PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

Phil. Trans. R. Soc. A (2010) **368**, 4177–4192
doi:10.1098/rsta.2010.0163

Retaining volunteers in volunteer computing projects

By Peter Darch[1,*] and Annamaria Carusi[2]

[1]*Computing Laboratory, University of Oxford, Wolfson Building, Parks Road, Oxford OX1 3QD, UK*
[2]*Oxford e-Research Centre, 7 Keble Road, Oxford OX1 3QG, UK*

Volunteer computing projects (VCPs) have been set up by groups of scientists to recruit members of the public who are asked to donate spare capacity on their personal computers to the processing of scientific data or computationally intensive models. VCPs serve two purposes: to acquire significant computing capacity and to educate the public about science. A particular challenge for these scientists is the retention of volunteers as there is a very high drop-out rate. This paper develops recommendations for scientists and software engineers setting up or running VCPs regarding which strategies to pursue in order to improve volunteer retention rates. These recommendations are based on a qualitative study of volunteers in a VCP (climateprediction.net). A typology of volunteers has been developed, and three particularly important classes of volunteers are presented in this paper: for each type of volunteer, the particular benefits they offer to a project are described, and their motivations for continued participation in a VCP are identified and linked to particular strategies. In this way, those setting up a VCP can identify which types of volunteers they should be particularly keen to retain, and can then find recommendations to increase the retention rates of their target volunteers.

Keywords: volunteer computing projects; distributed computing projects; public engagement

1. Introduction

In recent decades, there has been an explosion of scientific sub-disciplines and projects that involve the processing of large datasets or the running of computationally intensive mathematical models (Welsh *et al.* 2006; Hey *et al.* 2009). However, acquiring the necessary resources can prove very costly (Yao 2006). Another critical challenge that many scientists have recently faced is a political culture, both at the national and the international levels (for instance, the European Union), which increasingly calls on scientists to show a willingness to interact with and educate the lay public about their work (House of Lords 2000; Wynne & Felt 2007).

In order to meet one or both of these significant challenges, some groups of scientists have set up *volunteer computing projects* (VCPs) which ask members of the public (*volunteers*) to donate processing capacity from their personal

*Author for correspondence (peter.darch@ccc.ox.ac.uk).

One contribution of 15 to a Theme Issue 'e-Science: past, present and future II'.

Table 1. The five largest BOINC-based projects in terms of number of active volunteers as of 19 April 2010 (BOINC Stats 2010).

rank (by active volunteers)	project (host institution in parentheses)	scientific focus of project	active volunteers (to nearest 1000)	average floating point operations per second (teraflops, to nearest 10)
1	SETI@Home (University of California, Berkeley)	search for extra-terrestrial intelligence	178 000	730
2	World Community Grid (IBM Corporation)	variety of research areas	98 000	380
3	Rosetta@Home (University of Washington)	protein folding	44 000	100
4	Einstein@Home (University of Wisconsin–Milwaukee)	astronomy	39 000	210
5	climateprediction.net (University of Oxford)	modelling climate change	33 000	100

computers (Anderson 2003). In a typical VCP, a volunteer will download and process a block of data or a scientific model (a *work unit*) on their personal computers and, once complete, will upload the processed work unit to the project's central server. The first VCP was launched in 1996, and since then many have been set up in a variety of scientific fields. For instance, there are over 40 current VCPs in fields as diverse as biochemistry, physics and climate science, which use a piece of middleware called the *Berkeley Open Infrastructure for Network Computing* (BOINC) that provides a framework for scientists wishing to set up a VCP (BOINC Stats 2010). Table 1 shows the five largest BOINC-based projects in terms of the number of active volunteers: each has in excess of 30 000 active users from over 200 countries (often including a large number of volunteers with little or no other apparent links to scientific institutions—genuinely lay members of the public) providing a computing capacity that is equivalent to that of a powerful supercomputer.

In the light of the challenges outlined above, VCPs tend to have one or two of the following goals (this will be explained in greater detail in §2):

— to use the donated computing capacity to produce scientific results that can lead to journal articles, conference presentations and other scientific publications; and
— to engage and educate members of the public about their field of research.

An improvement in the retention rate of volunteers should help in the pursuit of both of these goals, given that it should increase the rate at which work for the project is done, and that being more involved with a project will provide more opportunities for a volunteer to learn about the underlying science. Indeed, it appears that there is substantial scope to achieve this: as of April 2010, nearly five out of every six volunteers who have registered for BOINC projects no longer participate (BOINC Stats 2010).

