MIS Quarterly

SPECIAL ISSUE: SOCIOMATERIALITY OF IS & ORGANIZING

A TRICHORDAL TEMPORAL APPROACH TO DIGITAL COORDINATION: THE SOCIOMATERIAL MANGLING OF THE CERN GRID

Will Venters
London School of Economics, London, UNITED KINGDOM {w.venters@lse.ac.uk}

Eivor Oborn
Warwick University, Coventry, UNITED KINGDOM {eivor.oborn@wbs.ac.uk}

Michael Barrett
University of Cambridge, Cambridge, UNITED KINGDOM {m.barrett@jbs.cam.ac.uk}

Appendix A

In total, 70 interviews were undertaken for this study. More were held with LCG than with CMS as these were needed for a technical understanding of the WLCG and also because CMS's documentation provided a large amount of detail of analysis practices. The choice of interviewees aimed to be representative of the range of jobs descriptions and seniority levels within the collaborations and WLCG. Interviewees were chosen through attending meetings and approaching people who appeared relevant and knowledgeable, examining member lists, and through snowball sampling (asking interviewees for suggestions of other relevant people). For pragmatic reasons, interviews were mostly of physicists and CERN members associated with UK institutions (although not necessarily based in the UK) or working directly at CERN. This limitation does not significantly affect the findings as UK participants are highly active within experiments and WLCG. Further, UK institution's participants are international, reflecting the diversity evident at CERN.

Interviews were semi-structured with open-ended questioning on the working of the grid and its use in particle physics, with probing to explore the practices (see Appendix B for details of questions). Particular emphasis was given to asking interviewees to recount what happened in doing physics analysis. Most interviews lasted around one hour.

The following table provides details the job descriptions of those interviewed.

Table A1. Details of Job Descriptions for Interviewees

Job Description	Code
Lecturer in particle physics undertaking CMS analysis	i1
CMS representative to UK WLCG, reader in particle physics undertaking CMS analysis	i2
Professor of particle physics undertaking CMS analysis and member of UK WLCG board	i3
Post-doctoral Research Assistant undertaking physics analysis on CMS	i4
Post-doctoral CMS physicist and developer of software for CMS	i5
Post-doctoral Research Assistant undertaking physics analysis on CMS	i6
PhD student undertaking physics analysis on CMS	i7

Table A1. Details of Job Descriptions for Interviewees (Continued)	
Job Description	**Code**
Physics analysis statistical package developer (for CMS)	i8
Management of LCG experimental interfaces (WLCG and CMS)	i9
CMS software developer	i10
CMS software developer also undertaking CMS analysis	i11
CMS software developer	i12
CMS software developer	i13
CMS software developer also undertaking CMS analysis	i14
Involved in writing CMS simulation software for analysis activity using the grid	i15
WLCG middleware developer	i16
WLCG developer	i17
WLCG site manager	i18
WLCG development and workload manager development	i19
WLCG developer	i20
WLCG developer	i21
WLCG developer	i22
WLCG site administrator	i23
WLCG middleware developer	i24
Lecturer in Particle Physicist who works on WLCG	i25
WLCG storage manager	i26
WLCG deployment manager	i27
WLCG middleware developer	i28
WLCG deployment staff	i29
WLCG project manager	i30
A previous senior manager of the integration of experimental software with grid services	i31
Post-doctoral particle physicist and software developer	i32
WLCG middleware developer	i33
WLCG middleware developer	i34
WLCG middleware test and quality assurance systems	i35
WLCG pre-production systems developer	i36
WLCG middleware developer	i37
WLCG core developer	i38
Board Member of UK part of LCG	i39
WLCG middleware developer and systems administration.	i40
Particle physicist and developer of an LCG grid monitoring application	i41
Senior manager of LCG; also a particle physics professor	i42
WLCG developer	i43
Particle physicist and computing coordinator for an experiment	i44
WLCG developer	i45
Particle Physicist and developer involved in documentation and writing user guides	i46
WLCG developer	i47
WLCG deployment staff	i48
Manager of WLCG data center and senior WLCG member; previously a particle physicist	i49
WLCG systems administrator	i50
Particle physics lecturer	i51

Table A1. Details of Job Descriptions for Interviewees (Continued)	
Job Description	**Code**
One of the overall directors of WLCG	i52
Theoretical particle physicist	i53
WLCG security specialist	i54
WLCG deployment	i55
Particle physics user support	i56
WLCG quality manager	i57
Board member of UK part of WLCG	i58
Press Officer (WLCG related project)	i59
WLCG deployment	i60
Senior member of UK part of WLCG	i61
Press Officer (CERN - WLCG)	i62
WLCG research and development project manager	i63
Particle physics professor and board member of UK WLCG with oversight for other experiments use of WLCG	i64
Particle physics professor and board member of UK WLCG with oversight for other experiments use of WLCG	i65
Particle physicist with wide understanding of grid	i66
Head of computing for a CERN experiment (ATLAS)	i67
Professor of Particle Physics and member of UK WLCG board	i68
Particle Physics post-doctoral researcher	i69
Particle Physics post-doctoral researcher	i70

Appendix B

Informants were asked a limited number of core questions (see examples below), and probed for elaborations and explanations. The questions were tailored for different interviewees but with emphasis on understanding work practices and technology of the grid as the goal for each interview. Interviews were recorded (with permission) and professionally transcribed, with transcriptions checked against the recordings. Where recording was not possible, notes were taken. As a snowball sampling method was used, many interviews ended with the question, "Who should we talk to next?"

