Managing Stakeholder Requirements in High Performance Computing Procurement

John Rooksby[1], Mark Westcombe[2], Mike Pidd[3], Ian Sommerville[1]

[1]Department of Computing, Lancaster University
[2]Attivation
[3]Department of Management, Lancaster University

High Performance Computing (HPC) facilities are provided in the UK at a national level. These facilities are amongst the best in the world, supporting world class research across a spectrum of disciplines. EPSRC has led the procurement of the two current facilities: CSAR (Computing Services for Academic Research) which was implemented in 1998, and HPC*x* which was implemented in 2001. In this paper we consider the issues involved in HPC procurement and present recommendations for managing stakeholder requirements in future procurements of national HPC facilities. These recommendations are organised as a process, of which the top level is presented here. The process takes an iterative, 'spiral' approach affording rounds of eliciting, balancing and validating requirements. Previous procurements have been successful, but with the increasing costs of HPC, an increasing effort must be made to ensure that this remains so. Our focus here is on how stakeholders, particularly within the academic scientific community, can be better involved in the decision making processes and ultimately how procurement strategy can rigorously meet the demands of the potential users.

1 Introduction

High Performance Computing (HPC[1]) facilities in the UK have acted as a focal point and catalyst for high quality research in science [1]. These facilities are amongst the best in the world; HPC*x* entered the list of the world's top five hundred high performance computers [2] in 2002 at number 9, and is currently ranked at number 12. An older facility, CSAR, is now ranked at number 120.

HPC rankings such as the top five hundred list are useful for the analysis of the industry and of available facilities [3]. However, comparison of high performance computers is difficult because they are designed for different purposes and placing their performance on a linear scale does not acknowledge their individual strengths. The list is based around a particular benchmark that does not well reflect scientific applications. The scientific focus of HPC*x* means that the ranking system is in fact hostile to it.

When procuring a new HPC facility, the aim is not simply to buy a faster machine to the ones previous, but to buy one best suited to the codes to be run on it. Decisions about the particular qualities of a new machine are difficult, particularly when there are a wide variety of potential codes from a large number of users. The machine must suit multiple stakeholders, come in within a limited budget and prove value for money. In this paper, we discuss the issues involved in procurement of HPC facilities for the UK research community, paying close attention to how stakeholder requirements can be met. We propose a requirements process that is geared to eliciting and reconciling the differing views of a wide range of HPC stakeholders.

2 Aim of the Project

The aim of our work is to determine and analyse the challenges facing procurement with specific emphasis on the process of gathering requirements and negotiating conflicts between them. We seek to design a practical framework to assist in responding to these challenges effectively in future HPC projects. In consultation with a range of stakeholders, a structured representation of the challenges facing the next HPC procurement is being produced. We are designing and documenting a flexible methodology to facilitate the elicitation of HPC requirements and needs. The

[1] We use the term HPC in a broad sense, where terms such as supercomputing, or high end computing are also sometimes used.

methodology must address the issues relevant to a range of stakeholders, be usable by the procurers, be justifiable (to stakeholders and reviewers), fit with other processes used in procurement, and be applicable to other HPC procurement exercises.

The project is to be delivered in autumn 2003. This paper presents an overview of the issues faced and a top level of the methodology. The project progress to date includes ten stakeholder consultations and a series of reviews.

3 Background

Previous procurements led by EPSRC include CSAR (Computer Services for Academic Research) in 1998 and HPC*x* in 2002. CSAR is a national high performance service run on behalf of the research councils by Computation for Science (CfS) at the University of Manchester. CSAR high performance computers currently include a 512 processor origin3800 called 'Green' and a 816 processor Cray T3E-1200E called 'Turing'. Turing is to be decommissioned in December 2003 and a new 256 Itanium2 processor SGI Atix called 'Newton' is to be provided. HPC*x* is a service provided by a consortium of the University of Edinburgh, CCLRC (the Council for the Central Laboratory of the Research Council) and IBM. The HPC*x* service provides a 1280 processor Power4 IBM cluster.

These procurements have been seen as considerable successes, achieving the aim of supplying a service to support world leading research. However, looking more closely at the recent procurement of HPC*x*, despite overall success, it is apparent that there was still room for improvement. The lessons learned show there were problems with time slippages, some of which were necessary to ensure competition was maintained, and others which could have been avoided. With hindsight it is also possible to say that the procurement happened too soon, missing a leap forward in technology. Stakeholder involvement was also problematic, with regional seminars and user questionnaires being limited in their usefulness and more, continuous involvement being needed. A more structured approach could also have been taken to user requirements, and a better approach to the developing and refining of requirements could have been taken. The HPC market place could have been better understood and a full survey could have had led to a better linkage of requirements and budget and allowed EPSRC to encourage particular suppliers to bid and possibly identify potential partnerships.