Nevertheless, there has been little consideration in the literature about how to improve retention rates, and there appears to be some confusion about how to proceed. The conclusions of papers produced by the scientists and software engineers about their experiences of working on VCPs are at odds with the results of a survey of volunteers in BOINC-based projects carried out in 2006 and 2007 regarding volunteers' motivations to continue participating in VCPs: while the former tend to ascribe volunteers' motivation to continue participating in the project primarily to the competitive aspect of the credit system, with volunteers keen to amass more credits and see themselves and their teams move up the league tables (e.g. Christensen *et al.* 2005), approximately three-quarters of respondents to the latter said that they were motivated to participate in a project because 'The science is important and beneficial', compared with fewer than one in six who indicated that a 'Fair and quick granting of credit for work done' or 'Getting more credit from this project than from others' were important factors in deciding whether to continue participating in a BOINC project (BOINC Project 2007).

Drawing on a qualitative case study of such a VCP (climateprediction.net), this paper seeks to provide recommendations for improving retention rates of volunteers. In particular, given that different VCPs may have different mixes of goals, and that some volunteers may make a particular contribution to certain of these goals and other volunteers may make a larger contribution to other goals, the research presented in this paper attempts to:

— identify which particular classes of volunteers might make an important contribution to a project's particular mix of goals;
— understand what motivates these volunteers to continue participating in a project; and
— recommend strategies that VCPs should pursue to retain these volunteers.

Before proceeding with the rest of this paper, it should be noted that there are some technical issues to consider before deciding whether to set up a VCP in the first place. These include the following.

— Issues relating to members of the public, in particular whether:
  (i) the task can be run on hardware and software that a standard personal computer is likely to possess (Anderson 2003);
  (ii) potential volunteers are likely to trust that participating in the project will not put sensitive data at risk; and
  (iii) the middleware and applications are such that volunteers' computers are vulnerable neither to malicious attacks nor to other risks of damage (Marosi *et al.* 2007).
— Issues of economy, in that the gains to the researchers of using volunteer computing should warrant the costs of the following (Kondo *et al.* 2009):
  (i) staff salaries for installing, programming and maintaining server software;
  (ii) purchasing and maintaining hardware; and
  (iii) transferring data from the central server to volunteers' computers, and back again (for instance, in terms of bandwidth required).

Table 2. Classification of VCPs according to their main goals.

example of projects	computational power sought	strong apparent commitment to public education?
climateprediction.net; World Community Grid	relatively large scale	yes
Einstein@Home	relatively large scale	no
BRaTS@Home	relatively small scale; sporadic	no

2. Goals of volunteer computing projects

In the discussion below, VCPs will be grouped into three broad categories depending on their mix of goals: these categories will be used later in this paper to match the particular classes of volunteers identified in the case study of climateprediction.net to the VCPs to which they might prove particularly useful.

The first group of VCPs consists of those with the goals of both seeking to process large-scale datasets and to educate large numbers of the lay public. One example of these is climateprediction.net, which set out to run between one and two million climate models on volunteers' computers but which also has a very strong commitment to public education, with a website containing a great deal of information about climate science and about the models volunteers are running on their computers, as well as a programme of talks and lectures to schools and universities (Christensen *et al.* 2005). Another notable VCP that combines these two goals is the World Community Grid, which has a range of research projects with a strongly humanitarian focus from a wide variety of scientific disciplines (such as medical research) and approximately 100 000 active volunteers (World Community Grid 2010).

The other two categories of projects focus primarily on accessing computation resources and do not have such a clear commitment to public education. Such projects tend to fall into one of the two following categories.

— Large-scale projects with a continuous supply of work units for volunteers. These have often been set up with grants awarded by public funding bodies, and have a number of personnel in scientific institutions working full-time on their operation. An example is Rosetta@Home, which seeks to determine the three-dimensional shape of proteins, is based at the University of Washington, and has approximately 44 000 volunteers (Rosetta@Home 2010).
— Small-scale projects, where work is available to volunteers on a more sporadic basis. An example is BRaTS@Home, which studies gravitational rays, and is run at the University of Missouri. It has only approximately 50 active volunteers as participation is by invitation only (BRaTS@Home 2010).

These classifications are summarized in table 2.

3. The climateprediction.net project

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. It received funding from the Natural Environment Research Council as part of its e-Science Programme, and from the Department of Trade and Industry (now the Technology Strategy Board). As mentioned above, it has two overriding goals (Stainforth *et al.* 2002).

