

This paper was presented at the conference:

Oxford eResearch 2008

11-13 September 2008

University of Oxford

Conference website:

<http://www.oii.ox.ac.uk/microsites/eresearch08/index.cfm>

Conference papers collection:

<http://ora.ouls.ox.ac.uk/objects/uuid%3A64aa6f39-7e81-4d42-a008-ee2d7524bd67>

Conference organisers:

Oxford Internet Institute

University of Oxford

1 St Giles, Oxford OX1 3JS

<http://www.oii.ox.ac.uk/>

email: enquiries@oii.ox.ac.uk

Oxford e-Research Centre

University of Oxford

7 Keble Road, Oxford OX1 3QG

<http://www.oerc.ox.ac.uk/>



A CLIMATE OF TRUST: Perceptions of legitimate authority in a volunteer computing project

Peter Darch

Annamaria Carusi

Marina Jirotko

Oxford e-Research Centre, University of Oxford

September 2008

Prepared for:

The Oxford e-Research Conference 2008

ABSTRACT

Volunteer computing projects are computationally-intensive scientific research projects which seek to meet their need for data processing capacity by asking members of the public to download and process project data on their personal computers. This paper presents findings from a study of the interactions in an online forum of participants in such a project, climateprediction.net. It focuses upon the beliefs of one group of participants, who have been identified as particularly critical to the success of volunteer computing projects, about the organizational policies of climateprediction.net and what are appropriate and inappropriate ways for other participants to behave in the forums.

INTRODUCTION

In recent decades, there has been an explosion of scientific subdisciplines and projects which involve the processing of large quantities of data or the running of computationally-intensive mathematical models (Galison & Hevly 1992; Welsh *et al.* 2006). To meet their computing needs, and seeking to avoid the high costs associated with supercomputers (Yao 2006), some scientists have set up scientific projects which ask members of the public to donate spare capacity from their personal computers. These projects have been referred to variously, and interchangeably, as 'volunteer computing' (Christensen *et al.* 2005, p. 8), 'public-resource computing' (Stainforth 2002a, p. 38), and 'global computing' (Anderson 2004, p. 4). The capacity potentially available to scientists is vast. In June 2008, the research firm Gartner estimated that there are now over one billion personal computers in use around the world (Winter & Pettey 2008). This, combined with rapid increases in the power of such computers, now means that the collective capacity of such machines now far exceeds that available in scientific institutions and supercomputing centres (Anderson 2003).

Since the first volunteer computing project was set up in 1996, scientists have recruited many members of the public as volunteers. The most successful in this respect is SETI@Home (part of the Search for Extraterrestrial Intelligence), which was launched in 1999 (Korpela *et al.* 2001). As of the end of August 2008, SETI@Home can count on almost 200,000 active volunteers, whose computers collectively contribute an approximate average of 450 TeraFLOPS, or 450 trillion floating point operations, per second to SETI@Home [1]. This rate of data

processing is approximately equal to the world's third most powerful supercomputer [2].

The team behind SETI@Home developed a piece of middleware called the Berkeley Open Infrastructure for Network Computing (BOINC). BOINC is a piece of middleware, which provides a framework for scientists who wish to set up a volunteer computing project and also provides an interface which volunteers download in order to communicate with the project (Anderson 2004). As of August 2008, there are fourteen projects using BOINC, in fields such as biochemistry, physics, mathematics, and climate science. Additionally, there are 55 projects using BOINC which are in the Alpha or Beta phase of development [3]. Collectively, there are over 320,000 active users of BOINC, a scale which massively outstrips that of any other middleware used in volunteer computing projects [4].

This paper will set out some preliminary findings from an ongoing qualitative study of one such project, namely *climateprediction.net*, which was set up by a group of atmospheric physicists at the University of Oxford to model climate change (Stainforth *et al.* 2002b). A particular focus is on perceptions of project volunteers regarding the way in which the project organizes its participants and in which ways someone can act with legitimate authority, in the sense of whether they would approve of an action performed by this individual. Legitimate authority is closely bound up with the notion of trust: trust in somebody's ability to perform the act competently, and trust that they will not perform the act with malevolent intent.

A motivation for this study comes from issues which may arise in future developments of volunteer computing projects, especially *climateprediction.net*

There is a recognition of substantial scope for improving the efficiency of BOINC-based projects (Anderson & Fedak 2006). Of particular interest are proposals that such projects should move from their current setup, where volunteers' computers communicate only with a project's central server (a 'client/server model' (Pourebrahimi *et al.* 2005, p. 1)), to a more 'Peer-To-Peer' (P2P)-based network where volunteers' computers could communicate with each other (Costa *et al.* 2008). It will be argued that features unique to climateprediction.net make a move to a P2P network seem especially promising for this project, but to move successfully would require understanding attitudes of volunteers regarding the project's organization and assignments of roles to various volunteers. Although all participants are of value to volunteer computing projects, some play a particularly critical role in the projects' success, and this paper will focus on one group of such volunteers. First, however, it is necessary to give a more detailed presentation of the main features of BOINC and climateprediction.net.

