



## Co-researching as a Driver for Technological Innovation: Computing and Cultural Heritage

David Arnold

### Abstract

A recurring question in defining publicly funded research programmes is “what research should be funded publicly and what should be funded through the private sector?”. The way this question is usually answered has placed emphasis on seeking to ensure that research that has a payback in purely economic terms is funded (at least in part) by those most likely to receive the economic benefit, but evaluating “who benefits?” and “how?” is not always easy to deduce from the research questions posed.

**Key Words:** research and innovation; computing; cultural heritage

### 1. Introduction

Computer scientists seek to develop methods that are applicable to classes of problems which inevitably means that solutions invented in one field can be applied to problems that have the similar structure in other fields. The recognition of this process of generalisation has been fundamental to the development of computer science as a discipline over the past 70 years. Formal recognition can be seen in seminal works such as Dahl, Dijkstra, & Hoare (1972) and Wirth (1976) and involves understanding that computation is achieved through developing structured representations of knowledge (facts, information, objects, etc.) and methods to analyse and manipulate those representations. The title of Wirth’s book (“Algorithms + Data Structures = Programs”)

summarises this fundamental underpinning of the discipline that remains just as applicable 40 years later.

There are however equally inevitable consequences for those who would seek to quantify the beneficiaries of individual pieces of computing research. If a method is generalisable to a class of problems then the potential beneficiaries of the research are all those stakeholders who have problems in that class and anyone who might provide solutions to those stakeholders.

The decision-making for research funders is therefore somewhat complex—if they wish to ensure that the beneficiaries contribute to funding the research then there needs to be a way of evaluating the degree of benefit. To take the example of Finite Element Analysis (or Finite Element Methods or FEM), this set of techniques can easily be traced back at least 50 years, as a way of improving the analysis of complex engineering designs (e.g., Strang & Fix, 1973), but many would argue that conceptually similar techniques predated the invention of the digital computer with the publication of Lewis Fry Richardson's book on numerical techniques in weather prediction. Richardson (1922) proposed that carrying out a large number of similar computations for a grid of cells that represented the earth's atmosphere could be used to trace the influence of each cell on the next and simulate the weather's behaviour into the future.

The generalised technique of FEM is now applied in a wide variety of fields to simulate the behaviour of complex systems that can be represented as a set of elements and the set of behaviours through which one cell influences the next, along with appropriate ways of handling conditions at the boundary of the simulated system. Applications (and hence beneficiaries) of research in these methods include not only weather forecasting and civil engineering (engineering structures like dams, bridges, buildings, etc.), but the design of complex parts and structures for manufactured products (cars, ships, aircraft, etc.), fluid dynamics in areas such as aeronautics and environmental pollution spread, to thermodynamics for simulating heat loss and environmental control systems. Several areas might be interrelated—for example the influence of high winds on engineering structures would involve fluid dynamics and structural engineering properties.

The solution of problems formulated to use FEM requires the solution of sets of equations relating the properties in each cell to the effect on its neighbours.

There are therefore three rather different aspects to set up an FEM experiment—the set of parameters and equations, the definition of the grid of cells in 2D, 3D or higher dimensionality, and the development of the computing engine capable of solving the set of equations that relates them. For any individual piece of finite element analysis, each of these three areas has different stakeholders, requires different sets of knowledge and hence potentially requires different research projects to acquire that knowledge.

For the computer scientists researching the computational method, the acid test might appear to be the capacity to solve the individual computational problem that this experiment poses and this is certainly the starting point for the research. However the greater success relies in generalising the approach so that the same computing solutions can be reapplied to new problems and potentially to new fields. The development of new generalised solutions is an essential part of creating new computational services and from that providing value and creating wealth.

The engineers and scientists undertaking research to inform the understanding of the relationships between the parameters of the system will typically be experimenting with simplifications of the complex systems to isolate and understand the relationships between different elements. Such experiments might address material properties in isolation from the particular situation in which those materials are used and then use that knowledge to formulate the set of equations that describe the behaviour of a more complex assembly of materials in real objects. Success is then achieved when the set of equations can be shown empirically to simulate the actual behaviour under controlled conditions—probably using scale models for larger structures or prototype manufactured parts for example. An additional level of success is however when the knowledge of (e.g.) materials that is gained in one field can be successfully replied in a different field.

For those working in the application field—bridge designers or those forecasting the weather—success is clearly measured in those fields—bridges designed that can be shown to be economic to build and proven to be still fit for purpose before they are built, or improved weather predictions.

