

Introducing Pegasus: An ethnographic research project to study Grid technologies' use by the UK particle physics community.

Will Venters, Tony Cornford

London School of Economics,

Houghton Street,

London,

UK.

w.venters@lse.ac.uk

Abstract. This paper introduces a three year EPSRC research project to study the use of Grid technologies by the UK particle physics community as it prepares for the launch of a major new experiment at CERN. The paper describes the aims of the project, outlines its methodology, and provides an overview of its expected contribution. The research explores Grids within particle physics not as a static platform, but as contested and political accommodation. The paper describes the theoretical lenses we employ in researching the contested, political nature of the GridPP project alongside its technical realization and achievements.

Introduction

The relevance of this research to this conference theme lies in the exploration of a successful Grid development within physics (GridPP) and the possible insights that can be offered to other communities as they engage with and start to use grid technologies. The approach of the work uses the lens of social constructionism (Berger and Luckmann 1966) exploring how individuals and groups interact to construct their understanding of the Grid, and through this influence the form the technology takes. The paper describes the research context, the theoretical lenses we employ in researching the contested, political nature of the GridPP project alongside its technical realization and achievements (Hlistova 2004), and the form of contribution we seek to make.

Research context - GridPP

GridPP is the UK particle physics Grid and has been under development and in operation for a number of years. At this point in time the development of this Grid, undertaken by the physics community is in large part driven by the need to analyse the unprecedented amounts of data (some 15 Petabytes per year) that will be produced by the LHC (large hadron collider)

experiments currently under construction at CERN¹ and which is due to begin operation in 2007. To process this data GridPP envisage requiring 100,000 computers forming its associated grids by 2007, spread across the globe and incorporating a number of grid infrastructures of which GridPP, from the UK, is one (Faulkner, Lowe et al. 2006). Clearly the technical tasks required in achieving this will be significant, however our interests lies rather in the processes and practices of bringing such Grid technology, taken as a socio-technical system rather than purely technical artefact, into widespread use. We are in particular concerned with the ways that Grid technology is accommodated into the working practices of physicists; how it is shaped by these ‘users’ (albeit highly technical users); and how this in turn shapes the LHC experiments and more broadly the character of experimental particle physics.

Particle Physics is a very appropriate area for this research and has, for a long time, exhibited an ability to assimilate new high performance computing resources within its work practices in a successful and highly pragmatic way. As an interviewee in our preliminary research highlighted “... *the particle physics community is usually a few years ahead in terms of the amount of computing it is trying to use and the amount of data it is trying to manage*”. CERN for example accepted the problems of working with pre-production supercomputers from the days of the CDC 6600 through to the CRAY X-MP (Jones 2004). Later they pioneered work on the Web (Berners-Lee and Fischetti 1997), shifted early to use Open-source (Linux) server-farms; and more recently have led early development of the Grid (Lloyd 2003). Yet throughout these transitions their focus has been resolutely upon the application of these complex (and somewhat temperamental) technologies to doing their science: “*the computer scientists will put together the most elegant thing in the universe but it will never work... Physicists will come up with the most hacked solution in the world...but it will work.*” (*experimental particle physicist involved in GridPP*). Furthermore in doing this they have demonstrated their ability to work pragmatically as a globally distributed community that thrives on debate and discussion (Knorr-Cetina 1999).

The nature of experimental particle physics work is of course quite distinctive. The experiments that underlie the field require very large capital investment beyond the possible budgets of most individual national science programmes, or their human resources. Due to their cutting-edge nature and their complexity they also have very long lead in periods on their construction, followed by long periods of operation. Intriguingly, such experiments are achieved within organizational structures which have very few formal lines of authority, working as collaborations (virtual organizations) of different people and institutional actors with different (though broadly aligned) aims – a very different kind of organization from those traditionally constructing such complex technical systems (for example NASA or National Rail).