BOINC AND climateprediction.net

The climateprediction.net project was set up, and is run by, a team of atmospheric physicists and computer software specialists at the University of Oxford. Its aim is to quantify the uncertainty involved in predicting future climate change. This uncertainty arises from difficulties in obtaining reliable estimates for many of the parameters involved in mathematical models of climate. The original intention was to run between one and two million versions of a climate model, which simulated future climate, each using different estimates for the model's parameters, to assess the level of variation generated by the model in predictions of future climate change

(Stainforth et al 2002a). The project began in 1999, and in 2002, the decision was taken to transform it into a volunteer computing project. Following Alpha and Beta testing phases in late 2002 and spring 2003, climateprediction.net was launched to the public in September 2003. In August 2004, the project moved to using BOINC, and continues to use this middleware [5]. As of August 2008, there are approximately 32,000 active users [6].

Currently, climateprediction.net offers three climate models, which have been developed by the UK Meteorological Office. One is called the HADCM3, which models atmosphere and ocean temperature from 1920 to 2080. It is expected that this would take four months to run on a volunteer's computer. It saves its progress on the computer after every six model days (known as a 'checkpoint') and after every model decade, simulation results are uploaded to the central project server, along with a 'restart dump', which is a file that would enable an aborted or crashed model to be completed on another computer. Another model, HADSM3, runs for 45 model years and does not involve such a complex model for ocean temperature. It is expected to take 15-30 days for completion, and checkpoints occur after every three model days. The third model, HadAM3, is to season variations in climate, and consists of one model year with checkpoints occurring after each model day. One climateprediction.net model is known as a 'work unit'. A typical work unit for climateprediction.net is substantially longer than for almost any other BOINC-based project. For instance, a work unit for SETI@home or Einstein@home 'can typically be completed in a matter of hours' (Christensen *et al.* 2005, p. 9). This will be an important point to bear in mind in the next section, where some of the challenges faced by climateprediction.net are discussed.

In order to run these climateprediction.net models on their computer(s), project volunteers need to download and install a copy of the BOINC interface. The interface allows the volunteer the ability to set certain preferences, including the proportion of CPU time they wish to allocate to each BOINC project. With respect to climateprediction.net, the volunteer is able to choose which type of model they wish to run.

If a climateprediction.net volunteer has difficulty with installing or running BOINC, or encounter difficulties with their climateprediction.net model, they can seek advice on one of the discussion forums found on the project's website. These provide the main arena for project participants to interact, and for those involved with the running and administration of climateprediction.net to communicate with the volunteers. Some are labelled to encourage participants to share problems they have, and offer opportunities for participants to offer encouragement or technical advice to others; others have been set up with the intention of participants sharing and discussing results from their models, or to discuss general issues relating to climate science. The responsibility for maintaining order lies with a group of project volunteers, who are known as 'moderators'.

Another key feature of the climateprediction.net project is that it, in common with other BOINC projects, records statistics about the quantity of model data processed by project volunteers and their computers. Two key indicators are total credit (which measures the total quantity of data processed on an individual's computer) and the total number of models completed. In climateprediction.net, credits are awarded for

each model-year completed on a volunteer's computer. These are very often a source of pride for project volunteers, who will sometimes report their statistics in a signature placed at the end of forum posts. Furthermore, volunteers can group together and form teams, with one volunteer being designated 'team founder' or 'team leader' and very frequently, these teams have websites and forums of their own. On the climateprediction.net website, there are lists ranking both individual participants and teams according to credits acquired, or Recent Average Credit (the statistics for teams are simply aggregates of the individual statistics).

PEER-TO-PEER NETWORKS: POSSIBILITIES AND SHORTCOMINGS

The use of software such as BOINC has made it substantially easier for many scientists to access spare computing capacity; nevertheless, many shortcomings have been identified. Some of these shortcomings have been attributed to the server/client setup of BOINC-based projects, particularly relating to risks involved in the network being completely dependent on the server, and costs incurred by the BOINC project. Bottlenecking at the server is a particular issue, necessitating the purchase of extra bandwidth by the project (Costa *et al.* 2008), and vast quantities of memory need to be bought by the project to store the status of already-issued work units, and any restart dumps which have been uploaded. This latter need is particularly acute for climateprediction.net. This project's very lengthy work units means that research dumps are substantially larger than for other BOINC projects, and, furthermore, means that climateprediction.net suffers a 'large attrition rate' relative to other BOINC projects, making research dumps particularly important if data is not to be lost when volunteers abort partially-completed models (Christensen

et al. 2005, p. 12). Additionally, the BOINC server/client network has a single failure point: the failure of the central server will have implications for all volunteers' computers requesting new tasks, or could lead to a catastrophic data loss.

Some form of P2P networks have been suggested as offering improvements, giving project scientists access to spare bandwidth between volunteers' computers to transmit data (Pan 2008; Pourebrahimi *et al.* 2005), or excess capacity on their computers to store data. Furthermore, risks of failure would be distributed throughout the network: the failure of a node in a P2P network would be likely only to affect a small part of the network (Costa *et al.* 2007), and the chance of a single catastrophic data-loss would be eliminated (Kim *et al.* 2006).