Example questions for those undertaking analysis activity:

- What work are you doing?
- What are you doing now [today]?
- Are you "doing physics" – if so what, why, how?
- What does your job involve?
- What are the challenges in your job?
- Do you face challenges in using the grid?
- What do you do when there are problems with the grid?

Example questions for those who develop software for LCG or CMS:

- What are the main problems with the grid?
- What is your opinion of the middleware?
- Do you develop software and applications? What kinds? How do they work? What do they do?
- How do you develop them? How do you identify the problem? How do you come up with a solution?
- What challenges does an ordinary user face in using the grid?
- Is there a boundary in your work between doing physics and doing computing?
- When does computing stop and physics start?

Appendix C

Extensive documentation was reviewed in order to understand physics analysis, WLCG, CRAB and CMS. The following table describes the key sources used. In addition to these formal sources various informal and unpublished emails, documents and presentations were reviewed. Online sources may have changed.

Table C1. Key Sources

Document	Description	Code
http://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCRABTutorial	Wiki page of various CMS analysis activities (including using CRAB).	d1
https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideCrabFaq	Wiki page of CRAB frequently asked questions.	d2
http://www.hep.wisc.edu/cms/comp/crab/	University teaching page on using CRAB.	d3
http://cms.cern.ch	The CMS website providing copious technical information on CMS for its physicists (requires CERN-ID).	d4
http://wlcg.web.cern.ch/	The website for WLCG providing copious technical information (requires CERN-ID).	d5
https://espace.cern.ch/WLCG-document-repository/default.aspx	WLCG document repository with various public documents including some technical documents.	d6
http://www.gridpp.ac.uk/	This site details the UK's contribution to WLCG. GridPP was heavily involved in the development of many areas of WLCG, and its members contribute around 20,000 computers to the WLCG. These pages provide technical overviews.	d7
http://cerncourier.com/cws/latest/cern	The CERN courier is CERN's official magazine and includes details of experiments progress and computing developments. Relevant articles were reviewed.	d8
www.ggus.org	GGUS and its documentation. This service has changed considerably since the research period.	d9
Burke, S., Campana, S., Lanciotti, E., Litmaath, M., Lorenzo, P. M. E., Miccio, V., Nater, C., Santinelli, R., and Sciaba, A. 2012. "gLite 3.2 User Guide (Manuals Series)," EGEE, wLCG, Geneva.	This is the 137 page user-guide for WLCG. It details the technical architecture of WLCG including job submission using gLite and WLCG's workload manager (rather than CRAB for example).	d10
Faulkner, P. J. W., Lowe, L. S., Tan, C. L. A., Watkins, P. M., Bailey, D. S., et al. 2006. "GridPP: Development of the UK Computing Grid for Particle Physics," *Journal of Physics G: Nuclear and Particle Physics* (32), pp. N1-N20.	Research paper outlining the development and operation of WLCG (within the UK). This paper has 197 authors hence author list is truncated.	d11
Gutsche, O. 2007. "CRAB: Introduction," CMS/ Fermilab, Batavia, Il.	A PowerPoint presentation of 25 detailed technical slides teaching CMS physicists how to use CRAB.	d12
Gutsche, O. 2007. "CRAB: Debugging Techniques," CMS/ Fermilab, Batavia, Il.	A PowerPoint presentation of 26 detailed technical slides teaching CMS physicists how to debug their crab.cfg files.	d13
Grid Log Files	Two years of WLCG log files were made available, amounting to around 4Gb of files. These were reviewed statistically (though not extensively).	d14
http://cmsdoc.cern.ch/cms/ccs/wm/scripts/Crab/	This site provides the full archive of the CRAB source code. The CRAB source code contains around 110 python script files, and numerous data and configuration files (including a generic crab.cfg).	d15

Table C1. Key Sources (Continued)