Our research is based on the belief that lessons from this community’s development and use of their Grid can offer insights for other fields with less history of engaging with technology or developing virtual organizational forms as they enter the field of e-science and e-social science (Foster, Kesselman et al. 2001). And yet we are aware that this may be difficult for a range of reasons.

Firstly particle physicists have focused on developing Grids to achieve results in the area of physics, not as a generic tool: as one respondent put it when reflecting on their role in developing the generic grid technology, “*They won’t really solve the problem of the Grid...*”

¹ <http://www.cern.ch>

it's a waste of time giving the money to particle physicists because they'll solve their own problems and they won't solve the problems for the biologists...". This leads to the second reason, the lessons we may draw from this community are embedded in the 'doing' of particle physics and within its complicated indeed unique context of massive physical apparatus, stunning data volumes and traditions of large collaborative teams.

Therefore, if there are lessons to be drawn from projects such as GridPP that can be usefully shared with the wider e-science community we argue that it is necessary to look beyond the simple replication of a technological infrastructure within another domain. Instead GridPP should be considered as a socio-technical infrastructure such that the context within which GridPPs practices and accomplishments exist may be better understood. From this position, and with a perspective that establishes a grid as a reflection and construction of its context, valuable insights can be gained. In brief we can learn about e-science and grids from physics, but not by attempting to become physicists! We therefore turn to fields within the social sciences to provide theoretical framework for our analysis.

Theoretical Point of Departure

In this research we take constructs within the field of information systems as our study's theoretical point of departure. Our preliminary work for the project drew on actor network theory as its theoretical lens (Hlistova 2004) to highlight the interdependence of GridPP as a technical infrastructure with the wider physics infrastructure and the funding regimes. In particular it exposed the socio-political realities which inscribe themselves in the GridPP technology through the funding regimes which enable its activity.

The Pegasus project's research aims are somewhat different from this preliminary research. The project is funded by the EPSRC as part of a programme to address the "*Usability Challenges to emerge from e-Science*" and which aims to address the following three broad questions (taken from the EPSRC funding programme²)

1. *What are the most appropriate design methods, approaches and techniques for e-Science systems?*
2. *What methods are appropriate for understanding scientific practices and what are the best ways of presenting these understandings to designers?*
3. *What sort of tools can be used to support the design of e-Science systems?*

To address these questions we draw upon the socio-technical traditions from the field of information systems. The project thus exploits a strong strand of recent work within the information systems field looking at the nature of infrastructures and their role as a constitutive element of organizations of various types (Star and Ruhleder 1996; Ciborra 2000; Hanseth 2000; Ciborra 2002). Informed by such work, we see infrastructures not as static technical platforms driven by engineering rationality and fixed common standards, but rather as political accommodations and arenas within which issues of work practices, resources, and organizational mission are surfaced and, to a degree, resolved. This leads us to focus on the improvised, and bricolaged (Ciborra 2002) actions which are employed by the developers and users of GridPP alongside consideration of their (more or less) systematic design and management practices.

GridPP is an emerging infrastructure for experimental particle physics research, in particular in relation to LHC experiment. To study such emergence implies a variety of area of concern

² <http://www.epsrc.ac.uk/CMSWeb/Downloads/Calls/UsabilityChallengesFrome-ScienceCall.pdf>

in the empirical aspects of this research. For example it is relevant that GridPP collaborators have developed their own management processes, practices and tools alongside the technical development activity and that such management tools, largely, grow out of the their established ways of doing science, rather than being informed by the more general traditions of IT development and management. Another example that we have noted in our preliminary work is the way that engagement in Grid projects has implications for individual physicists career trajectory, and how knowledge and experience in this area effect status.