Common to papers arguing for the superiority of P2P networks to client/server networks in volunteer computing projects is their heavy reliance upon the results of simulations of hypothetical P2P and client/server networks to support their conclusions (Costa *et al.* 2008; Krishnan *et al.* 2003). These simulations rest on the assumption that a change from a client/server to a more P2P-based network would be unproblematic, in particular that there would be the same number of individuals willing to participate before and after the change. However, it is very plausible that this would not be the case in reality, as P2P networks require different patterns of trust on the part of participants (Bursell 2005; Wallach 2002), including confidence in the security of their computers' multiple connections with other, largely-unknown, potentially malicious, volunteers. There are therefore serious grounds for suspecting that moving to a P2P setup could result in a reduced uptake amongst the general public.

This is usually treated simply as a challenge of software engineering, i.e. how to produce software that will protect volunteers from abuse (Pan 2008), but this may not be enough to secure trust of members of the public. Another, more promising, way to secure volunteers' confidence could be to restrict the ability to propagate data to a limited group of peers who are trusted, and endorsed by, the central project authorities ('Super-Peers'), so that volunteers' computers will communicate only with these Super-Peers (Costa *et al.* 2007).

In addition to understanding the nature and extent of the trust of members of the public in a Super-Peer system, considering such a system raises new issues which might impact on its success. Some of these relate to perceptions of the legitimacy of those who might be designated as Super-Peers. For instance, who might volunteers consider as possessing the appropriate attributes to become a Super-Peer? Would an individual feel aggrieved if they were not designated as a Super-Peer? Do people have strong opinions regarding how the roles of various individuals in the project should be positioned in relation to one another? Considerations of such issues should form an important part of any decisions regarding the introduction of elements of a P2P network into a project such as climateprediction.net, as perceptions regarding whether the project is well-structured will influence volunteers beliefs about whether continuing in the project is worthwhile. It is hoped that the results presented below will at least begin to contribute to such an understanding.

THE STUDY

Methodology

The approach in this study follows that labelled ‘virtual ethnography’. Rather than seeing ‘online’ phenomena as following inevitably from the features of the technology, they should instead be understood as artefacts of highly context-dependent, situated practices (Dominguez *et al.* 2007; Hine 2000; Markham 1998).

A corpus of data was assembled for analysis for this project, consisting of: semi-structured face-to-face interviews with those involved in developing and maintaining the climateprediction.net software; a qualitative questionnaire aimed at climateprediction.net volunteers consisting of a number of open-ended questions, with participants (thirty, so far) recruited by threads posted on the climateprediction.net forums by two moderators; articles published by the developers of climateprediction.net; and threads from climateprediction.net’s online forums. The questions for the interviews and questionnaire sought to encourage respondents to talk about experiences in the climateprediction.net project and, in particular, perceptions of, and interactions with, others involved in the project. Threads were selected from the forums according to how promising they seemed in terms of revealing participants’ attitudes towards others in the project. When quoting from forum threads, partial anonymity will be employed, in which those quoted will be assigned pseudonyms different to those they use in the forums, although it will be indicated if they are a forum moderator or team leader.

Approaches from discourse analysis were used when handling the data (Gill 2000), treating it not as accounts of underlying feelings and perceptions of the respondents but, in line with the approach of virtual ethnography, as contextually-produced artefacts of particular sites of social interaction (Abell & Myers 2008). Various systems of coding were developed, based on initial readings of the texts, threads, and transcripts, and influenced by the notion that individuals' accounts actively construct the roles and identities of themselves and others (Potter & Wetherell 1994; Taylor 2001). Additionally, a number of attempts were made to group participants based on various pieces of information available on the participants' profile pages on the climateprediction.net website, such as their Recent Average Credits, team membership, and the other BOINC-based projects in which they are involved.

Some Results

This section will focus on a particular group of project volunteers, namely those who are involved with a very large number of BOINC projects. For the sake of convenience, these will be collectively referred to as HNPs (volunteers in a 'High Number of Projects'). The vast majority of climateprediction.net volunteers participate in up to nine other BOINC projects, and these projects tend to be well-established, having already passed through Alpha and Beta testing phases. By contrast, the HNPs are involved with 15, 20 or even more than 30 BOINC projects, many of which are in an Alpha or Beta testing phase.

Such individuals are very valuable to BOINC projects in a number of ways. To projects in the Alpha or Beta testing phases, they provide feedback about

experiences of using the project software. In addition, for such relatively new projects, they provide an early, stable base of volunteers, which can be very valuable in terms of establishing the project, for instance if these volunteers promote the project on their teams' forums. To more established projects, such volunteers can prove valuable for testing new features or extensions of the project, for instance when climateprediction.net transferred to BOINC itself. As these volunteers play such a critical, and unique, role in various BOINC projects, studying their interactions is particularly important with respect to some of the questions motivated previously in this paper.

This section will present data from one thread, which will be referred to as 'Thread F', in the online climateprediction.net forums, occasionally bringing posts in other forum threads to clarify or develop points. Thread F was chosen as a focus primarily because it became a site for a dispute in which some HNPs played a prominent role in which they revealed a number of their attitudes about the roles which others could legitimately play in the climateprediction.net project.