— To produce scientific results. The main project of climateprediction.net is based on a particular climate model, the HadCM3, which models atmosphere and ocean temperature from 1920 to 2080 (Gordon *et al.* 2000). The project aims to use this model to produce a probabilistic forecast for future climate change. There are also a number of other sub-projects, for instance predicting the impact of various climate change mitigation strategies.
— To educate members of the public about the science of climate change.

The climateprediction.net project began in 1999, and in 2002 the decision was taken to transform it into a VCP. 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.

A single run of a climateprediction.net model is known as a *work unit*. A typical work unit for climateprediction.net (based on the HadCM3 model) is expected to take approximately four months to run on a standard personal computer, and as such, is substantially longer than for almost any other BOINC-based project. For instance, work units for SETI@Home or Einstein@Home are expected to take a matter of hours, rather than days, weeks or months, to complete on a standard personal computer (Christensen *et al.* 2005).

Using BOINC as a framework has allowed climateprediction.net to incorporate several features for project volunteers. One of these is an interface through which volunteers can communicate with the project, download work units and upload completed models and that also allows volunteers to register with other BOINC projects, so some volunteers participate in other projects simultaneously with climateprediction.net.

Online forums are another key feature for volunteers. 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 *credit system* is another feature. Credits are points that are awarded to a volunteer after every 10 model-years are completed on a volunteer's computer (this is in contrast to many other BOINC projects with shorter work units, where credits are usually awarded only after a work unit is completed).

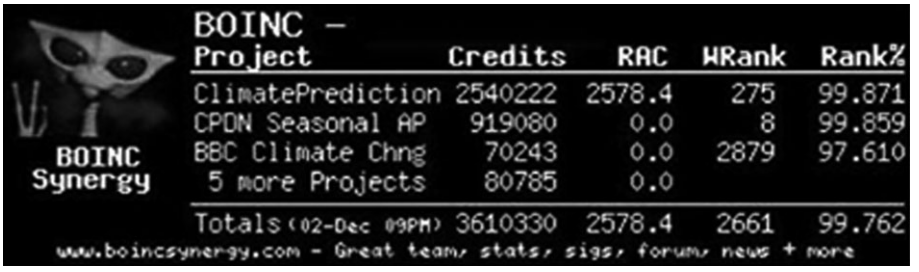

Figure 1. A climateprediction.net volunteer's signature. This gives the volunteer's current total credit score within the main climateprediction.net project and two climateprediction.net sub-projects, the RAC and the rank of the volunteer in comparison with other volunteers both as an absolute figure and as a percentile.

A volunteer is assigned two scores related to credit:

— *total credit* awarded since the volunteer became involved in the project; and
— *recent average credit* (RAC), which is a weighted average of the credit awarded to a volunteer, with the weighting given to the credit decaying exponentially over time, so that the greatest weighting is given to the credit most recently awarded.

Credits are very often a source of pride for project volunteers, who will sometimes report their credit score in a signature placed at the end of forum posts (figure 1). Furthermore, volunteers can group together and form teams, which very frequently have websites and forums of their own. On the climateprediction.net website there are lists ranking both individual participants and teams according to their credit scores (the statistics for teams are simply aggregates of the individual statistics). It should be stressed that policies for awarding credits are determined at the discretion of the individual project: credits from one project are not equivalent to those in others.

Finally, a screensaver (figure 2) allows volunteers to track the progress of their climate models over time.

4. Methodology

The research presented here is based on a qualitative study of climateprediction.net, loosely following the methodological approach of studying online communities set out in Hine (2000). The research was directed towards developing a typology of volunteers, particularly focusing on:

— the behaviour that different types of volunteers display in VCPs (for instance, the number of projects that they are involved in, the amount of processing capacity they make available to the projects, and the number and content of contributions to online VCP forums);
— what motivates different types of volunteers to participate, and to continue participating in VCPs;