Internationally, the ways in which procurements take place differ significantly. Many of the top HPC facilities in America regularly procure new machines, and have strong relationships with a particular vendor and a large input into how the technologies are developed. EPSRC are in a different position, where physical location of a facility is not predefined and European competition laws limit the ways in which relationships can be built with individual vendors. These factors have a positive effect in putting EPSRC in a strong position to negotiate on prices, but does mean that those with strong technological understanding are often not core to the procurement team.

Internationally, there have recently been efforts into developing so called 'science driven architectures'. Until recently, the HPC top five hundred league table [2] was dominated by machines designed for weapons simulation and weather forecasting. These machines were not specialised for scientific research and devoted only small parts of their time to it. The UK facilities were not the very fastest, but provided what was probably some of the best levels of service to researchers. The arrival of the Japanese Earth Simulator in late 2001 caused a great shake up. The Earth Simulator was at the time more powerful than the top twelve American machines put together, and still continues to massively dominate high performance computing. Moreover, the Earth Simulator was optimised for scientific research. The American response, in what some have dubbed 'computnik' recognising echoes of the space race, has been to concentrate on building 'science driven' architectures that can compete with Japan [4]. The cost of procuring the Earth Simulator, and of running it are massive and advocates of HPC in America are going to lengths to obtain comparable budgets. The stakes have been raised greatly, and we assume it is unlikely that the UK can compete at a cash for cash basis. HPC is also developing rapidly in Europe (outside the UK) and new competitiveness is coming from countries such as China. It must be said that countries are not always directly competing, and UK researchers do (potentially) have access to machines such as the Earth Simulator. To keep UK science at the forefront we need not assume that the UK needs a bigger machine than other countries, but we must look carefully at how scientists are best supported and we must develop a sound strategy to enable that support.

4 Requirements in HPC Procurement

Like many large system procurements, the complexity, scale and time frame of HPC procurement is considerable. During the years between project initiation and a running HPC installation, needs and requirements evolve. Project management staff, the priorities of users and funding bodies, the relevant user needs and the available technologies can and do change. In addition, some domain specific challenges face HPC. These need to be comprehensively analysed and described in order that future HPC procurement projects can be put on a stable footing. In particular the process of negotiating different, conflicting needs for HPC should be addressed.

In this section we will discuss the central factors relevant to HPC procurement. We will introduce these factors individually (4.1) and then examine their interrelationships (4.2).

4.1 Factors

The procurement of an HPC facility involves the elicitation and balancing of many requirements. These requirements relate to a number of factors, the core of which are described here. Most of these factors are closely interrelated and their classification negotiable, which will be considered in the next section. Here the factors described are classified as stakeholders, the market, benchmarks, science, technology, and budget.

Stakeholders

In any large technology project there will be a large number of stakeholders with differing interests and involvement. The term stakeholder is used broadly in technology projects to refer to anyone who should have some direct or indirect influence on requirements [5]. It is good practice to allow anyone who sees them self as a stakeholder to be a stakeholder, but they are often not quite so forthcoming and the difficult problem is how to enrol and effectively engage with stakeholders. HPC procurement is a large project, in terms of cost, timescale, and the people it will affect. The types of stakeholder are far ranging, including tax-payers, researchers, engineers, employees and many others. The procurer of HPC facilities must actively recognise different types of stakeholder and actively engage with them. The key stakeholders to requirements are the actual or potential users of HPC facilities, and it is these that we consider of central importance to the requirements process. It is important that stakeholders understand and as far as possible agree the benefits of a facility to be procured, and that support for the project is gained.

The Market

HPC has a limited number of vendors, and a tender for a new facility will predominantly attract bids from the 'usual suspects'. However vendors do disappear and new ones enter the market. It is also likely that there are vendors that are not within these usual suspects who are capable of supplying the technology, and could be encouraged to do so.

HPC at this level is not a simple off the shelf procurement, but requires huge custom built machines. As such, it is difficult to know exactly what the market can offer at what cost. The is plenty of evidence to suggest Moore's law (stating that processor capacity will double around every 18 months) is relevant to HPC. This might imply that a four times more powerful machine could be procured at the same cost to one procured three years previously. However, three factors muddy this rule: firstly that the increases in processor capability are not smooth increments but happen sporadically (lessons learned from HPC*x* show that entering the market before a leap forward in technology can happen, unfortunately it is probably only possible to recognise these leaps with hindsight); secondly that the demands from HPC by users change in the nature of calculations needed; and thirdly that increases in demand for processor power often outstrip the increases they can supply.