This thread originated as a request for help from a project volunteer ('Rob') whose HADCM3 model persistently crashed at a relatively late stage in its running. A number of volunteers offer advice on how the volunteer should proceed. A consensus is quickly reached that the model was unsalvageable, and that the volunteer should move on to another project. Rob expresses disappointment, as it was the first model he had not been able to complete. Consolation is offered by two forum moderators ('Jim' and 'Eileen') that results from the crashed model would still be of value to the scientists behind climateprediction.net. Indeed, Jim states that

figure reported by climateprediction.net on its website as the total number of models completed by volunteers actually includes all models that have reached 2050, and not 2080, in its reported count of the number of models completed by project volunteers. This provokes a response from two HNPs ('Jet' and 'SG'), both of whom say that it was wrong to define such models as completed when teams only count models which had reached 2080 as complete, and that there is a real need for consistency and a lack of ambiguity about the measurement of individual, team and project statistics. It is the conduct of Jet and SG in this thread which shall be dealt with below, looking at how they show their beliefs regarding who does and does not have legitimate authority to behave in particular ways, assign tasks to others, and thereby attempt to construct a social hierarchy.

HNPs Define Their Legitimate Authority

In the forum thread considered here, Jet and SG participate in a particular, and limited, way: although the thread was started ostensibly to solicit advice on an issue related to the day-to-day running of the BOINC on a participant's computer, they refrain totally from offering any such advice (or subsequently consolation) to Rob, and participate only in the discussion about the measurement of statistics, where they strongly advocate the importance of strict and universal rules relating to the measurement of statistics.

First, they act to exclude from the topics they seek to present themselves as having legitimate authority to talk about the day-to-day running of BOINC, and climateprediction.net models, on volunteers' computers, even though they could be

expected to be very familiar with running the software as they have amassed a high number of credits, and are long-standing volunteers. In Thread F, they do not offer any advice. Indeed, members of HNPS seldom offer such advice on the forums, and in the instances they do, they actually present their advice in such a way that reaffirms the notion that they lack the legitimate authority to do so in general. The following post (from another thread) exemplifies their approach:

‘I have 2 Sulphur wu’s [work units] at the moment...In general, if you have 2 wu’s...I’d leave them to run. No need to abort them.’ (Post L/005)

Whenever an HNP offer advice in the climateprediction.net forum, they cite an instance of personal experience and suggest it is very similar to the predicament of the participant to whom they are giving advice by suggesting common elements, thereby justifying their authority to give the particular piece of advice; by contrast, many threads by non-HNPs simply assert advice. For instance, advice given by non-HNP Jim to Rob begins: ‘The model may have reached as far as it is going to go’ – no personal experience cited here – and then offers advice: ‘It is generally worth restoring from a backup’ (Post F/003). Such a lack of justification can be understood as an implicit, general claim to authority relating to the giving of on the part of some non-HNPs; by contrast, the practice of HNPs of always citing a personal experience lays no claim to authority for giving advice beyond the specific instance where they do.

Instead, they seek to cast themselves as possessing legitimate authority to discuss and to seek to influence climateprediction.net’s policies and organization of volunteers. Firstly, they are very vocal in Thread F regarding the measurement of volunteers’ statistics, arguing that consistent, well-communicated rules are critical to

ensure that volunteers' behave in the best interests of the project. For instance, SG writes:

'If climateprediction.net report anything past 2050 as complete, this leaves us singing from different hymn sheets...If the rules of the game are to change, they need to be changed for all teams' (Post F/008)

And Jet:

'Let's get this crystal clear right now and not leave any misunderstandings, now or in the future...it would be dreadful to find people would 'tactically' prematurely abort a model at 2050 just to show how great their team is, or whatever.' (Post F/027)

Jet's use of the imperative in the latter post is suggestive of attempts on his part to define what he considers appropriate behaviour of others, legitimate ways in which others can and should act, and how he, and other HNPs, does so is considered in the next section. However, it must first be noted that in so doing, the HNPs are claiming legitimate authority to define the roles and relative positions of project volunteers and assign tasks to them. In other words, they claim legitimate authority to advocate and enforce their views on the social organization of climateprediction.net.

HNPs Define the Roles of Others

This section will focus particularly on the interactions in Thread F of Jet with the previously-mentioned Eileen, Jim, and Rob, as well as another project volunteer ('Sheila'). Both Eileen and Jim are long-standing forum moderators. Sheila is the

leader of a relatively large team, and posts a particularly high number of messages relating to teams, welcoming new recruits to the team, and congratulating team members on reaching particular thresholds for total credits acquired or the Models Completed count. Rob has no official status apart in the climateprediction.net project apart from that of project volunteer.