Document	Description	Code
CMS 2005. "CMS: The Computing Project Technical Design Report," CMS, Geneva.	This 169 page document was the initial technical design report outlining how CMS would use computing for the WLCG. This includes discussion of workload management (defined differently from WLCG).	d16
Carminati, F., Cerello, P., Grandi, C., Van Herwijnen, E., Smirnova, O., and Templon, J. 2002. "Common Use Cases for a HEP Common Application Layer (HEPCAL) - LHC Computing Grid," CERN, Geneva.	This document sets out the standard UML use-cases for the WLCG. It is common across all experiments and thus describes the generic work of physicists using the WLCG. It also defines the key terms such as "jobs."	d17
Heavey, A., Lassila-Perini, K., and Williams, J. (eds.). 2006. *The CMS Offline Workbook*, CERN, Geneva.	This was the 409 page soft-bound handbook given to each physicist within the CMS collaboration that provides a comprehensive manual for doing physics analysis of CMS data.	d18
Spiga, D. 2007. "CMS Workload Management," *Nuclear Physics B-Proceedings Supplements* (172), pp. 141-144.	Research paper written by CMS developers providing an overview of the purpose of CMS.	d19
Spiga, D., Lacaprara, S., Bacchi, W., Cinquilli, M., Codispoti, G., Corvo, M., Dorigo, A., Fanfani, A., Fanzago, F., and Farina, F. 2008. "CRAB: The CMS Distributed Analysis Tool Development and Design," *Nuclear Physics B-Proceedings Supplements* (177), pp. 267-268.	Research paper providing technical detail of the operation of CRAB for CMS.	d20
Magini, N., Ratnikova, N., Rossman, P., Sánchez-Hernández, A., and Wildish, T. 2011. "Distributed Data Transfers in CMS," IOP Publishing.	Research paper discussing data transfer between sites of CMS.	d21
Codispoti, G., Mattia, C., Fanfani, A., Fanzago, F., Farina, F., Kavka, C., Lacaprara, S., Miccio, V., Spiga, D., and Vaandering, E. 2009. "CRAB: A CMS Application for Distributed Analysis," *IEEE Transactions on Nuclear Science* (56:5), pp. 2850-2858.	Research paper detailing CRAB architecture and usage of CRAB until 2009.	d22
Andreeva, J., Anjum, A., Barrass, T., Bonacorsi, D., Bunn, J., Corvo, M., Darmenov, N., De Filippis, N., Donno, F., and Donvito, G. 2004. "Use of Grid Tools to Support CMS Distributed Analysis," in *Nuclear Science Symposium Conference Record* (4), pp. 2029-2032.	Early paper discussing CMS proposals for interfaces to WLCG.	d23
Britton, D., Clark, P., Coles, J., Colling, D., Doyle, A., Fisher, S., Irving, A., Jensen, J., McNab, A., and Newbold, D. 2004. "A Grid for Particle Physics—From Testbed to Production," GridPP.	Basic introduction to WLCG and CMS analysis activity.	d24
Grandi, C. 2004. "CMS Distributed Data Analysis Challenges," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* (534:1), pp. 87-93.	Research paper describing the data challenges used to test the WLCG's capacity for CMS analysis.	d25
http://pprc.qmul.ac.uk/~lloyd/gridpp/ukgrid.html	Website providing a real-time-monitor of UK WLCG sites.	d26
http://gridtalk-project.blogspot.co.uk/	Blog produced by various WLCG and CERN computing members reporting on e-Science conferences.	d27
Fanfani, A., Afaq, A., Sanches, J. A., Andreeva, J., Bagliesi, G., Bauerdick, L., Belforte, S., Bittencourt Sampaio, P., Bloom, K., and Blumenfeld, B. 2010. "Distributed Analysis in CMS," *Journal of Grid Computing* (8:2), pp. 159-179.	Research paper detailing CMS analysis system exploiting the WLCG. This includes a discussion of CRAB.	d28

Table C1. Key Sources (Continued)

Document	Description	Code
Britton, D., Clark, P., Coles, J., Colling, D., Doyle, A., Fisher, S., Irving, A., Jensen, J., McNab, A., and Newbold, D. 2004. "A Grid for Particle Physics—From Testbed to Production," Glasgow University Physics Department Working Paper, GLAS-PPE/2004-05.	Research paper outlining basics of running a grid for particle physics	d29
Bird, I. G., Jones, B., and Kee, K. 2009. "The Organization and Management of Grid Infrastructures," *IEEE Computer* (42:1), pp. 36-46.	Research paper authored by the WLCG leader, and the director of the WGEE project. The paper outlines the fundamentals of the WLCG from the perspective of its leaders.	d30
Jones, C. 2004. "Computing at CERN: the Mainframe Era," *CERN Courier* (44:7), pp. 32-35.	An article in the CERN Courier (the CERN magazine available throughout CERN) discussing the history of computing at CERN	d31
Barnier, M., Calmy-Rey, M., Curien, H., and Aymar, R. 2004. "Infinitely CERN, 1954-2004: Memories of Fifty Years of Research," Geneva, CERN.	A large 300 page hardback book of photographs and articles celebrating 50 years of CERN. Published by CERN and funded by various sponsors it is considered an official account and sold in the CERN museum.	d32
http://indico.cern.ch/conferenceDisplay.py?confId=113877	WLCG UK Deployment Team meeting minutes from November 23, 2010.	d35
Faulkner, P. J. W., Lowe, L. S., et al. 2006. "GridPP: Development of the UK Computing Grid for Particle Physics," *Journal of Physics G: Nuclear and Particle Physics*. (32), pp. N1-N20.	This research paper was produced by the UK WLCG computer specialists (it has over 100 authors) and outlines their contribution and achievements in developing their grid.	d36
Knobloch, J., Bird, I., Bos, K., Brook, N., et al. (eds.). 2005. *LHC Computing Grid: Technical Design Report*, Geneva, CERN.	The technical design report setting out how the initial grid services would be constructed and prepared for the LHC launch date.	d37
Innocente, V. 2003. "CMS on the GRID: Towards a Fully Distributed Computing Architecture." *Nuclear Physics B-Proceedings Supplements* (120), pp. 113-118.	Research paper detailing early decisions in the data analysis architecture for CMS	d38

Appendix D

The following tables provide evidence of the analysis process.

Table D1. Data Analysis Approach		
Steps	**Tasks**	**Outputs**
1. Gaining a familiarization with the setting under study.	Produce summaries of the key activities centered on the key themes from the data as temporally unfolding.	Table E2. Case study description. Summary documents. Table E7.
2. Elaborate what was happening within the case in mangle of practice terms, exploring intentionality, resistance, and accommodations.	Produce a detailed description of the different groups, and their intentionality. Elaborate the summary documents.	Summary documents. Table E2, E3, E4, E7, and E8.
3. Considering the interesting mangling within this the work practice of the macroactors identified. Revising the theoretical perspective to focus on temporality.	Seeking the harnessing of material agency within the case study. Identifying the various resistance faced, and how such resistance was tuned. Exploring the temporal emergence of such practices.	Table E3. Detailed summaries and documents. Sparse versions of the analysis document. Further interviews.
4. Engaging in a systematic analysis of the mangling using the theory of the mangle of practice extended to consider agency as a chordal triad of agency.	A systematic analysis was of the empirical material and interviews using the developed theoretical framework incorporating Emirbayer and Mische's (1998). The chordal dimensions of the agency within the case are identified.	Tables E5 and E6. Analysis section of this paper.