Such a view of infrastructure extends the focus of this work beyond the accomplishment of any specific technical aims of GridPP (such as the ability to submit LHC analysis jobs, or calculate the data necessary for Monte Carlo simulations), to consider the grid as an infrastructure that is imbricated in the working practices of experimental particle physicists (Sassen 2002). This then leads us to frame the project as a balance of research that addresses *action of doing particle physics (research question 2)* alongside the *action for doing particle physics (research questions 1 and 3)*. The perspective we take is thus that GridPP is itself constitutive of science rather than simply its enabler, a tool or an unproblematic ideal 'platform' (Latour 1987). The research questions posed above concern not simply GridPP's technical character and outcomes (success as technology), but rather how (and to what extent) it comes to form part of the everyday lives of individual experimental particle physicists and the teams and groups they work within.

Considering GridPP as an infrastructure suggests that "success" would lead GridPP to be transparent in use, supporting tasks across space or time (Star and Ruhleder 1996), and invisible within the ongoing working practices of the experimental particle physicists. It is only upon breakdown in such transparency (Ciborra 1996) that infrastructures make their existence known (and perhaps lead to questioning of their "usability"). And yet such transparency requires a dynamic alignment with the working practices of the community of particle physicists who have membership of the infrastructure. In this way the infrastructure is embroiled with the community of practice (Wenger 1998) of its members, and so needs to link with the conventions of this group's practices (Star and Ruhleder 1996). This suggests that the community is shaped by, and shapes the Grid infrastructure in use; just as "the way that cycles of day-night work are affected by and effect electrical power rates and needs" (Star and Ruhleder 1996).

In this research we are thus led to reexamine the questions of usability since "usability" does not simply imply a tool which is designed by a third party to be "usable", but also the extent to which the practices of the community have accommodated (or domesticated) it or have changed to reflect the presence and potential of the tool. Only through a longitudinal and ethnographic study can such issues be surfaced.

Star and Ruhleder (1996) further argue that infrastructures are built upon history and wrestle with the "inertia of the installed base", inheriting its weaknesses and strengths. So GridPP and its technical elements, architecture and standards embody the conflict of historic practices, revealing both previous successes and scars of conflict within their form. This accretion of history and the dynamics of remaining operational and coherent in the face of new demands and requirements, leads to the concept of infrastructures as 'drifting', "as an infrastructure is used, new ways of using it are discovered; it drifts and "meets" other infrastructures" (Hanseth and Braa 2000). We suspect such drift will be a significant part of our research, as GridPP comes to be used alongside the LHC experiment, and forms part of wider Grid enabled physics research across the globe.

The usability questions outlined above direct our project to consider other elements of particle physicist's practices. As in all usability research, it is necessary to understand the way in which various actors interact with each other and technical components. Within the field of human-computer interaction, Hollan *et al* (2000) suggests that a specific focus on distributed cognition "provides a radical re-orientation of how to think about designing and supporting human-computer interaction...specifically tailored to understanding interactions among people and technologies". We see distributed cognition (Hutchins 1995), and the methodological insight of Hollan *et al* (2000) as providing a complementary theoretical foundation for the empirical research. We thus now outline the methodological approach adopted.

Methodological Approach

Influenced by both distributed cognition (Hollan, Hutchins et al. 2000) and our socio-technical perspective towards infrastructure, the research project will combine ethnography of the work practices of experimental particle physicists, and of those involved with the design and implementation of GridPP.

A qualitative longitudinal research methods is chosen as appropriate because of the focus on discourse and communication within particle physics (Knorr-Cetina 1999). We employ ethnography as a main research technique in order to gain an account of the situated, contextual nature of the commingled design and use of GridPP. Ethnography employs direct observation by researchers embedded within the context over an extended time period (Hammersley and Atkinson 1995; Agar 1996). Researchers on this project will spend two years charting the development of GridPP, employing audio and video recordings of meetings, interviews, and through extended periods spent with Particle Physicists across the UK and at CERN, we will also employ video-conferencing to undertake interviews with global participants as the community are regular users of this technology. We begin this work with a pilot study of the use of Grid technologies on the CDF experiment at Fermi National Accelerator Laboratory near Chicago.