To predict the market is therefore not a sound strategy, but to negotiate and investigate is. It is desirable to build relationships with vendors, but not to the extent that competition is harmed. Market intelligence is not only about knowing what is available, but being able to get the most suitable technology at the best prices.

Benchmarks

To evaluate performance of different machines, benchmarks are used. Benchmarks are a useful way for deciding between machines, but they must be chosen and managed carefully. LINPACK is the standard benchmark in HPC, but is becoming less relevant to modern scientific computation as the types of calculation it favours are not those favoured by scientists. More relevant is to call for potential users to supply their own code for use as benchmarks. This means machines can be judged in terms of the calculations they are required to do. However, care must be taken that codes are well designed and representative. Lessons learned from HPC*x* showed that the codes were not always portable and vendors are often only able to return incomplete results. There are also political factors: selecting a particular user's code for use as a benchmark can potentially alienate others. The third type of benchmark is low level, where key performance parameters are assessed. A suite of benchmarks must be carefully chosen and managed. Care must be taken that the benchmarks are representative, and that vendors do not tinker with them to improve results.

Science

'Science driven architecture' is becoming a catchphrase in HPC. The idea is that the demands of science drive the creation and use of HPC facilities. It is certainly true that some architectures are more suited to scientific applications, but there are many types of scientific application in HPC and no single architecture perfect for them all. Facilities should also be accessible for other types of research, such as finance or quantitative analysis.

Turning the idea of science driven architecture on its head, it can also be a potential benefit for HPC to drive science (and other types of research): new technologies can afford new types of research, leading scientists in new directions. At a basic level, it is a positive step to encourage researchers who have not really considered HPC to think how applications could benefit their work.

Technology

We have discussed the issue that different technologies suit different applications and thus different stakeholders will often have different requirements. We must also add that HPC is run as a service, and so decisions are also made about how the technology is run, including how jobs are queued and prioritised. Related to this, decisions must also be made as to whether, and how, to include Grid integration. As a service, upgrades to technology must also be considered and planned.

Budget

The cost to HPC relationship is complex. It is never clear how a particular set of requirements will translate into cost, and prices quoted by vendors are also often negotiable. One thing for sure however is that costs are increasing. Although cost is falling relative to processor power, the demands for power are increasing at a greater rate.

With the spiralling cost of HPC it is tempting to look at partnerships. Possible partners are those already involved in HPC, for example bodies involved in weapons research or weather forecasting. Finding a partner would be a serious boost to finding the budget, but would also cause serious problems in the provision of service to researchers. Simply having to share time on a computer would reduce resources that are already in high demand. Partnering with military organisations would mean that the machine would have to be rebooted between use by different partners so that there is no gaining, accidentally or otherwise, of sensitive data by someone running a job after the partner. A partnership may bring in a large amount of money, but trade offs will be made.

4.2 Interrelationships between Factors

The factors given in the last section do not present an exhaustive list, but represent those we see as central. The categories we have used are in many ways arbitrary: the factors are not units fitting into discrete categories, but are an interconnected web of issues. The technological issue of different hardware being optimal for different codes is a central issue to almost every concern in this paper. As another example, stakeholders (considering the key stakeholders as users) and science are strongly connected, the users being the people who do the science. The factors must be addressed then, but not individually.

The viewpoint taken will effect the understanding of these factors. For example, a budget centred viewpoint might lead to viewing user needs as secondary to market prices, or a technology centred viewpoint might lead to computation speed being prioritised over specific stakeholder needs. We advocate a user centred viewpoint, seeing user requirements for technology as being the central resource of requirements and representation of the needs of research. The central task then is to balance these needs within the constraints of the market and budget that act upon them.

