# Research, Relativity and Relevance: Can Universal Truths Answer Local Questions

By Derek Law, University of Strathclyde

*Abstract: It is a commonplace that the internet has led to a globalisation of informatics and that this has had beneficial effects in terms of standards and interoperability. However this necessary harmonisation has also led to a growing understanding that this positive trend has an in-built assumption that "one size fits all". The paper explores the importance of local and national research in addressing global issues and the appropriateness of local solutions and applications. It concludes that federal and collegial solutions are to be preferred to imperial solutions.*

## Introduction

Professor Nick Moore of City University famously divided Europe into three groups when it came to developing electronic information services[1]: The Nimble North; the Messy Middle and the Sleepy South. The Nimble North consisted of small countries with resources limited by size which used their size to leverage those limited resources through sharing - the Nordic countries being the best example; the messy middle, including the UK, have rich resources but use a blunderbuss approach hoping that if you scatter enough resource something positive will happen, albeit in unplanned and not cost-effective ways; the sleepy south found it difficult to overcome issues of infrastructure and collaboration, and although producing some interesting and individual applications had no real sense of national policy or drive. But most of all what Moore demonstrated was the ability of small countries to innovate and develop, rather than just to accept that larger richer countries would deliver finished global products and services. It is important then to explore national versus global agendas; to consider whether user behaviour and needs are uniform; to discover how far open access and open source provide relevant content

and applications; and to determine how far local research can provide global input.

**Global Issues**

At a recent JISC/NSF (Joint Information Systems Committee/National Science Foundation) meeting Herbert Garcia-Molina[2] of Stanford University, outlined the global issues which in his opinion still needed to be addressed by research. First came the remaining physical barriers. As society moves ever faster to mobile access and ambient computing, the availability and usability and security of wireless remains an issue. At the same time the convergence of mobile phones, PDA's and laptops still remains to be worked out into an optimal set of tools.

There remain substantial economic concerns. Business models for digital libraries have not really been developed. Much work has been done on the economics of digital preservation, on the economics of scholarly publishing and on esoteric if somewhat retro initiatives such as Brewster Kahle's universal library downloaded, printed and bound for micropayments[3]. These economic concerns are also reflected in continuing issues over intellectual property structures. Although some international law applies here, much law is country specific. In general however the continuing ability of the entertainment industry to extend copyright periods poses significant difficulties for open scholarly discourse.

Information Loss was his next concern. The whole issue of institutional repositories and national repositories and the standards they should adopt remains unclear. There is no quality assurance procedure to define trusted repositories. And we have already lost much of the electronic information of the last fifty years. Whether on eighty column punch card, paper tape or proprietary hardware or software, it is already too late to recover much

data. For example, in the United Kingdom an oral history project was launched in recent years to recover the history of computer science once it dawned on that discipline that its pioneers were retiring and that there was no real paper trail to show how the science had developed.

Information overload is a much described but little resolved phenomenon. The tales of Google searches which produce hundreds of thousands of hits are legion but the solutions remain to be discovered. Value filtering requires much more work to become truly usable. As Garcia-Molina puts it "We have too much stuff, but not the right stuff[4]". Embedded within this issue is that of decontextualisation; the way in which the Internet search engines remove context and therefore relevance.

His last major issue to be addressed was that of service heterogeneity and the need always to ensure interoperability.