The clear implication is that the first two groups need the third not only for there to be meaningful criteria against which to measure success but also to give confidence to generalising their results for use in a different field or

even on different problems in the same field. But this in itself generates a catch-22 situation. Genuine research is exploring the unknown and hence any research, in setting out to discover new knowledge, runs the risk of failing to discover that knowledge. Why would busy professionals already engaged in delivering results within their chosen field go out on a limb and divert time and resources to participate in risky experiments and help those working in a different field to develop and perfect techniques and knowledge that will actually benefit later adopters and other disciplines?

These tensions were well recognised in Stokes' (1997) seminal work on "use-inspired basic research" which re-arranged the traditional conceptual model of research ranging from pure to applied into a  $2 \times 2$  matrix (Figure 1) depending upon the researcher's motivation to seek fundamental understanding, whilst having very specific application-related motivation for seeking that understanding. Pasteur's quadrant was that combination of specific applications in mind, with high intent to address them through an underpinning understanding of the fundamental science and engineering and was seen as the area most worthy of public research funding. It was no accident that this diagram was used as the frontispiece for the UK government's policy paper on investment in scientific research for the period 2004–2014.

In the 1970s a colleague researching in computer aided design (CAD) undertook a lecture tour in China and on his return reported that the most challenging question he had faced was "Given that labour is cheap in China, what are the benefits of CAD in comparison to using a 100 engineers with slide-rules." This incident highlights the inertia that accompanies the establishment of professional standards and best practice. Whilst this inertia might

Fig. 1: Stokes' (1997) classification of types of research.

Research inspired by:		Considerations of use?	
		No	Yes
Quest for fundamental understanding	Yes	Pure basic research (Bohr)	Use-inspired basic research (Pasteur)
	No		Pure applied research (Edison)

be considered an impediment to obvious and inevitable development of better practice, it should also be appreciated as the legitimate resistance to taking risks by adopting the unknown or unproven. Society places heavy emphasis on the needs for thorough drug trials before authorising their use routinely, which is sometimes seen by those awaiting new treatments as an irritating block. There are clearly judgements to be made.

Yet returning briefly to the use of CAD systems in China—it is clear that they have been adopted since those early conversations—for example in the design and build of the Birds’ Nest or Aquacube—both arenas built for the 2008 Olympics. At some point it became obvious that there were benefits in engaging with CAD products, but Chinese practitioners did not apparently see the point in engaging with the new technologies when they were in the early stages of development. Local circumstances made it the right professional decision for them to delay investment.

Demonstrating that getting engaged in multi-disciplinary research is worthwhile has always been a challenge for computer scientists. In the early days of the discipline applications-related professionals had not yet experienced the revolution that computing has brought in almost all aspects and fields of professional work and hence there was little track record to give credibility to the claims that engaging with computing would provide payback and be useful in the longer term. Even now the evidence of predicting where and for whom payback from research will occur seems inconclusive, with those deciding on whether to invest resources in developing new technologies being wary that they risk losing the investment whilst those that become involved once a technology is better established can benefit from the improvements.

## **2. Computer Science and the “Toy Problem”**

Moore’s Law is a well-established and well-known observation showing near exponential changes in aspects of computing technologies over time. Statements such as “the power of computing doubles every 18 months, whilst the cost halves over the same period” are commonplace and have been applied to: speed; memory capacity on a system; network capacity; number of CPUs/PCs globally; amount of information on the internet and others. Most of the applications of this formulation have been true for remarkably

long periods of time, particularly when the functionality is separated from the specific technology—"the memory capacity on a disc" assumes that discs are the technology used to deliver memory and this is likely to hit limits that would not necessarily affect "the memory on a system" for example.

The unchallengeable truth is however that the capacities of computing systems on which any specific computing research is undertaken, has been and, predictably, will be significantly less than the systems on which the results will be deployed once the research is completed. The history of computing is therefore one in which specific pieces of research have developed, for example, new data structures whose deployment on realistic ("real world") problems would imply the need for greater capacity than was available on the systems that were current when the research started. Similar observations can be made about the computation speed implied to make algorithms run in finite time with real data and the research is also undertaken in a climate where the user expectations of what is possible (and hence what "should" be provided) will also change over the period of the research.