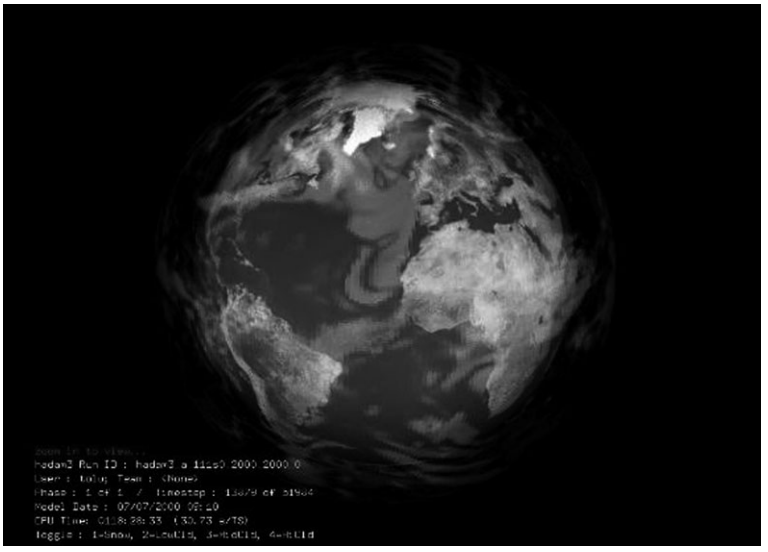

Figure 2. A typical climateprediction.net screensaver, showing the cloud cover and global temperature predicted by a volunteer's model.

— which particular features of VCPs affect this motivation, and why; and
— the perceptions that different types of volunteers have of other volunteers.

To this end, a corpus of data was assembled for analysis, including in particular:

— an open-ended questionnaire aimed at climateprediction.net volunteers with participants ($n = 35$) recruited by threads posted on the climateprediction.net forums by two moderators; and
— a large number of threads from the project's forums involving a wide range of volunteers (including both those who are currently active and those who have left the project) and covering a variety of topics.

Drawing on both sources of data helped to address, at least in part, three major issues that would have arisen if using questionnaires alone, because analysing forum threads enabled the following.

— A far larger number of volunteers to be studied.
— The study of some volunteers who are no longer active. It was possible only to recruit currently active volunteers for the questionnaire, as they could only be contacted either through the forums or messages sent to their climateprediction.net account (which they could receive only upon logging in). Forum posts by volunteers who are no longer active gave some insights into their reasons for no longer participating.
— The cross-checking of questionnaire responses. For instance, many responses referred to events that took place on the forums, which allowed for the mitigation of imperfect recall by respondents.

5. Findings and recommendations

In this section, results from the case study are presented. As a result of this study, a number of different classes of volunteers have been identified, based on the extent and nature of their participation in climateprediction.net and other VCPs: in other words, different classes contribute in different ways to the pursuit of the project's various main goals and objectives.

Owing to space constraints, the results presented below focus on three of these classes, with each such class selected on the basis that the volunteers in that class make a particular contribution to the main goals of one of the types of project identified in §2. These classes, along with their main characteristics and the goals to which they might contribute, are summarized in table 3.

In the discussion below, the nature and extent of the participation by each class in climateprediction.net will be described, and their particular value to VCPs will be explained. The specific reasons they continue to participate in the project will be outlined, along with how the particular features of the project (such as an educational component) or policies (for instance, relating to the attribution of credit) make their continued participation more or less likely. It has been found that almost all volunteers are fundamentally motivated by the idea that they are helping to produce scientific results that will benefit humanity, but different classes of volunteers are reassured (or otherwise) that this is the case in different ways: some features of the project can help to reassure some volunteers, whereas other volunteers are more influenced by other features or policies.

(*a*) *Super-Crunchers*

Super-Crunchers form approximately 10 per cent of all active volunteers within climateprediction.net (the figures given here are derived from a random sample of 1000 active climateprediction.net volunteers). This class of volunteers is characterized by the relatively large quantity of project data that they process on their own computers, often running BOINC on a number of computers simultaneously for many hours each day (some keep their computers running for all 24 hours every day), and they have frequently adapted their computers to increase their processing capacity. Unlike many other volunteers who also display a strong commitment to participating in VCPs (including the *Alpha-Testers*, and those who may become forum moderators in a particular VCP), the *Super-Crunchers* tend to restrict themselves to only a handful (usually between one and three) of VCPs, which allows them to accumulate a high number of credits in specific projects (rather than spreading their computational resources, and hence credit scores, more thinly over a larger number of projects).

As a result of this, the *Super-Crunchers* participating in climateprediction.net can be seen to make a particularly valuable contribution to the project's goal of producing publishable scientific results: approximately 60 per cent of all credit awarded by climateprediction.net has been awarded to the *Super-Crunchers*. Indeed, within this class, there is a great deal of variation in the level of work done among *Super-Crunchers*, with over 10 per cent of credits awarded to just 0.2 per cent of active volunteers.

