar, A., M. Wan, et al. (2002). *MySRB & SRB - Components of a Data Grid*. Proceedings of the 11th International Symposium on High Performance Distributed Computing (HPDC-11).
- Rapoport, R. (1970). "Three dilemmas in action research." *Human Relations* **23**(6): 499-513.
- Rosenhead, J., Ed. (1989). *Rational Analysis for a Problematic World: Problem Structuring Methods for Complexity, Uncertainty and Conflict*. Chichester, John Wiley and Sons.
- Scott, S. and W. Venters (2007). The Practice of e-Science and e-Social Science: Method, Theory, and Matter. *Virtuality and Virtualization*. W. Orlikowski, E. Wynn, K. Crowston and S. Sieber. Boston, Springer: 267-279.
- Shrum, W., J. Genuth, et al. (2007). *Structures of Scientific Collaboration*. Cambridge, MA, MIT Press.
- Smarr, L. (2004a). Grids in Context. In: *The Grid2: blueprint for a new computer Infrastructure*. I. Foster and C. Kesselman, Elsevier Inc.
- Smarr, L. (2004b). Grids in Context. *The Grid 2*. I. Foster and C. Kesselman. San Francisco, Morgan Kaufmann.
- Starkey, K. and P. Madan (2001). "Bridging the Relevance Gap: Aligning Stakeholders in the Future of Management Research." *British Journal of Management* **12**(Supplement 1): S3-S24.
- Tranfield, D. (2002). "Future Challenges for Management Research." *European Management Journal* **20**(4): 409-413.
- Traweek, S. (1988). *Beamtimes and lifetimes: The world of high energy physics*. Cambridge MA, Harvard University Press.
- van Aken, J. (2005). "Management Research as a Design Science: Articulating the Research Products of Mode 2 Knowledge Production in Management." *British Journal of Management* **16**: 19-36.
- Venters, W. and T. Cornford (2006). *Introducing Pegasus: An ethnographic research project studying the use of Grid technologies by the UK particle Physics Community*. Second International Conference on e-Social Science, Manchester, UK.
- Voss, A., M. Mascord, et al. (2007). e-Research Infrastructure Development and Community Engagement. *All Hands Meeting* S. Cox. Nottingham.

- Weber, M. (1930). *The Protestant Ethic and the Spirit of Capitalism*.
- Weick, K. (1990). Technology as Equivoque: Sensemaking in New Technologies. *Technology and Organizations*. P. Goodman and L. Sproull. San Francisco, Jossey-Bass Publishers: 1-45.
- Wenger, E. (1998). *Communities of practice : Learning, meaning and identity*. Cambridge, Cambridge University Press.
- Wladawsky-Berger, I. (2004). The Industrial Imperative. *In: The Grid2: blueprint for a new computing Infrastructure*. I. Foster and C. Kesselman, Elsevier Inc.
- Wouters, P. and P. Schroder, Eds. (2003). *Promise and Practice in Data Sharing*. Amsterdam, NIWI-KNAW.

9

Pegasus Working Paper - DRAFT

## NOTES

---

<sup>1</sup> [www.gridcomputingnow.org.uk](http://www.gridcomputingnow.org.uk)

<sup>2</sup> See <http://www.rcuk.ac.uk/escience/default.htm> for details of the UK's research focus in eScience. EU Framework 7 funding will also focus on multi-disciplinary activity within its eScience programmes: See [http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52005PC0119\(02\):EN:HTML](http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52005PC0119(02):EN:HTML). Also [www.gridcomputingnow.org](http://www.gridcomputingnow.org).

<sup>3</sup> Note the significant debates in the UK over “Research Impact” evaluations within the new Research Excellent Framework <http://www.hefce.ac.uk/Research/ref/>

<sup>4</sup> eu-EGEE .org includes a “business task force”; and the Particle-Physicists’ GridPP boasts links with industry ([www.gridpp.ac.uk/wider/index.html](http://www.gridpp.ac.uk/wider/index.html))

<sup>5</sup> <http://www.bbc.co.uk/rd/projects/prism/>

<sup>6</sup> <http://www2.besc.ac.uk/financial-services>

<sup>7</sup> <http://www.e-science.soton.ac.uk/pages/projects.htm> - G-Civil

<sup>8</sup> <http://www.wesc.ac.uk/projectsite/covite/index.html>

<sup>9</sup> Applications on EGEE (2009) - <http://press.eu-egee.org/fileadmin/documents/infosheets/final/en/applications-on-egee-final.pdf>

<sup>10</sup> <http://glite.web.cern.ch/glite/alien/>

<sup>11</sup> EGEE and Globus are alternative Grid software stacks (known as Middleware)

<sup>12</sup> <http://cordis.europa.eu/fp6/stepbystep/project.htm>