Anticipated Contributions

Drawing from the interpretive research tradition in information systems (Walsham 1993) the focus in the analysis will be on sensemaking, learning and organising (Weick 1995). At this initial stage it is difficult to anticipate the contributions our ethnography might make, however the project has set itself the objective of developing a framework of qualifying guidance for the development of e-science infrastructure. This framework aims to address the use of Grids by collaborations of disparate scientists working as a virtual collaborative organization similar to GridPP. Our aims is to provide, through ethnographic accounts of GridPP, a coherent set of studies of how the LHC requirements were translated into Grid development; how the potential of e-science and GridPP is understood and appropriated by particle physicists; and to understand how the doing of e-science is achieved within this unique context. We anticipate lessons from this project will have wider implications with the information systems field, particularly around the development of complex infrastructures such as those implied by "software as services", open-source and virtual organizations.

References

Agar, M. (1996). *The Professional Stranger: An informal introduction to Ethnography*. San Diego, Academic Press.

- Berger, P. and T. Luckmann (1966). *The social construction of reality*. London, Penguin Books.
- Berners-Lee, T. and M. Fischetti (1997). *Weaving the Web*. San Francisco, CA, Harper.
- Ciborra, C., Ed. (1996). *Groupware & Teamwork - Invisible aid or technical hinderance*. Chichester, John Wiley & Sons Ltd.
- Ciborra, C. (2000). *From Control to Drift: The dynamics of Corporate Information Infrastructure*. Oxford, Oxford University Press.
- Ciborra, C. (2002). *The Labyrinths of Information: Challenging the Wisdom of Systems*. Oxford, Oxford University Press.
- Faulkner, P., L. Lowe, et al. (2006). "GridPP: development of the UK computing Grid for particle physics." *Journal of Physics G: Nuclear and Particle Physics*. **32**: N1-N20.
- Foster, I., C. Kesselman, et al. (2001). "The anatomy of the Grid." *International Journal of Supercomputer Applications*.
- Hammersley, M. and P. Atkinson (1995). *Ethnography: principles and practice*. London, Routledge.
- Hanseth, O. (2000). The Economics of Standards. *From Control to Drift*. C. Ciborra. Oxford, Oxford University Press: 56-70.
- Hanseth, O. and K. Braa (2000). Who's in Control: Designers, Managers - or Technology? Infrastructures at Norsk Hydro. *From Control to Drift*. C. Ciborra. Oxford, Oxford University: 125-147.
- Hlistova, J. (2004). How is the Particle Physics Community Shaping the Grid in the UK? *Department of Information Systems*. London, London School of Economics: 38.
- Hollan, J., E. Hutchins, et al. (2000). "Distributed Cognition: Towards a New Foundation for Human-Computer Interaction Research." *ACM Transactions on Computer-Human Interaction* **7**(2).
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA, The MIT Press.
- Jones, C. (2004). "Computing at CERN: the mainframe era." *CERN courier* **44**(7): 32-35.
- Knorr-Cetina, K. (1999). *Epistemic Cultures: How the sciences make knowledge*. Cambridge, MA, Harvard University Press.
- Latour, B. (1987). *Science In Action*. Cambridge MA, Harvard University Press.
- Lloyd, S. (2003). "GridPP - From web to Grid." *Frontiers*(16).
- Sassen, S., Ed. (2002). *Global Networks: Linked Cities*, Taylor & Francis, Inc.
- Star, S. L. and K. Ruhleder (1996). "Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces." *Information Systems Research* **7**(1): 111-135.
- Walsham, G. (1993). *Interpreting Information Systems in Organizations*. Chichester, Wiley.
- Weick, K. (1995). *Sensemaking in Organisations*. London, Sage Publications.
- Wenger, E. (1998). *Communities of practice : Learning, meaning and identity*. Cambridge, Cambridge University Press.

Acknowledgments

Pegasus is funded by the UK EPSRC (Grant Number EP/D049954/1) under the e-science program. The project includes participants from University College London and LSE. The authors acknowledge the considerable contribution of Jana Hlistova's in undertaking the initial analysis.