First, the case of Rob shall be considered. In addition to writing the initial post of the thread soliciting advice, he goes on to make frequent posts relating to how a completed model should be defined, but it is clear that Jet does not regard these as legitimate contributions to the debate, and this can be seen by comparing Jet's responses to Rob's posts to his responses to Jim's. Both make posts almost simultaneously, with both apparently arguing along similar lines, seeking to resolve the controversy about what counts as a completed model by appealing to the good of science. They both argue that, as models approach 2080 (and pass 2050), they are unlikely to provide accurate predictions of what will actually happen, and hence results for the 2050-2080 time period are of reduced worth to the scientists and it is therefore legitimate for them to consider models which have reached 2050 as 'complete'. Some hours later, Jet posts a response, which is oriented solely towards the post by the moderator, Jim. This can be seen firstly because the post is specified as being in response to that of Jim's and also because Jet directly quotes a phrase from Jim's thread (' "please think about the physics and science of it." ' (Post F/020)). This was even in spite of Rob's post actually being addressed to Jet ('Hi [Jet], my fellow Team England (BOINC) team mate' (Post F/015)).

This is typical of Jet's behaviour towards others in this thread: respond to those posts about the measurement of statistics made only by moderators or Sheila, a team leader, and ignore the many made by others. Every post made by Jet on the issue prior to its apparent resolution (with one key exception, discussed in the next paragraph) either states that it is 'in response to Message ID [number of a post made by a moderator or Sheila]' and contains direct quotes from the posts of moderators, or Sheila, only, creating the strong impression that Jet regards the contributions to the debate of those who don't hold an official post as not having any value.

This can be further understood as also forming part of a strategy on the part of Jet to assign the task to these three individuals of communicating unambiguously and explicitly to the mass of project volunteers regarding which models count as 'complete'. The first part of the strategy involves the construction by Jet of a division between these three individuals, casting them as being able to understand how to best serve the good of science, and the broad mass of project volunteers, who are portrayed as lacking this understanding, instead needing to have their behaviour directed and regulated by the system of attribution of credits and other statistics. For example, when addressing Sheila, Jet states: 'Science is fair and obvious to you', but then goes on: 'it's so obvious that winning on credit is more important than (*sic*) the climate science and...It's quite clear that "completed models" is a top bragging item for many' (Post F/030).

The next part of the strategy involves Jet suggesting that any confusion among the volunteers about when a model can be counted as completed is damaging to the interests of the project, because it might induce volunteers to ' "tactically"

prematurely abort a model at 2050 just to show how great their team is' (Post F/027), and that the fault of such confusion lies with people such as the moderators or team leaders. For instance, Jet says to Eileen: 'The fact is you have given an impression that models at 2050+ are complete. I expect some are confused about this' (Post F/013). By accusing them of being the source of the confusion amongst the volunteers, Jet is also portraying them as already possessing authority to influence and direct the project volunteers. In turn, Jet is claiming authority to influence these three individuals regarding the way in which they control the project volunteers, and attempts to do so to get them to unambiguously make it clear to the project volunteers that only models which get to 2080 should be considered complete.

Briefly, then, the behaviour of various HNPs in Thread F, and elsewhere, suggests that they believe in: the critical importance of consistent and clear measurement of volunteers' statistics (count of models completed, credits etc); strict demarcations of when volunteers can and cannot act with legitimate authority; and, embedded within this, a social hierarchy where HNPs are able to discuss and influence climateprediction.net rules and policies, and designate moderators and team leaders as being responsible, and possessing the authority, to enforce these rules amongst project volunteers.

Theoretical Models

At this stage, theoretical frameworks which might help to explain the data presented in the previous section will be considered. This may be useful for aiding

understanding of the views and perceptions of HNPs and could also prove fruitful for informing future analysis of the actions of climateprediction.net's volunteers.

By claiming legitimacy to discuss and enforce the organization of (part of, at least) the members of the projects, it appears that Jet and SG seek to align themselves with the technoscientists (the atmospheric physicists and computing experts) involved in the project's setting-up and running. Studies of the conduct of scientists and technical experts regarding controversies such as the effects of nuclear power stations, lead to the development of the paradigmatic view in the academic literature that those who claim membership of a technoscientific community will act to draw a visible cultural boundary between themselves and the public-at-large, to claim that they, and they alone, possess the knowledge and the ability, and hence the authority, to participate meaningfully in processes relating to technoscience, and that any contributions by the public-at-large to these processes would be worthless (Irwin & Wynne 1996; Wynne 1996). Indeed, this has been supported by surveys asking scientists about their perceptions of the public (Cook *et al.* 2004; Frewer *et al.* 2003). Some recent work, however, has suggested that scientists believe that the public can contribute meaningful knowledge to technoscientific processes but that, nevertheless, the members of the public cannot legitimately participate in evaluative processes (Young & Matthews 2007).

This seems, initially, to be a promising model for explaining the discursive actions of Jet and SG in Thread F, and HNPs in the climateprediction.net forums in general. They seem intent on drawing a clear boundary between themselves and the project volunteers in general. They appear to acknowledge the value of knowledge offered by these volunteers when it relates to the day-to-day running of BOINC and

climateprediction.net models on volunteers' computers and, indeed, they, themselves, sometimes solicit such advice from project volunteers. However, they also appear to be excluding most project volunteers from the group of those who can legitimately talk about, evaluate, and seek to enforce, the organization and policies of the climateprediction.net project. Indeed, the volunteers are seen as lacking sufficient knowledge to evaluate how they should conduct themselves in the climateprediction.net project (hence the need for a rigid and clearly-understood system of rules, communicated by particular volunteers who hold positions such as moderator or team leader, for measuring volunteers' statistics) to direct their behaviour. However, this model does not appear to be sufficient to explain the behaviour described in the previous section, in the sense that it does not account for why Jet and SG, and other HNPs, appear to present themselves as lacking the authority to offer advice on the day-to-day maintenance of the project on volunteers' computers.