Table D2. Initial Statements Summarizing Key Points in Interviews Relevant to How the WLCG Was Made to Work†	
Examples of initial analysis statements written in Atlas.ti against interview transcripts.	
Automatic continuous production jobs where failure doesn't really matter because running 150,00 at a time continuous flow. CERN is a great place to work. Challenge of providing users with support and keeping everyone happy. Chaotic analysis: will the grid be able to support it? CMS users are also involved in testing the grid. Complaint of users not being satisfied. Complex job roles: Maintainers are often part time physicists as well. Computing is physics and the integration of computing into work practice. Control over the grid through programming. CRAB as an interface to a loose system... the grid. Define the Users as those who submit jobs. Experiments building their own software and not following formal processes. Experiments developing their own software. Experiments have developed their own systems, their own code. Experiments have power over grid and can find their own ways. Experiments Interfaces (e.g., CRAB). Experiments oversee the development of the grid. For PP this is vocational its Mecca; they are driven. GGUS available to support users. Grid is complicated, and with multiple application systems the cost increase exponentially. Grid organized as federations...across the world. Grid problems cannot always be predicted; problems with hardware. How the experiments write software for their own purposes. Job failure not sure where jobs are running. Job failure to run. Jobs failing. Lack of accounting, lack of plans. Maintainers opinion of CRAB as designed for the physics not generic grid. Materiality of the software code. Most of a physicists time is doing computing. Need for fault-tolerant design. Need the software to analyze the data. Network latency between sites avoided through dedicated communication links. No way for users to fix problems with the grid.	No one has done this before. People work together when its about the LHC working.... Perceptions of grid problems worse than reality. Poorly managed sites with poor staff to manage them. PP are savvy programmers and will produce things. PP people who write software. PP think they are good at computers but doesn't mean they know how to hold a keyboard. Problems with grid middleware quality. Problems with the workload manager for users. Scripts—spend time developing them, I've made my scripts public. Seamless technology of the grid. Sites which used to provide cluster computing now do so as grid. Somehow the grid systems try to cope with all the requirements for different communities. Super users and going to say who gets the resources. The complexity of the grid and its impact on everyone. The complexity of the underlying is hidden from everyone. The experiments have greater power and less dependency on the central services. The experiments requirements for the grid are unclear. The experiments wrote their own software. The foundations of grid are shaky. The functionality is drive by the applications. The grid is about getting the experiments to work. The legacy of the Linux PC was people expect it to work. We got used to it with UNIX workstations and need to relearn our old stuff from the mainframe. The pressure on experiments to control their own software . Grid is materially changing throughout the process. The role of analysis code and improving analysis code. The sites have different focus and reflect different experiments. The value of local sites and knowing about sites. Users don't know what is going wrong. Users initializing jobs and finding errors. Users need to understand the materiality of the grid. Volume of jobs using the grid.

†The aim was not theoretical synthesis but manageability.

Reference

Emirbayer, M., and Mische, A. 1998. "What Is Agency?," *American Journal of Sociology* (103:4), pp. 962-1023.

Appendix E

The following tables result from the analysis. They are organized into three temporal dimensions, present, future, and past (reflecting the analysis).

Tensions That Predominate Around Gaining Grid Transparency (In Present)

The following two tables provide evidence to support the assertion that tensions exist in gaining grid transparency.

Table E1. Resistance from the Grid Faced by Computer Specialists in the Present

Material Resistance	Examples from the Field
Coordinating the evolution of sites joining the grid	"It's the fact that [grid sites] are autonomous in a sense that they don't have to do what we tell them to do. So everybody has to agree that they will deploy a certain thing and these time lines and so on, so that the most difficult thing is just getting people to do things in a coordinated fashion and do it in a timely way." (i65) "A large number of services across a large number of sites, something is happening almost every day you are having to upgrade it or fix it or something like that. And, again, because of the nature of the grid, if you do something here, you do affect many other sites." (i27)
Heterogeneous hardware which require testing in use.	"They've all got slightly differences in terms of hardware they purchased….There are subtle differences which mean that the way they installed the operating system's different and it's all built up complexity. And you can't test all eventualities until it's actually deployed....You don't always know why something's failing." (i30) "There are all sorts of weird problems…the way that different components will interoperate" (i49)
Heating, cooling and networking	"Cooling power and also the network. Many sites have had huge problems with cooling, we have had unseasonably hot spells and things…as the resources that are needed grow, the cooling problem grows, the power problem grows, and then it looks like we might need to have a different solution. Regarding the network; we thought of this as being something that is fine, managed, reliable, never breaks. But you discover it does." (i27)
Poorly expressed requirements from the users	"The requirements are not clear, because they can't be….This is a scale we have never really been at before. I think it is physically impossible for the experiments to sit down and write a requirements document….I think probably the underlying problem is it is so distributed." (i60)
Lack of expertise in management of data centers	"There are a number of sites that are in the system now that don't have the experience of running you know, large scale time critical operations where you have to be called in you know, available to fix problems in the middle of the night because if you don't, you know, the accelerator stops working." (i49)
Challenge of testing without access to the experiments data and software	"We don't have access to the experiment software directories for example. So we can't run jobs as a user within that [experiment] will access their data. So if the new Middleware doesn't somehow interface correctly with something, we won't see that until a real user comes along and tests it." (i30)
Fire in a data center	The WLCG data center in Taipei suffered a fire and was out of action for two months: "It has been a disaster right now. the whole data center area are affected while it's the damage of the UPS battery cause the entire power system down and dust and smoke spread into other computer room in which all computing and storage facilities resided. Minor water leak have been observed while fire fighter trying to suppressing the fire in power room. We leave DC an hour ago, right now, the situation in data center are not acceptable for human to stay long." (email from data-center manager at the time)