He also saw need for further R&D work – and work on the research process - in a number of areas. The most difficult remained that of scalability. Many local projects were workable at the local level but simply did not scale up to be appropriate or even usable at national and international level. The swell of local developments also led to issues of consistency and interoperability. Dissemination of research remained a key problem. With so much research being undertaken throughout the world it was difficult to discover what was novel and what was being repeated. But a major concern remained archiving including such basic theoretical issues as the appropriate unit for archiving. There are issues surrounding the hidden web[5] and ensuring access to information and further issues related to collection building on specific topics. Nor is enough known about user behaviour and how far it might be culturally variable

**Local Issues**

The list of research activity as seen by Garcia-Molina from Stanford has been listed at some length in order to compare it with the concerns manifested at this conference. Some conference papers clearly address the global topics. There have been studies of user behaviour and content dissemination; there have been papers on contextualisation and on ICT strategies. But there has also been a persistent, and in my view welcome, analysis of local issues. There have been papers on the needs of semi-literate groups, including parliamentarians; on issues facing a multilingual society; on indigenous knowledge systems (an area increasingly seen as universally important); on health issues; on social equity and rural development.

At the launch of the Strathclyde Institute for e-Systems a paper by Nigel Kay[6] proposed a national e-agenda for Scotland. To some extent his concerns reflect Garcia-Molina's concerns: to achieve pervasive access; to address the skills shortage; to support innovation from R&D through Proof of Concept to new products and services creation; to tackle the legal issues; to accelerate uptake, 'e-incubation'; to effect synergies with e-government and e-learning; to change mindsets and culture; to use the technology itself to help. However in such a general list Kay could hope that the small country would again have the advantage of "A fully collaborative, joined up effort, done at speed."

At the same time and in the same department, Dennis Nicholson has been undertaking applied research to build a national digital library infrastructure ranging from a cultural portal to a national union catalogue,

to OAI harvesting and collection level description for federated searching[7]. Much of this work involves the development of standards often in partnership with organizations such as OCLC. This reflects the ability of small countries to take a holistic view of national information policy requirements, as predicted by Moore.

**Some local initiatives**

Although not research in the narrow sense, there is a number of local initiatives which have had global impact. Kahle's Internet Bookmobile[8] may not be the way ahead, but his reputation and enthusiasm have ensured that it has been given a broad international welcome from India to Egypt. He has taken 100,000 out of copyright books and put together an easy to use mobile technology which allows the books to be cheaply downloaded and bound on demand. The scale of the project is allowing him to work with other groups to digitise collections in other languages such as Arabic. Next he plans to turn to out-of-print but in copyright material.

The Greenstone Digital Library Software[9] is a suite of software for building and distributing digital library collections. It is produced by the New Zealand Digital Library Project at the University of Waikato and developed and distributed in cooperation with UNESCO as open source software and has some thirty language interfaces. It is specifically intended to empower users to build and then share their own digital libraries.

The Directory of Open Access Journals[10] is an Open Society Institute funded project enthusiastically developed by Lund University in Sweden. It began as a project simply to list the steadily growing number of peer reviewed and/or quality controlled electronic journals. These now number

1250. Under growing user demand the project has now expanded to include article level coverage and already over 320 journals are covered. In developing this service, the project has had to address and resolve a number of research issues.

**Local Content, Open Access and Peer Pressure**

The Open Access movement has had as its doughtiest champions Paul Ginsparg, Stevan Harnad and Peter Suber. At first presenting extreme arguments on the fringe of academia, their foresight is now seen as mainstream. A string of declarations and reviews over the last eighteen months has supported the position of open access publishing and institutional repositories. An original Budapest Open Access Initiative of 2002 was followed by the Bethesda Statement of June 2003, the Berlin Declaration of October 2003, the Wellcome Trust Position Statement of October 2003; the House of Commons Select Committee Report of August 2004, the US Congressional Committee on the National Institutes of Health proposal of September 2004 and most recently the Scottish Declaration on Open Access of October 2004. This of course shows the power of individual persistence, but it has also led to a flourishing research industry on the nature and impact of scholarly communication.

More generally work has been slow to begin on electronic collection building. Yet the sheer volume of electronic materials is growing rapidly[11] and requires thought to be given to policy on collection building as well as the technology and practices which will allow it to happen. Present academic research builds on the collections of the past: it then behoves us to build collections for the future. It has been claimed by Pinfield in the context of Open Archives that "the biggest challenge is getting content."