Instead, it will be argued that the model of 'legal authority' proposed by Max Weber, one of 'three pure types of legitimate domination' (Weber 1978, p. 31), offers a better fit. By a 'pure type of legitimate domination', Weber meant the existence of a valid social order, valid in the sense that individuals will orient their behaviour towards the norms of the social order. However, an important point to bear in mind below is that in *Economy and Society*, Max Weber argues that the existence of a legitimate social order does not necessarily imply that all actions will obey the social order. In the case of an act which transgresses the social order, the order is nevertheless recognised as legitimate if, for instance, the transgressor feels compelled, publicly and explicitly, to justify their behaviour.

The main features of a Weberian rational-legal social order are: clearly-defined rules to regulate behaviour, enforced by a hierarchical authority structure, with authority delegated down the structure; the existence of an impersonal order behind these rules which even the individual(s) in authority must obey; clearly-defined spheres of competence for each individual in the social order, who have the authority to perform any action which falls within their sphere, and are trusted to do so competently, and it is considered inappropriate for them to act outside of their sphere; and the existence of a reward system, with reward directly related to the quantity of work done towards the ultimate goals of the social order, as an incentive for individuals, and hence can be seen as foundational for the functioning of the social order (Weber 1946, 1978).

The HNPs studied above appear to support a social order whose organization is founded upon rules relating to how an individual's models are counted as 'completed' and credits are attributed. As measurements of work done, and the kudos gained by volunteers when they display their statistics on the online forums, they can be seen as functioning in the same way as a reward system in a Weberian legal social order, with individuals (and teams) being able to display their statistics to other volunteers in the forums to gain kudos, and hence can be seen as a way in which climateprediction.net can direct the behaviour of its volunteers through giving them incentives to behave in particular ways.

Furthermore, Jet's, and other HNPs' behaviour, in the forums appear to be akin to seeking to define spheres of competence both for themselves and for others, and furthermore, spheres which are organised into a social hierarchy. HNPs regard themselves as legitimately competent to discuss rules regarding how statistics are

measured and, given that they believe these form the basis of attempts to direct and govern the behaviour of project volunteers, it can be seen that the organization and policies of climateprediction.net may be supposed to fall within their spheres of competence; by contrast, giving advice on the day-to-day running of climateprediction.net models and software on volunteers' computers falls outside their spheres of competence. Weber stated that the recognition of the legitimacy of a social order by an individual does not always mean they will obey the social order, but that they will act in a way, such as explicitly offering justification, to acknowledge that they realise their transgression and to legitimate the particular instance when they do so. In a similar vein, although HNPs do occasionally offer such advice, they always relate it to a particular experience.

CONCLUSION

In this paper, some early findings from a study of volunteers in the climateprediction.net project were presented. Moves towards a more Peer-To-Peer-based network (and away from client/server models) have been advocated strongly in the literature as promising substantial improvements to the efficiency of volunteer computing projects in general, but it was argued here that such a move might not be as unproblematic in terms of volunteer-retention as is assumed in this literature. In particular, a P2P network would bring into being new relationships between volunteers, requiring different patterns of trust than in a client/server network. Some possible ways of dealing with this, such as a 'Super-Peer' system, motivate the importance of understanding volunteers' perceptions of other volunteers, including (or especially) beliefs about organizational policies of volunteer computing projects such

as who should or should not assume particular roles or be given the authority to act in particular ways, when deciding future changes to volunteer computing projects.

This paper started to address this. It looked in particular at one group of people, the HNPs, who have played a unique and critical role in BOINC-based volunteer computing projects: their recruitment is important to emerging projects, and their retention is important for existing projects. It was found that they possess strong views that clear, consistent, and well-communicated rules regarding the measurement of project statistics are necessary in order to direct the behaviour of project volunteers.

It was also argued that HNPs appeared to advocate, and indeed enforce, a social organization founded upon these rules, and of a form which matches very closely to a theoretical social order described by Max Weber, which he labelled 'legal authority' (Weber 1978, p. 31). One of the key features of this system is the assignation of spheres of competence, and the HNPs acted in a way which seemed to reveal strong beliefs that particular volunteers possess legitimate authority, and responsibility, to act in particular ways and not other ways. Of particular note was the way in which HNPs sought to enforce a social hierarchy, aligning themselves with those who set up and maintain climateprediction.net and designating forum moderators and team leaders as having the responsibility and legitimate authority to control and direct the volunteers.

Of course, this is not the complete picture: very different views regarding how projects should operate and how volunteers should act may be held by the many

other participants who have also made particularly large contributions to climateprediction.net and volunteer computing projects in general. These users include those who have donated a particularly large amount of computing capacity, or who regularly respond to requests for advice on the forum, or who enthusiastically recruit new volunteers from their friends and family, and questions regarding their perceptions of other volunteers should be addressed.