Table E2. Tuning the Grid through Harnessing CRAB in the Present		
Theoretical Concept	**In practice this is means…**	**Evidence from the Field**
Harnessing of material agency to tune the mangle in a dialectic of resistance and accommodation	Physicists harnessed CRAB, and through it the API of the WMS, to tune the allocation of jobs within the GRID to reflect their intentions.	"So what happens in reality. And this is something we clearly see with the coming of real data [from CMS and the LHC]. The user needs to go to a conference and so needs to do some activity like producing results, which should go in a paper, and publication and so on. So they need to access some data and they know, well they don't know but the CRAB points them to a couple of sites, say three sites. At some point the user sees in real life, that a bunch of jobs continuously fails in a site. From the user point of view they don't care if the jobs are failing due to the fact that the site is failing due to [being] misconfigured, or for other kind of reasons like overload of the site, or problem with data-storage which is broken. So jobs are failing. So the user wants a way to say that my jobs should not go to that site since [the site] is preventing me to produce the results I need to do. If I can black list this site I know I can produce my results quicker." (i12) "I quite often try and send the job somewhere else, not use that particular grid site." (i4). "[The grid] is mostly annoying because [particle physicists] are used to transparency, and the transparency goes away." (i70) "I use a tool called CRAB, which is a CMS tool which allows you to sort of specify places. You can specify places for it not to go to or places that it should go to. So I can try sending it somewhere else…you find out where the data is and then you tell the thing to send there." (i6) "The current grid only provide the basic infrastructure there, so from the physicists, from the experiment, we have to develop a lot of things ourselves on top of the middle ware, what we call production systems, that deals with all the special requirements from us." (i66)
	Tuning the WLCG by physicists creates resistance for the computer specialists	"We are trying to find the best solution to each particular problem domain, integrate it into our release so everyone can use it. Now if there was a user sitting in isolation they would probably have to use what we provide, they don't really have much choice. But these users don't really exist so much, they all work for experiments with lots of influence and resources and everything….So they can bypass stuff. Bypassing is probably a pejorative phrase, it is just they choose to use an alternative route.… But certainly one of [computer specialists'] big services is workload management. So the idea is that [WMS] takes all your jobs and manages them for you, submits them to the right place, so you send them there and forget about them until you all come back. But on [your experiments interface such as CRAB] you can implement most of this stuff, if you want to, to your own satisfaction. And we find people have done that." (i40) "[Experiments] could start trying to pull out some of these bits for their own use [remove data centers from the grid]. That would break the whole idea of having a grid computing environment which other people can share when it's not being used. But there is a risk because they're main goal is to have something which allows them to do physics. If they feel that's not being provided, then they will find other ways of doing it." (i30) "First of all there are a lot more of them, a lot more of the particle physicist users, potential users, than there are of us. So if they have a problem they can go and solve it themselves, rather than coming to us and saying 'why don't you add this into your code for us?' There are just so many of them. If they don't like what we have done they can go away and write something else." (i47)

Tensions of Developing and Sustaining Heterogeneity of Grid Infrastructures Which Orientates its Future

Table E3. Analysis of Human Intentionality Within the Case

Theoretical Concept	In practice this means	Evidence from the Field
Intentionality—presently none existent future states humans seek to bring about.	CMS physicists' intentions in using the WLCG are focused on the experiment and discovering new physics. Computing is a resource in achieving this.	"We have one goal, which is to deliver the CMS experiment and win a Nobel Prize, that's the goal. And we are all working towards that." (i31) "They are driven by one fundamental thing. They want their experiment to work when the beam gets into the accelerator, okay? And that transcends everything else they do." (i42) "My primary interest is physics, and I really see the grid as a tool to let me do physics." (i6) "My real interest is physics....I am not actually that interested in computers, per se, for me they are a tool. I enjoy programming up to a point, but for me it is a means to an end." (i7) "They work because of their passion to do science. So they, we strive to deliver the best result." (i69) "[CERN] is vocational…this is the Mecca…this is why we get up in the morning." (i27)
	A subset hold similar intentions but see the writing of software as a way to achieve this intention	"We discovered from [WLCG] there were some problems with the…middleware. Some requirements were not completely [aligned] with the experiments." (i12) "So the experiments, rather than take the risk of not having things in place, have taken it upon themselves to develop their own applications which do more than WLCG envisioned or intended…they have big development groups working on grid-based technology simply because the functionality's not there yet." (i30) "We don't want it that all physicists need to know all the details of the grid and also about the CMS infrastructure." (i13) "[Software] written by CMS to solve their problems." (i31) "[CMS] got so fed up with this, they went to write their own middle section, the interface between the middleware and the, and the application." (i2)
	The computer specialists intentions are broadly focused on production of a shared grid	"I hoped we would set up a facility to enable LHC analysis that would be able to be operated by computer professionals without further intervention by academics." (i65) "My overall intention for [WLCG] was to solve the LHC computing problem. That is, to contribute to the creation of the world-wide LHC computing grid so that we, wearing another hat—the LHC experiment collaborations, could handle the data that we expected and, ultimately discover or rule out the Higgs." (i3) "Each of the four [LHC] experiments could have somehow cooked up their own solutions and we would have some four things in parallel. [The grid] has brought us further and faster than we would have been."(i40) "In the vision the end user ought to be able to sit behind his terminal and say—run this job—and it comes back some time later. And behind his screen he can't tell whether it's a mainframe or a bunch of PCs off in Iceland or something." (i60) "[WLCG's (UK)] vision is to create a computing infrastructure for UK particle physicists and originally proposed to support the LHC experiments but now supports other particle physics experiments, other physicists, researchers from numerous other disciplines and even small to medium enterprises." (d7)