Table 3. The classes of volunteers discussed in this paper and the types of VCPs to which they might make a particularly important contribution.

class of volunteer	description	typical characteristics (note that these are not fixed criteria for inclusion across all VCPs, for instance, because different VCPs attribute credit in different ways)	VCPs to which this class of volunteers might make a particular contribution
Super-Crunchers	volunteers whose computers process a particularly large quantity of project data	RAC > 1000; involved in 0–2 BOINC-based projects in addition to climateprediction.net	those seeking large-scale computational power
Lay Public	volunteers who have little or no other contact with scientific institutions outside of VCPs	RAC < 100; involved in 0–2 BOINC-based projects in addition to climateprediction.net	those seeking large-scale computational power; those with a strong commitment to public education
Alpha-Testers	volunteers who test new features of VCPs (e.g. new models, updated software, etc.) or who participate in alpha testing phases	involved in 12 or more BOINC-based projects	those seeking small-scale, sporadic computational power

The *Super-Crunchers* enjoy the prestige that high credit scores bring within the online community of project volunteers (both in climateprediction.net's own forums, and the forums of the teams they join), and this is what primarily motivates their continued participation in climateprediction.net. They regularly make posts in these forums when they have reached particular milestones (for instance, once they have a particular number of credits (usually measured in hundreds of thousands, or even millions) or completed a particular number of climate models), for which they receive praise from a broad cross section of other volunteers. To complement this, a Facebook application was launched in summer 2009, which enabled credit milestones to be displayed on a volunteer's Facebook profile. Furthermore, the *Super-Crunchers* often display their credit scores in climateprediction.net (and other BOINC-based projects) in the signature they attach to forum posts (see figure 1).

(i) *Recommendations*

The *Super-Crunchers* of climateprediction.net have been retained owing to the way that the project's credit system and online forums operate. Its credit system has been relatively stable and consistent, and furthermore is updated on a daily basis, so that volunteers have a high degree of trust in its reliability and robustness as an accurate measure of project work completed on a volunteer's set of computers: this means that when a *Super-Cruncher* posts that they

have reached a particular credit milestone, other volunteers are willing to offer praise and congratulations because they are confident that it represents a genuine accomplishment. Equally important is the activity of volunteer online forums: because there are so many volunteers reading and posting on them, this reassures the *Super-Crunchers* that it is worthwhile to post about achieving credit milestones because these posts will be noticed by many others.

In general, if VCPs wish to attract and retain the interest of a body of *Super-Crunchers* to process data, the following appears from the above.

— *A VCP should include a system of attribution/scoring (such as the credit system)* that is

(i) easy to understand by volunteers and

(ii) regarded by volunteers as an accurate and reliable indicator of a particular volunteer's contribution to the project.

Such a system should

(i) be cumulative, so that an increase in score is associated with additional data processed;

(ii) be consistent throughout the duration of a project;

(iii) operate under rules that are equally applicable to all volunteers, so that all volunteers are scored on the same basis, allowing for easy comparison between volunteers; and

(iv) be updated on a regular basis.

— *A VCP should ensure that the scores achieved by volunteers are very visible to other volunteers.* This happens within climateprediction.net owing to the very active forums, and the display of credits on a league table on its website. It is not, however, inevitable that a VCP's forums will be used by a large number of volunteers. To nurture activity, those who run a VCP could post regularly, for instance, about the science underlying their project, issues relating to the science to stimulate debate (the fields of many VCPs relate to current political debate, such as climate change in the case of climateprediction.net), or post when volunteers have reached credit milestones in order to encourage a culture of praise for volunteers who achieve high scores.

(*b*) *Lay Public*

The label *Lay Public* is applied to the class of volunteers who have little or no contact with scientists or scientific institutions other than through climateprediction.net. They comprise approximately 80 per cent of active volunteers within climateprediction.net, but less than a quarter of all credits are attributed to them.

Compared with the *Super-Crunchers*, therefore, the *Lay Public* volunteers make a smaller contribution to the project in terms of producing scientific results. Nevertheless, they are very important to the project because they are the focus of its outreach efforts; climateprediction.net is a valuable way of communicating science to a group of people for whom scientific institutions may have few other opportunities to engage them.

It has been suggested that the credit and team systems play a critical role in the retention of these volunteers, because they enjoy the competitive element of accumulating credits and comparing them via league tables (Christensen *et al.* 2005). Indeed, it certainly appears that this is the case, because many *Lay Public* volunteers will post threads on the forums when they notice that credits have not been updated, with some even threatening to leave the project if they fear that the credit system is not stable.

However, this is not the whole story. *Lay Public* volunteers have a strong tendency to stress that they are not motivated by the accumulation of credits and, indeed, that they find the idea of being motivated by this offensive. Instead, their most frequently expressed motivation for participating in climateprediction.net is the notion that, by so doing, they are contributing to the good of science and society.