In addition, some other directions for future research to inform further developments in volunteer computing are suggested by the emergence, during the course of this paper, of the notion that HNPs perceive the system of measurement and publicization of volunteers' statistics as critical to the success of projects. Hence, HNPs could plausibly consider that any project which compromises this system might not be worthwhile to participate in. One example where a greater understanding of how, and to what extent, the HNPs believe credits and other statistics are important could prove beneficial, is the engineering of more lightweight, easier-to-implement, alternatives to BOINC. These are motivated by criticisms that the computing expertise needed to run a BOINC project often necessitates the employment of technical specialists, a cost that may lead to the exclusion of less well-funded institutions from the benefits of volunteer computing. For the sake of ease-of-implementation, most such proposed alternatives lack provision for something akin to the credit system (for instance, Baldassari *et al.* 2006). Might such an omission jeopardize a project's chances of recruiting HNPs?

By setting up volunteer computing projects, such as climateprediction.net, scientists have managed to exploit some of the vast, and increasing, unused computing

capacity on personal computers around the world. Nevertheless, there is massive scope for increasing the amount donated by members of the public, and the efficiency with which this donated capacity is used. In designing, implementing, and changing volunteer computing projects in the future, scientists and software engineers must consider why members of the public join and continue to participate in projects if progress is to be made towards fulfilling the immense potential of volunteer computing.

ACKNOWLEDGMENTS

- Milo Thurston & Tolu Aina, who work on running and maintaining climateprediction.net, for their time and for cyber-introductions to climateprediction.net volunteers;
- mo.v and Thyme Lawn, two climateprediction.net forum moderators, for advice on how to approach their online community and for promoting my project on the climateprediction.net forums;
- The many climateprediction.net volunteers who have given the time to answer questions I have put to them.

NOTES

[1] http://boincstats.com/stats/project_graph.php?pr=sah (accessed 24 August 2008).

[2]

http://www.top500.org/blog/2008/06/14/preview_31st_top500_list_world_s_most_po

[werful supercomputers topped world s first petaflop s system](#) (accessed 24 August 2008).

[3] http://boincstats.com/page/project_ranking.php (accessed 24 August 2008).

[4] http://boincstats.com/stats/project_graph.php?pr=bo (accessed 24 August 2008).

[5] <http://www.climateprediction.net/project/about.php> (accessed 24 August 2008).

[6] http://boincstats.com/stats/project_graph.php?pr=cpdn (accessed 24 August 2008).

[7] http://boincstats.com/page/project_ranking.php (accessed 24 August 2008).

BIBLIOGRAPHY

Abell, J. & Myers, G. (2008) 'Analyzing Research Interviews', in *Qualitative Discourse Analysis in the Social Sciences*, ed. R. Wodak & M. Kryżanowski, Palgrave Macmillan, New York, pp. 145-160.

Anderson, D. (2003) 'Public Computing: Reconnecting People to Science', *Proceedings of the Conference on Shared Knowledge and the Web*, Residencia de Estudiantes, Madrid, Spain. [online] Available at: <http://boinc.ssl.berkeley.edu/boinc2.pdf> (24 August 2008).

Anderson, D. (2004) 'BOINC: A System for Public-Resource Computing and Storage', *5th IEEE/ACM International Workshop on Grid Computing*, pp. 4-10.

Anderson, D. & Fedak, G. (2006) 'The Computational and Storage Potential of Volunteer Computing', *Proceedings of CCGRID'06*, pp. 73-80.

Baldassari, J., Finkel, D. & Toth, D. (2006) 'SLINC: A Framework for Volunteer Computing', *Proceedings of the 18th IASTED International Conference on Parallel and Distributed Computing and Systems*.

Bursell, M. (2005) 'Security & Trust in P2P Systems', in *Peer to Peer Computing: The Evolution of a Disruptive Technology*, ed. R. Subramanian & B. Goodman, Idea Group Inc., Hershey, Pennsylvania, pp. 145-165.

Christensen, C., Aina, T. & Stainforth, D. (2005) 'The challenge of volunteer computing with lengthy climate model simulations', *Proceedings of the First International Conference on e-Science and Grid Computing*, pp. 8-15.

Cook, G., Pieri, E. & Robbins, P. (2004) 'The scientists think and the public feels: Expert perceptions of the discourse of GM food', *Discourse & Society*, vol. 15, no. 4, pp. 433-449.

Costa, F., Silva, L., Kelley, I. & Taylor, I. (2007) *Peer-To-Peer Techniques for Data Distribution in Desktop Grid Computing Platforms: CoreGRID Technical Report TR-0095*. [online] Available at:

<http://www.coregrid.net/mambo/images/stories/TechnicalReports/tr-0095.pdf>

(accessed 24 August 2008).

Costa, F., Silva, L., Kelley, I. & Fedak, G. (2008) 'Optimizing the data distribution layer of BOINC with BitTorrent', *IEEE International Symposium on Parallel and Distributed Processing*, pp. 1-8.