Table E4. Analysis of Future Expectations of Material Trajectories on Present Agency

Theoretical Concept	Anticipated Future Influence	Evidence from the Field
Chordal composition of projected future materiality	The projected future of the LHC detector influences their actions in the present.	"We know there is an accelerator being built, it is going to turn on, it has got a beam coming through it at a certain time, and you can't miss it, you have got to be there. And that really focuses people's minds, they realize they have to work together and get the thing going." (i67) "A hard deadline is when the LHC is switched on. By that time the computing infrastructure has to be ready, so that is why there was just this big push." (i59) "CERN is put big effort and big money of course, into developing the computing power…in developing and pushing the grid technology because…we believe that… you can't do the science if you don't have the computing power." (i69) "[The LHC's anticipated the] first year will likely be characterized by a poorly understood detector, unpredictable machine performance, possibly inadequate computing infrastructure but also with the potential for significant physics discoveries. We…must be able to make that data…available to the collaboration so that their expertise can be brought to bear on detector, software, calibration and physics as effectively as possible." (d16) "[CERN is] where all this clever technology was invented, you know, real sort of original developments in computing to meet the challenges… that's the way it progresses…this is tricky now. By 2014, 2015, this will not be challenging anymore, okay? So the hump year, the year you've got to get right [is the launch year]." (i2)
	The projected future of ICT technology influences their actions in the present	"After the initial provisioning of the computing systems for the start of data taking, computing resources will continue to ramp up to support running at increasing luminosities, rising to full nominal LHC luminosity. It is expected that through Moore's law, by exchanging out-of-date computing system components at a roughly 3-year cycle, the required performance and resource increases can be obtained at a roughly constant yearly budget." (d16) "Moore's law, which only stated that the transistor budget grows from one generation to the next, will continue to be true, but both the problems with basic physics and the longer verification time needed by more and more complex designs may start to delay the introductions of new process technology." (d36) "There will still be a large amount of local tape-based storage in the LHC era." (d36) "Any functioning grid in the future will have to be heterogeneous, okay, and that's what we don't have in particle physics…the [WLCG infrastructure] takes a cross section through and…it's actually doing a very specific solution now, okay?…If you never [focus on the heterogeneous grid] and you only do [a physics grid], you run the risk of never ever building a useful system that will ever work." (i42) "As far as [high energy physics] is concerned, we can expect the current ten-gigabit technology to satisfy our needs for many years to come." (d36) "I'll use the word bespoke really. You know, where some of it is common and can be but some of it's just got to be done in a bespoke way." (i42) "They were only getting something like one megabyte per second, per job, which is a complete nonsense because we'll easily get, I know, 10, 15, 20 megabytes per second." (i8) "The expected rate of about 150Hz on disk, which CMS will reach at the start of the activity on 2007, implies that few PBytes of data per year will be stored and processed while the detector is collecting data...The most promising solution to cover all these task seems to be the grid paradigm." (d19)

Inertia of Different Installed Bases, Disciplinary Agencies and Conventions of Practice that Orientate to the Past

The following tables provide evidence of the inertia of different installed bases, of the disciplinary agencies and of the conversions of practice that orientates to the past.

Table E5. Examples of the Installed Base of Particle Physics Technologies Upon Which the WLCG Infrastructure Is Founded

Theoretical Concept	In practice this means…	Evidence from the Field
The installed material base imposing path dependency	The range of particle physics IT standards, technologies and material arrangements which orientates the WLCG's material form.	The LCG architecture will [run] "on grid infrastructures provided by the LCG partners. These infrastructures at the present consist of those provided by the Enabling Grids for E-scienceE (EGEE) project in Europe, the Open Science Grid (OSG) project in the U.S.A. and the Nordic Data Grid Facility in the Nordic countries." (d36, p. 29) "No one has done this before, if we did that again, a lot of the lessons we've learnt over the last three years, about how you set up various services, the configuration, the redundancy, we didn't know then, we didn't know how experiments would use the services, so you were more or less guessing what the configuration could be." (i27) "Existing SRM implementations currently deployed include CASTOR-SRM, dCache-SRM" (d36, p. 30) "We have developed the CASTOR Mass Storage System at CERN and by the middle of 2005 the system contained about 35 million files with an associated 4 PB of disk space. The system uses files and file systems as the basic operation unit. The new improved and rewritten CASTOR software is in its final phase and will be deployed during the second half of 2005....The problem of large numbers of small files in the system can only partly be addressed by the new CASTOR implementation, as the major obstacles are not CASTOR-specific but rather arise from limitations in the tape technology. The CASTOR MSS software is the CERN choice for the foreseeable future." (d36, p. 76) "The standard language for physics applications software in all four LHC experiments is C++. LCG software should serve C++ environments well." (d36, p. 90). "[LCG] runs over the standard academic network [with some] dedicated high speed connections." (Pearce and Venters 2012) "The grid is built on the same Internet infrastructure as the web, but uses different tools. Middleware is one of these tools." (d7) "Steve's Jobs"/tests orient present and future action: "There is a new set of tests targeted at UK Resource Brokers. These send (non ATLAS) 'Hello World' jobs to each UK RB every 10 minutes to execute on any UK CE." (d26) "Providing diagnostic information which can help the systems administrators debug their site." (d7)