This is an apparent paradox: the credit system plays a critical role in the retention of these volunteers, but they are primarily motivated by the idea of contributing to climate science. This arises because the role that the credit system plays here is not to create an element of competition but to reassure volunteers that the climateprediction.net project is progressing towards providing new scientific knowledge and that the particular models they are running on their own computers are contributing to this progress. The credit system provides such reassurance because the way in which it operates coincides with beliefs expressed by *Lay Public* volunteers about how science works: they believe that science is a *progressive accumulation* of knowledge about the world, and the progressive accumulation of credits is a marker of this.

By seeing both the project's total credit (which is displayed on the front page of climateprediction.net's website) score grow, and their own credit score increase regularly as part of this growth, they can find reassurance that:

(i) the project is progressing;
(ii) their own models are progressing; and
(iii) their own models' progress forms a part of the overall progress of the project.

(i) *Recommendations*

A VCP with a strong commitment to public engagement in science should seek to retain a large number of *Lay Public* volunteers. In light of the above, the following is recommended.

— Regular feedback is provided to volunteers about the project's progress, for instance through posting regular updates on the project's forums, or posting links to journal articles and conference presentations. This will provide reassurance to volunteers that the project itself is producing worthwhile, certified scientific results.
— Regular feedback is provided to volunteers about the progress of their work units. In the case of climateprediction.net, this takes the form of the screensaver, as well as a display on the BOINC interface saying how many model years have elapsed. This feedback need not necessarily be as

sophisticated as this because what the volunteers simply require is being able to see the progress of their work unit: a display of the percentage of the work unit that has been completed may suffice.

Although the two recommendations above will help to meet challenges (i) and (ii) mentioned above, they do not tie the progress of a volunteer's model to the progress of the project as a whole. Hence, the following is also recommended.

— A scoring system akin to the credit system of climateprediction.net is included. This is similar to what has been recommended in the case of *Super-Crunchers*; however, there are certain additional points that should be borne in mind if a VCP wishes to have a large number of *Lay Public* volunteers as well. In line with *Lay Public* volunteers' view of science, such a system should be cumulative and progressive. Additionally, the following can be stated.

(i) The project should display publicly, for instance on its website, the aggregate score amassed by all volunteers. This enables individual volunteers to tie their own work unit's progress in with the progress of the project as a whole.

(ii) The score for an individual *Lay Public* volunteer should be updated regularly. It is important to remember that *Lay Public* volunteers' work units are likely to progress many times more slowly than those of, for instance, the *Super-Crunchers*. This is a challenge that has been successfully navigated by climateprediction.net. As mentioned in §2, a climateprediction.net work unit is supposed to take a few months to run on a typical home computer (i.e. one that a *Lay Public* volunteer might be supposed to have), so if credits are awarded only once a work unit is completed, a *Lay Public* volunteer might not receive any credit for a long period of time, and hence lose motivation: as a result, climateprediction.net awards credits after every 10 model years are completed.

However, it should be cautioned that a VCP that prioritizes the processing of data and has a less strong commitment to public education should be wary of having a score system that is updated too regularly. It is in the interest of a VCP that its work units are run through to completion and awarding credits for only part-completed work units has been found to encourage some volunteers to abandon work units before they are completed: work units often crash part-way through and require effort on the part of the volunteer to restart, and in such a case a volunteer is more likely to try to rescue the work unit rather than abandon it for a new one if credits are awarded only for completed work units.

(*c*) *Alpha-Testers*

The final class of volunteers considered here is called the *Alpha-Testers*, so-called because it comprises those who are recruited by VCPs that are at the alpha stage of development. Many well-established VCPs also retain a group of such volunteers to be able to test new software or new types of work units (for instance, in the case of climateprediction.net, the HADam3 seasonal variation sub-project).

The *Alpha-Testers* in climateprediction.net are often registered with 15, 20, 30 or even more other BOINC-based projects, and appear to donate a very large quantity of computing capacity to VCPs (a comparable amount to the *Super-Crunchers*). As with the *Super-Crunchers*, the *Alpha-Testers* enjoy the status they have within a VCP, in their case as part of a small group of volunteers to which admission is by invitation only. It is the possession of such a status that motivates such volunteers to continue to participate and, as such, they are highly reliable volunteers; they are keen to maintain their reputation to ensure they continue to be invited to take part in alpha testing.