Dominguez, D., Beaulieu, A., Estalella, A., Gómez, E., Schnettler, B. & Read, R. (2007) 'Virtual Ethnography', *Forum: Qualitative Social Research*, vol. 8, no. 3.

[online] Available at:

http://160.45.170.223:90/ojs_fqs/index.php/fqs/article/view/274/602 (accessed 24

August 2008).

Frewer, L., Hunt, S., Brennan, M., Kuznesof, S., Ness, M. & Ritson, C. (2003) 'The views of scientific experts on how the public conceptualize uncertainty', *Journal of Risk Research*, vol. 6, no. 1, pp. 75-85.

- Galison, P. & Hevly, B. (ed.) (1992) *Big Science: The Growth of Large-Scale Research*, Stanford University Press, Stanford, California.
- Gill, R. (2000) 'Discourse Analysis', in *Qualitative Research with Text, Image & Sound*, ed. B. Martin & G. Gaskell, Sage, London, pp. 172-188.
- Hine, C. (2000) *Virtual Ethnography*, Sage, London.
- Irwin, A. & Wynne, B. (1996) *Misunderstanding science? The public reconstruction of science and technology*. Cambridge University Press, Cambridge.
- Kim, J-S., Bhattacharjee, B., Keleher, P. & Sussman, A. (2006) 'Matching Jobs to Resources in Distributed Desktop Grid Environments', *Technical Report CS-TR-4791 and UMIACS-TR-2006-15*, University of Maryland, Department of Computer Science and UMIACS.
- Korpela, E., Werthimer, D., Anderson, D., Cobb, J. & Lebofsky, M. (2001) 'SETI@home – Massively Distributed Computing for SETI', *Computing in Science & Engineering*, vol. 3, no. 1, pp. 78-83.
- Krishnan, R., Smith, M. & Telang, R. (2002) 'The Economics of Peer-to-Peer Networks', Working Paper, Carnegie Mellon University.
- Markham, A. (1998) *Life Online: Researching Real Experience in Virtual Space*, AltaMira, Walnut Creek, California.
- Pan, Y. (2008) 'Scientific computing on public computing platforms – practices and experiences', *IEEE International Symposium on Parallel and Distributed Processing*.
- Potter, J. & Wetherell, M. (1994) 'Analyzing discourse', in *Analyzing Qualitative Data*, ed. A. Bryman & R. Burgess, Routledge, London, pp. 47-66.
- Pourebrahimi, B., Bertels, K. & Vassiliadis, S. (2005) 'A Survey of Peer-to-Peer Networks', *The 16th Annual Workshop on Circuits, Systems and Signal Processing*.

Stainforth, D., Kettleborough, J., Martin, A., Simpson, A., Gillis, R., Akkas, A., Gault, R., Collins, M., Gavaghan, D. & Allen, M. (2002a) 'Climateprediction.net: Design Principles for Public-Resource Modeling Research', *Proceedings of the 14th IASTED International Conference: Parallel and Distributed Computing and Systems*, pp. 32-38.

Stainforth, D., Kettleborough, J., Allen, M., Collins, M., Heaps, A. & Murphy, J. (2002b) 'Distributed Computing for Public Interest Climate Modeling Research', *Computing in Science and Engineering*, vol. 4, no. 3, pp. 82-89.

Taylor, S. (2001) 'Locating and Conducting Discourse Analytic Research', in *Discourse as Data: A Guide for Analysis*, ed. M. Wetherell, S. Taylor & S. Yates, Sage, London, pp. 5-48.

Wallach, D. (2002) 'A survey of peer-to-peer security issues', *Proceedings of the 46th Annual Meeting of the International Society for the Systems Sciences*, pp. 42-57.

Weber, M. (1978) [1947] *Theory of Social and Economic Organization*, trans. G. Roth & C. Wittich, Oxford University Press, New York.

Weber, M. (1946) *From Max Weber: Essays in Sociology*, trans. H. Gerth & C. Wright Mills, Oxford University Press, New York.

Welsh, E., Jirotko, M. & Gavaghan, D. (2006) 'Post-genomic science: cross-disciplinary and large-scale collaborative research and its organizational and technological challenges for the scientific research process', *Philosophical Transactions of the Royal Society A*, vol. 364, no. 1843, pp. 1533-1549.

Winter, H. & Pettey, C. (2008) *Gartner Says More than 1 Billion PCs In Use Worldwide and Headed to 2 Billion Units by 2014*, [online] Available at: <http://www.gartner.com/it/page.jsp?id=703807> (24 August 2008).

Wynne, B. (1996) 'May the Sheep Safely Graze?', in *Risk, Environment and Modernity*, ed. S. Lash, B. Szerszynski & B. Wynne, Sage, London, pp. 44-83.

Yao, C.-H. (2006) *Grid Computation – The Fastest Super Computer in the World*, [online] Available at: <http://www.csa.com/discoveryguides/grid/review.pdf> (24 August 2008).

Young, N. & Matthews, R. (2007) 'Experts' understanding of the public: Knowledge control in a risk controversy', *Public Understanding of Science*, vol. 16, no. 2, pp. 123-144.