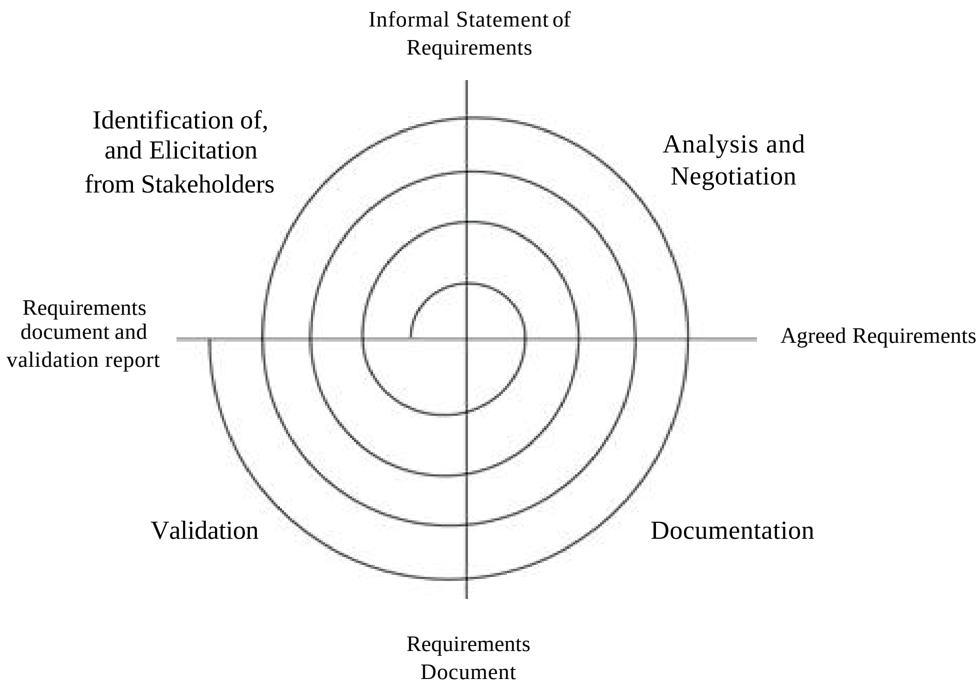

Figure1: A Top Level Spiral Process to Manage Requirements in Procurement of an HPC Facility

5 The Top Level Process

The aim of this project, as discussed in section 2, is to create a requirements process for use in procurement of an HPC facility. We have looked at lessons learned from the previous procurement, at the international context, and have outlined the factors we recognise that the process must address. In this section we discuss how the requirements process should be created and present a top-level view of this process.

5.1 Issues to Address

There are a number of issues the requirements process must overcome and account for if it is to be successful. The process must strike the right balance between being lightweight so as to be usable by the procurers, and being rigorous so as to be justifiable to stakeholders and reviewers. It should be possible to elicit requirements at different levels of detail and to tailor the process to the resources available for the requirements exercise; in particular it must be possible to maintain an optimum speed. The process should also correlate with the other processes involved in procurement, such as the creation of a business case. It is also desirable for the process to be generic.

Key aspects of the process will be the elicitation of requirements, and the balancing and deciding of requirements. To support requirements elicitation and balancing, there are a number of issues to be addressed that are general to the majority of large technology projects, and issues more specific to HPC. The issues surrounding requirements elicitation are how to recognise stakeholders and elicit high quality requirements from them. The issues surrounding balance involve how to judge and prioritise different requirements, and how to recognise and define core issues and assumptions, how to plan contingencies and how to maintain a memory of those requirements. Integral to both should be the assessment of risk. More specific issues to HPC revolve around the factors discussed in section 4 of this paper.

5.2 A Spiral Type Process

We believe that a ‘spiral’ type process is most suitable for this situation. The spiral process encourages multiple iterations of the same steps, starting rapidly and ending with longer more detailed iterations. As such it is a better informant of the creative design process where people do not operate in a top down manner, but start with a set of requirements and go about refining them. The spiral process allows requirements at different levels of detail to be captured and integrated during each round of the spiral. The spiral is also adaptable to limited and changing resources: because all phases may be accommodated in each round of the spiral, with planning it is possible to close the activity after any round. The process should ideally continue, for as many rounds as resources will allow, until results are deemed good enough for external review.

For the procurement of a national HPC facility there are a great number of uncertainties and past experience has shown that to act quickly and dynamically can be beneficial and sometimes essential to success. The process presented here is designed to offer a greater amount of planning and control to the procurement process, but not to inhibit flexibility. The spiral allows for the balancing and refinements involved in eliciting and managing stakeholder requirements, negotiating technologies with vendors, and putting all the necessary components in place to proceed. Simply, the process we present recommends working and reworking requirements until the necessary stakeholders are satisfied enough for procurement to proceed.