The consequence of this observation is that in retrospect the original research has often been conducted on what would seem inappropriately small scale tasks that can be seen as having little relevance to "current" challenges in the "real world". To offset this potential perception computer scientists have increasingly sought access to "real world" data that can be used to test specific algorithms or data structures (as indicated above computing research almost always involves an inevitable link between the methods and expectations of the data structures on which they are useful). Getting access to this data inevitably and increasingly requires negotiation about how the data will be used (including ethical considerations), how its integrity will be maintained so as not to present it as representing something that it wasn't designed to represent, as well as how providing access to the data might afford some benefits to the owners of the data.

In practice Wirth's formulation ("Algorithms + Data Structures = Programs") should probably now be reformulated as "Algorithm + Data Structures + Data = Systems". As data increases in volume so does: the investment required to generate it; the value of the data assets; and the computational challenges in analysing it. In the age of "Big Data" access to suitable volumes of appropriate data will increasingly become a pre-requisite for meaningful research into innovative techniques and technologies.

In short this sort of collaboration requires all parties to share very different perspectives on the data: what it represents and how it can and will be used. The more complex the collaboration the more perspectives are likely to be involved and the more investment each party is likely to need to make in adapting their own perspective to those of the others in the team. This can have very practical implications—for example in the knowledge that is formulated within a data structure and the investment needed by the users and other parties to organise their data into a format that allows the computer scientists to undertake the research.

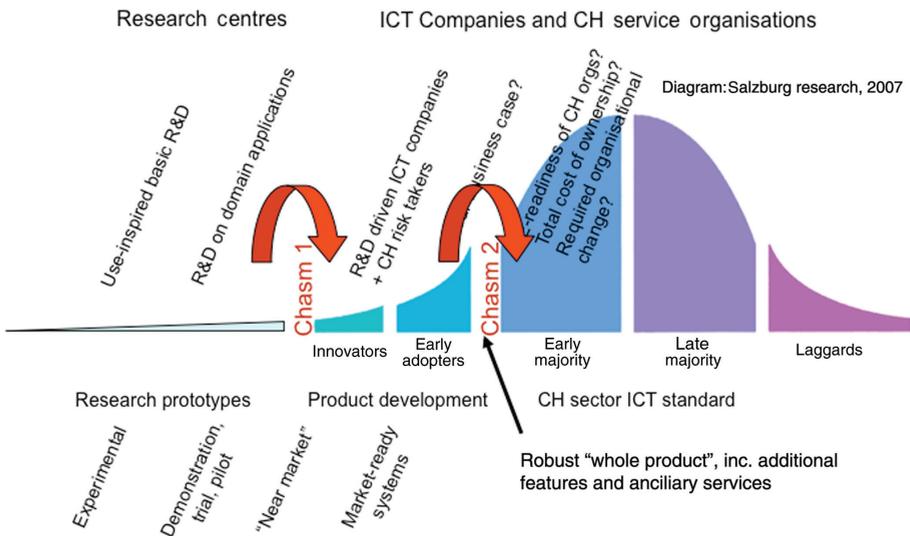
Going back to the FEM case to allow computational experiments with detailed meshes would require the bridge designers to be prepared to allow their designs to be represented as finite elements and the relationships between them, which could represent a significant amount of additional work. If that work is required when the technique is experimental and the FEM approach was not an expected part of the design process, then this might well represent extra and potentially unjustified cost to the project.

However for any innovation that fundamentally affects the way in which a profession can operate there have to be similar processes that can be traced through a number of stages. Normally the first stage would be for the computer scientists to develop a demonstration of the concept using a small-scale simplified exemplar of the problem domain. The challenge is to make the exemplar convincing of the potential to scale the approach up to the scale required in practice. That will typically involve orders of magnitude more data and increased complexity—for example to cope with special cases that the problem domain can require in a minority of situations. The innovative method won't be adopted routinely until it can cope with the normal challenges it would encounter in deployment and exposure to representative data, drawn from known case studies, is the only way of providing believability and hence to achieving adoption of the innovation.

### **3. Technological Innovation in the Cultural Heritage Sector**

There are many models of the adoption of technological innovation in differing sectors, but one aspect that seems to transcend sectors is that adoption and market penetration are human rather than technological challenges. Figure 2 is drawn from Arnold and Geser (2008). This is presented with no

Fig. 2: The assimilation of ICT research and innovation [from Arnold and Geser (2008)].



scale on the horizontal axis since the elapsed time to embed innovation will depend on the speed of development of the professions that contribute to that sector. Whilst any enterprise will have some office functions that are innovated through technology in this chapter sector innovation is considered as innovation which is linked to the essential characteristics of the sector, rather than to the processes that would be typical of any organisation. The salary system has the same characteristics and requirements virtually independent of the sector in which an organisation sits—an employee is an employee whether the service is plumbing or a museum. What is of interest in this section is any area of operations that bring processes or data characteristics that are fundamental to the sector's services.