Table E5. Examples of the Installed Base of Particle Physics Technologies Upon Which the WLCG Infrastructure Is Founded

Theoretical Concept	In practice this means…	Evidence from the Field
	The range of material artefacts impacting upon the grid's design	"The current CERN network is based on standard Ethernet technology, where 24-port fast Ethernet switches and 12-port gigabit Ethernet switches are connected to multigigabit port backbone routers (3Com and Enterasys). The new network system needed for 2008 will improve the two involved layers by a factor 10 and the implementation of this will start in the middle of 2005. Later the performance in latency and throughput can be further improved by using Infiniband products" (d36, p. 80) "Data coming from the experiment data acquisition systems is written to tape in the CERN Tier-0 facility." (d36, p. 29) Connections with tier 1 sites use "a detailed architecture based on permanent 10-gigabit light paths. These permanent light paths form an Optical Private Network (OPN) for the LHC Grid." (d36, p. 82) "General purpose connectivity between Tier-2s and Tier-1s will be comprised of a complex set of research initiatives world-wide that, as with the general Internet, will provide global connectivity permitting Tier-2–Tier-2 and Tier-2–Tier-1 communications to take place." (d36, p. 86) "…was deployed as an interface to the CASTOR storage systems at CERN and UAB Barcelona, to the Rutherford Appleton Laboratory (RAL) ATLAS storage facility, and to IN2P3's HPSS system." (d24) (On the hierarchical organization of the grid) "CMS was actively involved in the MONARC project, whose work led to the concept of a Regional Center hierarchy as the best candidate for a cost-effective and efficient means of facilitating access to the data and processing resources." (d37)

Table E6. Disciplinary Agency and Conventions of Practice of Particle Physicists

Theoretical Concept	In practice this means…	Evidence from the Field
A disciplinary agency of experimentation and individual fixing of problems	That problems with computing should be resolved by the individual, calling upon resources from other physicists.	"The closer it gets to switch-on, the more they'll do quicker ad hoc jobs to get it done." (i42) "The way particle physicists works is it's very good at getting stuff done." (i39) "They are, essentially, very dirty programmers…they really will use the fastest way to get at something." (i4) "In physics, you just develop something, you make sure you get it running." (i59) "So every physicist is a programmer." (i15) "it is all very home-made, home brewed." (i68) "Most particle physicists want to roll their sleeves up and get involved in all the technical detail of how it's done. And it's always been that way. You rarely would get somebody going to do their PhD in a particle physics group who said: oh I don't want to know about computing. I just want to do the analysis. It just doesn't happen.… It's always in the mentality." (i42) Also see articles (Galison 1997; Knorr-Cetina 1999; Pickering 1984; Traweek 1988) which detail the disciplinary practices.
A disciplinary agency of writing software as central to the practice of doing physics	Physicists program as part of their work, and are skilled at writing software.	"By and large, whether it's good or bad, we're all pretty technical when it comes to programming.…You might think physicists should be operating at much higher level of dealing with pictures and boxes for analyzing data. But by and large they don't. They just write C++ and that's the end of the matter. We like doing it." (i42) "We are experimentalists, so we need to find ways to write code to analyze the data. One could easily say, well, okay, why don't you give the code to software people, to write proper code? But you need to put your physics intuition in the code so we may not write a perfect code that software engineer would write but on the other hand, we write codes that are mainly directed by physics intuition." (i30) "You can do purely physics just sitting in front of your computer and programing. You have the data and then you program to analyze the data. So programming is major." (i85) "[A] lot of people who go into high energy physics at least end up being pretty good in information technology and some form of coding because it's so heavy IT-based. They'll develop various applications." (i30) "I know that grid people complain about a lot is there are all these physicists turned computer scientists who are essentially improvising a bit, and I can't judge whether that is right or wrong, but I do suspect that computer scientists would probably solve things completely differently than a physicist would." (i4) "It's clear that everybody does computing you know, you go and find a particle physicists, and he's sitting in front of screen somewhere doing some computing." (i65) "In order to complete, to complete his or her PhD, he should know, he should learn C++. Otherwise, he would find it very difficult to follow." (i69)

