Institution:	**University of Oxford**
Unit of Assessment:	**11 Computer Science and Informatics**
Title of case study:	**BOINC – Volunteer Computing (5)**

1. Summary of the impact (indicative maximum 100 words)

Research in the UoA has underpinned the development of the current version of BOINC (Berkeley Open Infrastructure for Network Computing), a technology to enable secure volunteer computing. The research was done as part of the climateprediction.net project that is currently managed as CPDN through the UoA, supporting international climate modelling. CPDN models climate change using donated cycles on users' computers, with almost 700,000 users registered by 2013. Significant work to develop BOINC in CPDN has enabled the public to engage with science more easily and conveniently. BOINC has become recognised as the key open-source tool for volunteer computing and is also available to companies to create their own grid networks. It has been used for a range of applications from driving experiments to find the Higgs particle to using home PCs to detect earthquakes.

2. Underpinning research (indicative maximum 500 words)

The initial launch of climateprediction.net used SETI@HOME, an early implementation of the BOINC base platform, and highlighted a series of deficiencies with this technology. A concern from users of volunteer computing was security and assurance that their home PC would not be affected by having others run simulations on their machines. This potential vulnerability was highlighted in a paper by Oxford computer scientists in 2004 [1]. It was therefore imperative that BOINC was enhanced to ensure that assurances could be provided, or the concept of volunteer computing would not succeed. Deficiencies were also highlighted in the effectiveness of communication techniques between the individual simulation cycles and the controlling system to ensure dynamic evolution of the simulations. Details were highlighted in a paper focusing on the challenges of volunteer computing with lengthy simulations [2], as well as the design of an infrastructure for distributed servers to manage the data communication [3]. The former paper presented various issues with running lengthy work-units and large, complex applications in volunteer computing. It discussed the challenges to be overcome in constructing the climateprediction.net project, in terms of: porting the scientific models; ensuring the model produces checkpoints; the breaking up of the models into smaller, more manageable chunks; creating the infrastructure and building; and retaining the user community.

Oxford University staff were involved in the CPDN project from both an Atmospheric Physics perspective for the model development and a Computer Science perspective for the work on BOINC, with the latter team comprising Andrew Martin, Andrew Simpson, Carl Christiansen and Tolu Aina from the UoA. Christiansen oversaw the major redevelopment of BOINC; Aina left the UoA in 2010 and is now the BOINC chief developer. Drs Martin and Simpson are academic staff in the UoA whose expertise lies in distributed computing and security.

The major CPDN contributions to BOINC have been:

- Integration of the LibCurl HTTP library. This library is used for the transferring of files via HTTP to and from the project servers, and as such is fundamental to BOINC as a distributed computing project. It was a huge task to integrate this into BOINC, and all current projects using BOINC benefit.
- Support for 3-D fonts in OpenGL.This is used by almost all projects that offer graphics and have 3D visual models running on users’ machines.
- Adapting zip/unzip libraries to BOINC. Many projects use zip/unzip to handle large numbers of files more efficiently and enhance data communication. The files that are distributed to participants are in a zipped form, which reduces both the total size of the data held by the project and the volume of data transferred to the participants. The zip/unzip libraries enable the files to be automatically handled on the participant's side both prior to the start of the computational model, and whilst uploading the results to the project servers. Many projects

use this feature to handle the large numbers of files commonly produced in a BOINC project more efficiently and to enhance their data communications.

The UoA and their collaborators also worked on the porting of the scientific codes as the original codes were meant to be run on clusters or large infrastructures, not single processors. This involved taking code developed by physicists over many years in, typically, a million lines of Fortran code, and rewriting or modifying it to run across multiple platforms for individual processors. Whilst BOINC is a community developed open-source tool to date, the early work for CPDN was instrumental in its transformation into the widely used tool that it is today. In particular, by developing the security aspects of this platform, it has been able to offer the security assurances required by the users.

3. References to the research (indicative maximum of six references)

*[1] David Stainforth, Andrew Martin, Andrew Simpson, Carl Christensen, Jamie Kettleborough, Tolu Aina, and Myles Allen. Security Principles for Public-Resource Modeling Research, Proceedings of the 13th IEEE Conference on Enabling Grid Technologies (ENTGRID), Modena, Italy, June 2004. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.97.8980.

In this paper, the crucial steps taken in the climateprediction.net project to address the security concerns inherent in the design of a volunteer computing project are described.