If Cultural Heritage is the “significance of the evidence of the past in the present” then it has interesting characteristics in that significance in the present effectively means that everyone has their own, individual cultural heritage which changes over time and in that understanding the cultural significance of the evidence requires access to the evidence. Knowledge of past cultures requires research on the evidence; the representation of the knowledge in suitable formats to allow analysis of the information and analysis

methods to work on the data—directly comparable to the “Algorithm + data structures + data = systems”.

The Cultural Heritage sector is justifiably measured (some would say “slow”) in its response to technological innovations for a number of good reasons. Most obviously conservators are sceptical of new technologies because there is much evidence that they are transitory—new storage media are continually replacing the old—and the effort required to maintain digital resources is normally on top of the resources used in conserving the physical cultural assets. For the benefits to outweigh the additional costs is therefore a different challenge from other fields. A design department moving from paper-based systems to CAD might take the view that in the future the maintenance of digital records would replace the need to maintain libraries of large format plans for example and hence that adopting CAD could be expected to release resources elsewhere in the workflow. In practice the value-added has been more in the ability to be responsive to clients than to a saving on the costs of individual designs.

For cultural heritage, however, these savings through using digital technologies to effect fundamental workflow changes are harder to see. There can never be an assumption that adopting the technologies will lead to less of a requirement to maintain the original artefacts, so perceived benefits must lie in elsewhere—and they do, at least in the minds of computer scientists!

From the Curators perspective (and that of other professionals) the benefits might lie in (for example):

1. Improved techniques for the detection of cost-saving maintenance interventions
2. Detection of relationships between dispersed pieces of evidence of the past, that would assist curators in their research; in the authorship of exhibition narratives; in communication with the public.
3. Establishing provenance; detecting stolen artefacts and improving security for artworks.

There are, largely manual, processes for all of these in existing practices which are more or less effective and the evidence that large-scale digitisation will provide a fundamental improvement is scant. In all projections to make improvements digitally will require substantial investments of time

and money to establish the infrastructure of data resources which the computational methods will require to deliver results.

#### **4. Computing, Cultural Heritage and Digital Humanities**

Taking the above definition of Cultural Heritage as the “significance of the evidence of the past in the present” then inevitably, as with our health and educational history, we all have our own frame of reference as a combination of the society within which we were born and our experiences since birth. In addition, knowledge in these fields is a combination of the factual (or near factual) and interpretations of context—for example how an artefact might have been used within a society which may be largely undocumented, or of the cultural or religious significance of an event or ceremony where the heritage would be classed as intangible, even if there is tangible documentation of instances of such ceremonies.

To the computer scientist this combination of Cultural Heritage “knowledge” as an uncertain mix of informed opinions, culturally-based inherited narratives and linguistically-based interpretations presents a class of data that may be shared across the arts and humanities but would be less apparent in scientific research or even in the social sciences. In this context the significance of any piece of information is cultural. For example the interpretation of the symbolic significance of colour depends upon the culture in which it is found. Red may be used to represent “danger” or “fire” in many western societies but represents other values elsewhere. It is associated with “weddings” and “good luck” in many oriental cultures and with “mourning” in parts of Africa. Interpreting texts and imagery that contain cultural symbolism through the use of colours therefore requires cultural awareness, that may or may not be geographically-based, or based within diaspora or, with the advent of the internet, within professional or social virtual communities.

Representing the “knowledge” embodied in cultural information therefore requires the computing system, with all the inherent qualities of deterministic behaviour and black and white facts (represented with the decisiveness of the binary world of noughts and ones), to deal with uncertainty, incomplete data and, potentially conflicting, opinions. The uncertainty is compounded by the nature of cultural contexts—much of the evidence of the past is in

commemoration of religious belief or heroism in conflict, both areas encouraging very strong and potentially conflicting views of the past and of its current significance. Cultural Heritage thus represents a wide range of information contexts that are very different from the more traditional realms of scientific computing or business processing.