As with the other classes, the way in which the credit system operates impacts on the *Alpha-Testers*' motivation to participate, but their view of the credit system is very complex. On the one hand, they are unconcerned with their own acquisition of credits: their effort is spread across many projects, rather than focusing on acquiring a large number of credits in a few projects like the *Super-Crunchers*, and, furthermore, any credits that may be earned during alpha testing phases are not automatically transferred to their overall credit scores for a particular project. On the other hand, they appear to believe that all other (non-*Alpha-Tester*) volunteers are motivated primarily by the acquisition of credits and express a great deal of concern when they feel that credit system policies are unclear or may operate in a way that encourages behaviour amongst volunteers that works against the project's interests (for instance, if they believe the way in which credits are attributed will encourage volunteers to abandon work units before completion): if *Alpha-Testers* believe that the credit system works in such a way, they are liable to withdraw from a VCP because they believe that their efforts would be better used elsewhere.

(i) *Recommendations*

Small-scale projects (*such as BRaTS@Home*). As discussed in §1, these projects often only have work units available for volunteers sporadically, so their progress is usually irregular and credits (in the case of BOINC-based projects) are not available in large quantities. As a result, they are not likely to attract either *Lay Public* or *Super-Cruncher* volunteers. However, *Alpha-Tester* volunteers might prove useful volunteers for such projects. Because these projects have substantially less work to process than larger-scale projects such as climateprediction.net, they do not need such a large base of volunteers. Instead, an alternative strategy might be to restrict participation to a selected group of volunteers, as is the case with BRaTS@Home. In such a case, it may be advised to target *Alpha-Testers*, because

— they are likely to be attracted to a project that only allows selected volunteers to participate;
— as mentioned above, *Alpha-Testers* are not motivated by having others recognize the quantity of work they are able to do for any particular project, and hence are unlikely to be put off by the sporadic availability of work; and
— they can be counted on to do work for a project when asked, because they are keen to establish a reputation as reliable volunteers to maintain their status.

Whether such a project has a system of attribution for work done (such as a credit system), or how such a system operates, should not matter for the retention of *Alpha-Testers*.

Large-scale projects (*such as climateprediction.net*). When such a project is at a preliminary alpha-testing phase or an existing project is testing new features such as new types of work units, the project's software engineers and scientists have a great deal of work to do regarding development, testing, and refinement of project software, and as a result may defer decisions about what form a system of attribution for volunteers, such as a credit system, might take (what to measure, e.g. number of completed work units or total processing donated to the project, how regularly volunteers' scores are updated, how flexible policies regarding scores are to be, whether there are league tables, etc.).

Nevertheless, *Alpha-Testers*' concerns regarding acquisition of credits should be taken into account. It is therefore recommended that software engineers on large-scale projects ensure that they convince *Alpha-Testers* at an early stage of alpha testing (either in a new project or of a new feature of an existing project) that their proposed credit system will

— have firm, consistent and transparent rules; and
— reward work that benefits the project, for instance, by ensuring that volunteers have the incentive to run work units through to completion rather than abandoning them uncompleted if they have difficulty running them on their computer (for instance, if they crash regularly). Furthermore, it is recommended that *Alpha-Testers* are involved with the development of a project's scoring system, in order to ensure they are satisfied that it will operate in the best interests of the project.

6. Conclusion

Different VCPs aim to achieve different goals. Based on a case study of one VCP, a typology of volunteers was developed with the aim of identifying which volunteers make especially strong contributions to the particular goals of a VCP. This paper focused on three classes of volunteers within the typology (the *Super-Crunchers*, the *Lay Public*, and the *Alpha-Testers*) and it was explained how each class might contribute to a specific desired outcome of a VCP (see table 3). For each of the three classes, consideration was given to what motivates volunteers from that class to continue participating in a VCP, and how features or policies of a VCP might work to provide this motivation. This meant that a direct link could be made between the main goals of a VCP and recommendations for the features and policies which should be incorporated into a project (for instance, how a credit-style points system might operate or whether participation should be by invitation only) in order to help pursue these goals.

However, it should be noted that the recommendations in this paper are only designed to help the pursuit of broad goals relating to the size of the datasets that a VCP intends its volunteers to process and whether or not it seeks to educate members of the public about its underlying science. The choice of what strategies a project should employ may involve more complex considerations than those covered in this paper. For instance, a VCP may have other goals besides those

discussed above, an example of which is AlmereGrid, a VCP whose field is medical sciences, which has the goal of promoting civic cohesion within the city of Almere in The Netherlands (AlmereGrid 2010).