*****[2] Carl Christensen, Tolu Aina, David Stainforth. The Challenge of Volunteer Computing With Lengthy Climate Modelling Simulations, Proceedings of the 1st IEEE Conference on e-Science and Grid Computing, Melbourne, Australia, 5-8 Dec 2005.

This paper presents various issues with running lengthy workunits and large, complex applications in volunteer computing.

* [3] N. Massey, T. Aina, M. Allen, C. Christensen, D. Frame, D. Goodman, J. Kettleborough, A. Martin, S. Pascoe and D. Stainforth. *Data access and analysis with distributed federated data servers in climateprediction.net* , Advances in Geosciences, 8, p49-56, 2006.

This paper discusses the project data handling in terms of the distribution of computational work, the collection and storage of the results, and the subsequent analysis.

4. Details of the impact (indicative maximum 750 words)

The primary impact is that the UoA's underpinning research enabled *large-scale public engagement with science* by improving the security and usability of BOINC: allowing users to be confident of BOINC's ability to harvest compute cycles without compromising the security of their PCs, and making modellers able to develop more complex codes to run out in the community and return timely results. It enabled the public to engage in science in a way that had not been achieved before with thousands of users observing a changing globe graphic on their machines, showing the results of their own models running on behalf of the climate scientists [A]. CPDN was, and still is, producing climate modelling results which are used by the Met Office.

BOINC [B] is managed as a community developed open-source tool from Berkeley University. The UoA's contribution is acknowledged by the current head of this group [C]. BOINC is now widely used by scientific communities across the world. According to www.boincstats.com, 2,599,338 volunteers have installed the BOINC software on their machines and there are currently 272,459 active users.

The global reach of the CPDN project using BOINC is significant. CPDN has 20,000 active users contribute approximately 27,000 active hosts, with a combined power of 35 TFlops. Of the 82 BOINC projects, in May 2012 climateprediction.net was the 4th most popular by work-units in progress and 5th in attracting new users [B]. Over 129 million simulation-years have been performed since the project's inception and registered users are located in 221 countries. Although the project has been running since September 2003, 9 of the 10 busiest days have been since January 2008, and currently around 30-40 new users join each day. Since 2008, over 100,000 users have joined and over 60% of the total computational cycles in the project have been donated [A]; 40,702 distinct individual users have successfully completed one or more model simulations.

The computing time required for all the model simulations run successfully since 2008 is equivalent to a 32,220 core machine running full time for one year and producing 100% successful results. An estimate of the value of this CPU time is $22.5M, based on the rate of the Amazon Elastic Compute Cloud Standard Spot instance ($0.08/hour default).

The following quote [C] from David Anderson, Director of the BOINC Project, indicates his belief in the impact of CPDN:

'*I'd like to congratulate and thank all the people at Oxford who made it happen, and all the volunteers who courageously ran huge climate models on their PCs. CPDN has been a huge success. There's no more worthwhile scientific goal than investigating the fate of Earth, and CPDN has made critical contributions to this investigation. CPDN inspired BOINC; when I read Myles Allen's original (1999) paper it got me very excited, and I immediately contacted him, wanting to get involved. CPDN's unique requirements had a big impact on BOINC's design. Carl Christensen, who for several years did the heavy lifting of getting CPDN working and keeping [it] going, has also contributed greatly to BOINC, and more recently so has Tolu Aina. I'm extremely proud to have worked with these guys and the rest of the CPDN group. Congratulations all around!!'*

BOINC has enabled industry to utilise a technology to create huge computing resources (exemplified by the World Community Grid) without the high cost of procuring conventional supercomputing facilities. It has also enabled the public to play a crucial role in solving challenging scientific problems by allowing them to sign up as volunteers to run models on their machines with results often viewable through graphical interfaces and by joining communities of volunteers through the use of portals and websites. The BOINC website details uses of the technology for an international array of projects, of which the following are examples:

GPUgrid.net is a distributed computing infrastructure devoted to biomedical research. Thanks to the contribution of volunteers, GPUGRID scientists can perform molecular simulations to understand the function of proteins in health and disease. In 2012, a team of scientists using GPUGRID announced that they had made crucial steps in simulating the AIDS virus maturation for the first time. Using computational techniques, researchers have shown how a protein responsible for the maturation of the virus releases itself to initiate infection. This event is at the root of the whole maturation process and if HIV protease can be stopped while it is still becoming mature, then the virus particle as a whole cannot become mature. Accessing the nascent structures of HIV protease provides a novel and critical target for the development of ARVs in the fight against HIV/AIDS [F]. GPUGrid currently has 1,994 active users with BOINC installed on their machines providing compute cycles to simulate cancer, AIDS and neural disorders.