For computing research cultural heritage becomes the test-bed for computing solutions that seek to represent uncertainty, conflicting truths, and linguistically-based and culturally-based interpretation of texts, imagery and objects. In fact this is the world of the “Internet of Things”—objects out of context in a sea of loosely linked and often conflicting descriptions. Cultural heritage knowledge is a prime example of the uncertainty of knowledge and relates very closely to the world of multiple certainties of knowledge—the increasingly worrying world of absolute certainty in specific beliefs, even where the wider secular society might often consider them “radical” or “extremist”.

Given that the volume of data on the internet appears to obey a variant of Moore’s Law and double every year, the challenging fact is that half the current information on the internet wasn’t there a year ago and no human was able to know what was there then. If social media and the internet are the jungle in which threats to humanity hide, then we need much better ways of understanding the cultures and communities in which cultural opinions are expressed—as well as much better ways of encouraging mutual respect and co-existence on our increasingly shrinking planet.

## 5. Conclusions

In this chapter I have argued the absolute necessity of close working between applications communities and the world of Computing Science. In the era of “Big Data” this cooperation is designed to ensure the relevance and practicality of computing solutions aimed at a scale of problem that no human can undertake during their life time. In practical terms this means developing solutions that no human can ever completely validate so that technologies must both propose a solution and analyse whether the results proposed are in fact the appropriate ones (optimum, correct or whatever other measures of success are appropriate).

As soon as we begin to deal with data that embodies the value sets and cultural expression of different ethnic groups this situation is compounded by orders of magnitude. We are told that there are something like 7000 languages that are currently endangered. Humankind is inevitably and inexorably losing our ability to understand past societies ways of expressing their cultural values and for endangered languages this means present minority communities' values.

It is my hope and my belief that in co-developing techniques that recognise and address different cultural values, in the relatively safe environment of cultural heritage, we can also better understand the ways in which different value sets meet and either, in the best of worlds, evolve through mutual respect and understanding, or in less accommodating circumstances instil the potential for civil conflict, extremism and terrorism.

Those who have supported the early stages of this journey, by empowering computer scientists and cultural professionals to work together on ICT research that targets cultural heritage knowledge, are from my perspective making a fundamental contribution to the future of humanity. This may sound grandiose, but for those of us who embarked on careers of research in Computing Science in the heady days of the 60s when everything was possible, it is no more an unlikely aspiration of science fiction than the idea of running one's life on a laptop, via mobile communications from a moving train in a different country—as I am doing whilst writing this text!

I started this chapter by highlighting the challenges of justifying the investment in publicly-funded IST research and pointing out the challenges in quantifying the beneficiaries of any research ahead of time. At the micro level this remains really difficult, but since any search for new knowledge might fail, consideration at the micro level is rarely the right level to justify investment in a research programme. Given a range of individual projects within a programme it has to be understood that whilst the individual projects may succeed or fail the sum total of knowledge continues to grow even from analysing failure.

I hope I have argued successfully that investment in understanding the types of knowledge represented in Cultural Heritage stands a very realistic chance of a payback not only to the Heritage sector but to humanity in general. The world feels a brighter place whilst those who control and administer research

funds continue to recognise Cultural Heritage as a test-bed for computer scientists to develop new solutions that respect cultural diversity, opinion and historic context. That effort can only succeed with the commitment of experts who understand the cultural context for the data and artefacts being considered and the complexities of interpretation when working with that data. The journey will be, predictably, long but there can be few more worthy causes or valuable prizes to pursue.

## References

- Arnold, D.B., & Geser, G. (2008). *Research agenda for the applications of ICT to cultural heritage*. EPOCH publication. Budapest: Archaeolingua.
- Dahl, O.J., Dijkstra, E.W., & Hoare, C.A.R. (1972). *Structured programming*. London: Academic Press Ltd.
- Richardson, L.F. (1922). *Weather prediction by numerical process* (2nd ed - 2007). Cambridge: Cambridge University Press.
- Stokes, D.E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, D.C.: Brookings Institution Press (ISBN: 0815781776).
- Strang, G., & Fix, G.J. (1973). *An analysis of the finite element method*. Englewood Cliffs, N.J.: Prentice Hall (ISBN 0-13-032946-0).
- Wirth, N. (1976). *Algorithms + Data structures = Programs*. Prentice-Hall Series in Automatic Computation. Englewood Cliffs, N.J.: Prentice Hall (ISBN: 0130224189).

## Note

---

The author, prof. David Arnold, passed away on October 25, 2016.