In this paper, we are discussing the top level of the requirements process. The generic nature of the spiral at this top level means that different fine grain approaches can be used beneath it. While we are recommending a specific approach based upon stakeholder viewpoints, alternative approaches can be used with the spiral if preferred. The proposed spiral model, presented in figure 1, contains four phases and states the output of each. The first phase is 'identification of, and elicitation from stakeholders' with the output as an 'informal statement of requirements'. The second phase is 'analysis and negotiation' with the output as 'agreed requirements'. The third phase is 'documentation' with the output as a 'requirements document'. The final phase is 'validation' with the output as a 'requirements document and validation report'. These phases are now described. We seek in this paper to give an overview of the process, and therefore the following descriptions do not spell out the fine detail.

Identification and elicitation of requirements from stakeholders

The first stage entails the identification of stakeholders and the eliciting of requirements from them. The users and potential users of HPC are probably the most important stakeholders to consult in terms of requirements. They can be identified by looking at who has previously used HPC, but care should be taken to also think about who new users may be. It is likely that asking recognised potential users who other potential users might be, will be a valuable method of identification. It is also suitable to advertise for stakeholders to come forward. The eliciting of requirements can be done by asking what the stakeholder wants. This can be done by one on one interviews, in group workshops or by questionnaires. It is likely that a combination of these will be most appropriate. Which ever method is used, it must be well designed so as to elicit the best possible requirements in the most manageable form.

As the process progresses, this stage will be repeated. This reflects that not all stakeholders can be consulted at the same time, that new stakeholders will become apparent during the process and that stakeholders may have to be returned to.

Analysis and negotiation

This second stage involves the balancing of requirements. The requirements must be analysed and understood, before they can be negotiated and prioritised. It is likely that some requirements will be unrealistic and that others may clash with each other. It is essential to build a coherent requirements model, but also to allow for some flexibility. Negotiation and analysis of requirements could involve the stakeholders, or might be in a closed room. Involving stakeholders will allow greater involvement and accountability but risks deadlock and will almost certainly increase the process timescale. A possibility is that stakeholders are involved in some but not all of the iterations of the process.

As the process progresses, this stage will be repeated. This reflects new inputs from stakeholders and the fact that requirements will have to be modified in conversation with the market and budget.

Documentation

The third phase involves the documentation of requirements. Documents will have been produced as part of the previous stages but documents must be produced to explicitly state decisions that can then be used for validation. Iterations of this phase will be necessary reflecting iterations of previous stages.

Validation

The final stage of each cycle will concentrate on validating requirements. The procurers will have done some validation of requirements at the analysis and negotiation stage, but here they can be checked for feasibility given cost, technology, time and whether they meet the needs for keeping UK research at the forefront. Factors such as cost are fairly hard constraints although vendor prices and procurer budgets sometimes hold some flexibility. Other factors are more qualitative and demand careful consideration.

Validations will start as internal validations, where tough questions should be asked by, ideally, a series of stakeholders with different viewpoints. It is better that problems arise at this point than at external validations (such as the 'Gateway Review' by the Office of Government Commerce). Iterations of this stage, and the spiral itself, should continue until procurers are ready to face external review.

6 Further Work

This paper presents an overview of a process for managing requirements in high performance computing procurement. The paper does not seek to spell out the fine grain detail, which is still work in progress. An aim of presenting this paper is to gain feedback from the e-science community that can be fed back in to the project. The final process model is to be delivered in Autumn 2003.

7 Conclusion

In this paper, we have discussed procurement of UK HPC facilities, concentrating on the management of stakeholder requirements. A top level view of a process to make explicit these requirements has been presented. The process is a spiral, going through several iterations of each stage until completion. This is a realistic model for a situation where all the stakeholders may not be known at first, and some back and forth will occur between the requirements, the marketplace and the budget. The process given here is to give a flavour of the final contribution, and not to spell out the detail. Procurements of UK HPC facilities have to gain stakeholder approval and support, and this paper addresses ways of achieving that.

References

[1] Anon (2002) High End Computing: Strategic issues for the UK. A Report from the High Performance Computing Strategy Committee.

[2] www.top500.org

[3] Dongerra JJ, Meuer HW, Simon HD, Strohmaier E (1997) Changing Technologies of HPC. Future Generation Computer Systems 12: 461-474

[4] McCurdy CW, Stevens R, Simon H, Kramer W, Bailey D, Johnston W, Catlett C, Lusk R, Morgan T, Meza J, Banda M, Leighton J, Hules J (2002) Creating Science Driven Computer Architecture: A New Path to Scientific Leadership. http://www.nersc.gov/news/ArchDevProposal.5.01.pdf

[5] Sommerville I (1995) Software Engineering (Fifth Edition). Wokingham: Addison-Wesley