LHC@home is a platform for volunteers to help physicists develop and exploit particle accelerators such as CERN's Large Hadron Collider, and to compare theory with experiment in the search for new fundamental particles. By contributing spare processing capacity on their home and laptop computers, volunteers may run simulations of beam dynamics and particle collisions in the LHC's giant detectors [D]. This project currently has 13,794 active users with BOINC installed on their machines providing compute cycles.

The **Quake Catcher Network** (QCN) is a project that uses Internet-connected computers to do research and outreach in seismology. Users can participate by downloading and running a simulation program on their computers, and this project currently has 1,199 active users with BOINC installed on their machines providing compute cycles [D]. Laptops connect to the QCN over the Internet. The laptop monitors the data locally for new high-energy signals and only sends a single time and a single significance measurement for strong new signals. If the QCN server receives a series of these times and significance measurements all at once, then it is likely that an earthquake is happening. The website for this project shows the location and triggered results from seismic activity, including recent activity on the west coast of America. This real time monitoring of activity using PCs as the base for detecting movement enables the public to be engaged in the detection of earthquakes [G]. A version of QCN is currently being used in Taiwan to enable them to benefit from volunteer computing to detect seismic activity as conventional earthquake monitoring techniques are known to be expensive.

Rosetta@home determines the 3-dimensional shapes of proteins in research that may ultimately lead to finding cures for some major human diseases. By running the Rosetta program on home computers, it speeds up and extends the research in ways that wouldn't be possible otherwise. It is also used to design new proteins to fight diseases such as HIV, Malaria, Cancer, and Alzheimer's. This project currently has 27,471 active users with BOINC installed on their machines providing compute cycles. The findings allow the Rosetta lab, run by David Baker at the University of Washington, to design proteins that do not exist in nature. Some new proteins could deactivate viruses such as the flu—as Dr. Baker's lab is trying to do for this year's H1N1 strain—by attaching to and smothering the sections of the pathogens that harm human cells. Dr. Baker has stated that the project's biggest recent breakthroughs have been in creating catalysts, which selectively speed up chemical reactions and which regulate almost every biological process. One catalyst in development, for instance, is an enzyme that could slice apart genes in female mosquitoes, potentially preventing malaria transmission without using toxic chemicals.

It is estimated that the total CPU time donated to run models using BOINC is 1,016,099,474,571 seconds. This is equal to 32,220 years of CPU time (equivalent to running a 32,220 core machine full-time for one year). The examples above are indicative of how BOINC and 'citizen science' have enabled the public to engage with science and to assist with the scientific breakthroughs achievable by using knowledge and modelling to drive experimental activity, thus reducing experimental costs.

Such is the definition of REF impact that while some of the *scientific* results of BOINC projects are admissible, such as the contribution of CPDN to the global warming debate, others might be classed as being purely of academic interest. It is, however, clear that *public engagement with science*, the thing most directly enabled by the UoA's research, is covered by the definition.

5. Sources to corroborate the impact (indicative maximum of 10 references)

[A] *Website for the Climate Prediction project: http://www.climateprediction.net.*

[B] *Website detailing uses and publications for BOINC:* http://boinc.berkeley.edu/

[C] Email from David Anderson (February 2013) stating the three main contributions by the Oxford CPDN team, held on file.

[D] Lopez-Perez, Juan Antonio. Distributed computing and farm management with application to the search for heavy gauge bosons using the ATLAS experiment at the LHC. PhD thesis, January 2008.
Provides details of the results from using volunteer computing for the LHC.

[E] Cochran, E.S., J.F. Lawrence, C. Christensen, and R. Jakka, The Quake-Catcher Network: Citizen science expanding seismic horizons, *Seismological Research Letters*, 80, 26-30, 2009.
Provides details of results from using volunteer computing for the quake catcher activity.

[F] S. K. Sadiq, F. Noé, and G. De Fabritiis, Kinetic characterization of the critical step in HIV-1 protease maturation, PNAS, published online November 26, 2012.
This publication describes the work using BOINC on HIV research.

[G] *The Quakecatcher website shows results from the remote sensors installed on users machines:* http://qcn.stanford.edu/qcn-map.