Table E7. Disciplinary Agency and Conventions of Practice of Computer Specialists		
Theoretical Concept	**In practice this means…**	**Evidence from the Field**
A disciplinary agency of computer science with elegance and sophistication	Focused on building a coherent scalable grid with elegance and sophistication. Long history of developing and using computers.	"Developers want to deliver something that is perfect." (i60) "We have a complete specification. The specification changes slightly as we go along, as does the design. But we still have an overall plan." (i47) "Computer scientists are reengineering this type of thing, from scratch…what you really want is a combination, you want computer scientists working within the science community…you want software engineers…People who actually get experience in big products and stuff." (i5) "I think the whole challenge really is to make everything resilient, efficient and in the same way that we've done computing in the past, get to that sort of level of efficiency." (i58) "There's always been a long history of computing …at the forefront of computing... there's always been a strong computing, we've always had lots of computer experts cause we've always had system managers and people." (i65) "[For a long time,] there's been a body of people that have been looking at parallel computing/performance computing, distributed computing and so on you know, over time, looking at clustering, looking at having sites working together." (i49) "We've done distributed computing for many years…we've worked out a way of doing it and we're using grid to actually make that easier and improve it and extend it." (i54) "A lot of emphasis was put into designing architectures, in which you separate the concerns very much…of what the physicists will have to write and [what the] computer scientist has to write." (i15)
A disciplinary agency of running efficient systems	Computer specialists are responsible for developing and running the grid efficiently.	"[We] will have some unit tests…we also have a set of system tests as well....Once we have actually got a build based on a tagged set of components then we go through tests, the testing time. Once we have actually got something that we are happy with…we have a consistent build …is then submitted." (i17) "The biggest thing is making it failsafe. That it'll operate 24 hours a clock without anybody having to be called. There's a lot of discipline in that when you're dealing with more than a couple of boxes. We're dealing with, today, three or four thousand and it will double in the next two or three years." (i49) "We use methodologies but just up to the limit that its appropriate for what we're trying to do…very often the spec is a know *a priori*." (i42) "We have the basic principles of software engineering that [you] go from having an automated system, to a test infrastructure, to using any tracker type of tool, for bug reporting, task management, etc. So this is mandatory." (i70) "There's this army of people in the background keeping it going, hopefully diminishing in numbers but at the hardware level, you're going to obviously have people dealing with failures and you're going to have to have the usual support help desks and the people responsible for fixing the middleware bugs." (i64)

References

Galison, P. 1997. *Image and Logic*, Chicago: University of Chicago Press.
Knorr-Cetina, K. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*, Cambridge, MA: Harvard University Press.

Pearce, S. E., and Venters, W. 2012. "How Particle Physicists Constructed the World's Largest Grid: A Case Study in Participatory Cultures," in *The Routledge Handbook of Participatory Cultures,* A. Delwiche and J. Henderson (eds.), London: Routledge.
Pickering, A. 1984. *Constructing Quarks: A Sociological History of Particle Physics*, Chicago: University of Chicago Press.
Traweek, S. 1988. *Beamtimes and Lifetimes: The World of High Energy Physics*, Cambridge, MA: Harvard University Press.

Appendix F

This appendix includes various evidence samples.

Item F1. A sample of CMS's online Frequently Asked Questions (FAQs) (d2) detailing CRAB.cfg options (truncated). Note CE stands for Computing Element—a computer-processor on the grid. SE stands for Storage Element—a data-storage device (e.g., Hard disk/Tape Drive) on the grid

```
If you want/need to select/deselect some site, you can use: (see Crab FAQ for
more info)
□  Ce_black_list – (refuse access to all the listed CEs, allow all others)
□  Ce_white_list – (allow access only to those CEs listed)
□  Se_black_list – (remove the selected SE from the list of sites hosting
data)
□  Se_white_list – (select only the SEs listed)
-----------------------------------------------------------------------------
## CE Black List: all the CE whose name contains the following strings (comma
## separated list) will not be considered for submission.
-----------------------------------------------------------------------------
So, in summary, if you want to force your jobs to go a specific site (eg if
you want to test the site), use "SE_w/b_list".  If instead you want to access
some dataset but you want to avoid a site (because you don't trust it), use
"CE_w/b_list".  In addition, se_w/b_list cannot be used with None as input
dataset.
```

Item F2. A Wiki page (d3) showing other CMS physicists how to simulate a CMS Higgs-Boson discovery using simulated data. Within this script it directs the user to se_white_list.

```
Create the CRAB configuration file: Demo/MyTrackAnalyser/test/crab.cfg and
give it the following contents, replacing the items in brackets <…>
appropriately:
…
[EDG]
se_white_list   = <location found with dataset discover page (e.g.,
srm.unl.edu)>
…
```

Appendix G

The following table provides a glossary of the technical terms and acronyms used within the paper.

Table G1. Glossary of Technical Terms and Acronyms

Term	Definition
Computer Cluster	A large numbers of usually homogenous co-located servers running the same software to provide high performance computation. Also known as server-farms. These cluster computers ranges from few networked personal computers to supercomputers such as Titan, a cluster with 560,640 processor (www.top500.org/).
CMS	A collaboration of physicists who built, and run the Compact Muon Solenoid detector, one of the four experiments on the LHC.
Computing Job	A "job" is the running of a software application on the grid; it is a block of computation run on the grid. Usually a job is broken into large numbers of separate jobs which run on separate processors of the grid.
CRAB	CMS Remote Analysis Builder: A grid job submission system written by members of the CMS collaboration.
GGUS	Global Grid User Support: A global messaging system which allows users of the grid to gain support from the global network of maintainers without knowing their location/details.
Grid Computer	A distributed infrastructure connecting of large numbers of sites containing computers and storage devices running grid middleware.
LHC	Large Hadron Collider: A machine to accelerate particles (hadrons) to close to the speed of light and then collide them within the four experiments distributed around its 27 km ring.
Middleware	The software run on the grid's computers which organizes and provides the utility computing service. The term middle refers to its place between the operating system and the application.
Monte Carlo	An analysis technique.
Site	A data center housing large numbers of computers and disks which contribute to the grid.
WMS	The Workload Management Service is the part of the grid middleware which organizes the allocation of computing jobs to the computers of the grid.