It should also be cautioned that the recommendations provided above are based on the study of one VCP alone, and therefore care should be taken if applying them in the context of other VCPs. For instance, other distributed computing projects might require the incorporation of very different policies or features to motivate volunteers, because they involve very different relationships between volunteers and projects from those discussed above: one example of such a project is GalaxyZoo, in which volunteers participate by actively classifying galaxies, as opposed to the more passive form of participation involved in climateprediction.net (http://galaxyzoo.net/). Furthermore, it should be borne in mind that climateprediction.net may find it easier to interest potential volunteers than other VCPs given that climate science is very much in the public eye, and therefore its goals and work may be more easily explained to volunteers than many other fields which involve computationally intensive work (for instance, the Rosetta@Home project focuses on protein folding, a field of investigation whose significance may be less obvious to the lay public). Scientists considering VCPs in such fields should therefore be aware that they may need to employ other strategies in addition to those described above in order to encourage and maintain the interest of volunteers.

Nevertheless, it is hoped that the research and conclusions presented here will help those involved in the running of VCPs to navigate the many options that they face regarding what features and policies to incorporate into their project.

Particular thanks are given to Milo Thurston and Tolu Aina, who work on running and maintaining climateprediction.net in the Computing Laboratory at the University of Oxford, 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; and the many climateprediction.net volunteers who have given their time to answer questions I have put to them.

References

AlmereGrid. 2010 Help de wetenschap een handje. See http://almeregrid.nl.

Anderson, D. 2003 Public computing: reconnecting people to science. *Proc. Conf. on Shared Knowledge and the Web, Madrid, Spain*. See http://boinc.ssl. berkeley.edu/boinc2.pdf.

BOINC Project. 2007 BOINC users' survey. See http://boinc.berkeley.edu/poll_results.php.

BOINC Stats. 2010 Detailed user, host, team and country statistics with charts. See http://boincstats.com.

BRATS@Home. 2010 BRATS@Home gravitational lensing. See http://maxwell.dhcp.umsl.edu/brats.

Christensen, C., Aina, T. & Stainforth, D. 2005 The challenge of volunteer computing with lengthy climate model simulations. *Proc. 1st Int. Conf. on e-Science and Grid Computing, Melbourne, Australia, 5–8 December 2005*, pp. 8–15. Washington, DC: IEEE Computer Society.

Gordon, C., Cooper, C., Senior, C., Banks, H., Gregory, J., Johns, T., Mitchell, J. & Wood, R. 2000 The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustment. *Climate Dyn.* **16**, 147–168. (doi:10.1007/s003820050010)

Hey, T., Tansley, S. & Tolle, K. 2009 *The fourth paradigm: data-intensive scientific discovery*. Redmond, WA: Microsoft Research.

Hine, C. 2000 *Virtual ethnography*. London, UK: Sage.

House of Lords. 2000 *Science and society*. London, UK: The Stationary Office.

Kondo, D., Javadi, B., Malecot, P., Cappello, F. & Anderson, D. P. 2009. Cost-benefit analysis of Cloud Computing versus desktop grids. *Proc. IEEE Int. Symp. on Parallel and Distributed Processing, Rome, Italy, 23–29 May 2009*, pp. 1–12. Washington, DC: IEEE Computer Society.

Marosi, A., Gombas, G., Balaton, Z. & Kacsuk, P. 2007 SZTAKI desktop grid: building a scalable, secure platform for desktop grid computing. In *Making grids work* (eds M. Danelutto, P. Fragopoulou & V. Getov), pp. 363–374. New York, NY: Springer.

Rosetta@Home. 2010 What is Rosetta@Home? See http://boinc.bakerlab.org/rosetta/rah_about.php.

Stainforth, D. *et al*. 2002 Climateprediction.net: design principles for public-resource modelling research. *Proc. 14th IASTED Int. Conf. on Parallel and Distributed Computing and Systems, Cambridge, MA, 4–6 November 2002*, pp. 32–38. Calgary, Canada: ACTA Press.

Welsh, E., Jirotka, 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. *Phil. Trans. R. Soc. A* **364**, 1533–1549. (doi:10.1098/rsta.2006.1785)

World Community Grid. 2010 World Community Grid—about us. See http://www.worldcommunitygrid.org/about_us/viewAboutUs.do.

Wynne, B. & Felt, U. 2007 *Taking European knowledge society seriously: report of the Expert Group on Science and Governance*. Brussels, Belgium: Commission of the European Communities.

Yao, C.-H. 2006 Grid computation—the fastest super computer in the world. See http://csa.com/discoveryguides/grid/review.pdf.