[12] In the context of e-collection building the challenge is perhaps that of building collections of coherent content.

**Building research collections for the future**

In the past building collections was relatively straightforward. The papers of distinguished academics were collected from their studies after death; manuscripts and books were purchased from rare book and manuscript dealers, contacts were cultivated in the hope of donations. The vary stability of the paper record allowed patience and often multiple opportunities to determine where papers gravitated to. Nor were the collections only paper, but sometimes physical objects. The University of Hull famously added Philip Larkin's lawnmower to its library collections[13].

The issues are much more daunting when it comes to electronic materials and largely revolve around media formats and preservation. But we do precious little in terms of what would have constituted collections in the past. E-drafts of documents and paper; e-correspondence between researchers; personal files on a pc rather than in a filing cabinet; the electronic equivalent of lab books are all falling through the net. While we feed off the collections of the past we generally fail to reflect on how the born digital collections of the future will look. Nor do we consider how the material will be held. The absence of agreed repository standards must be a major cause of concern. Ironically, as in so many things one can see a potential solution in looking back to the experience of the past to develop thinking on the future. The Maori tradition is an oral one and they have developed a quite specific set of criteria to guide the selection of the keepers of that oral tradition[14]:

1. Receive the information with utmost accuracy
2. Store the information with integrity beyond doubt

3. Retrieve the information without amendment
4. Apply appropriate judgement in the use of the information
5. Pass the information on appropriately

These seem a perfect guide to the preservation requirements of tomorrow's e-collections.

## A typology of collections

Thinking on collections has most fruitfully taken place within the context of the Digital Library Federation http://www.diglib.org/dlfhomepage.htm) and they have produced interesting work for example on strategies for developing sustainable and scalable digital library collections. Greenstein[15] proposes four types of collection

- local digitization projects that produce surrogates for analogue information objects;
- data creation projects that produce information resources that have no analogue equivalent and are in this respect "born digital";
- the selection of existing third-party data resources for inclusion in a collection either through their outright acquisition or by acquiring access under some licensing arrangement; and
- the development of Internet gateways comprising locally maintained pages or databases of web-links to third-party networked information

This typology allows an exploration of the nature and extent of what is, should be and could be made available.

### *Digitised Surrogate Resources*

It is a commonplace that not all existing collections will be digitised. Scale, copyright and value are argued to make such conversion implausible. It is certainly the case that at present we tend to see projects delivering

selected subsets of collections rather than the whole. Digitised resources can be further sub-categorised beyond Greenstein's single overarching category, because the motives for digitisation are very varied. Improved access, preservation, aggregation of scattered material, and more are all reasons for creating digital collections, as the following examples of the sub-categories show.

***Surrogates of rare items: the British Library***

One of the best examples of this is the British Library's Treasures collection (http://www.bl.uk/collections/treasures.html), where rare treasures are made more accessible to the public (and indeed to scholars), using software to mimic page turning. This collection contains such miscellaneous material as Magna Carta, the Lindisfarne Gospels, the Gutenberg Bible and the notebooks of Leonardo Da Vinci. What these great documents have in common is their rarity and their public prominence. The e-collection acts as a surrogate to allow these iconic treasures to be open to all.

***Surrogates for whole or part collections: The Springburn Virtual Library***

During the summer of 2000 it became apparent that the Springburn Community Museum in Glasgow faced closure for financial reasons. Although the collections were to be transferred to the Mitchell Library in Glasgow, this much loved and popular local resource would be separated from its community. A project was put in place to ensure that the museum's rich collection of local photographs would still be accessible to the local public over the internet. Funding was secured to digitise a representative selection of materials from the collections and to lay the foundations for Springburn Virtual Museum. Images were chosen to convey the social and economic history of Springburn, notably

community and tenement life and the important local railway industry. http://gdl.cdlr.strath.ac.uk/springburn/. As a result a community threatened with the loss of a resource has had at least a subset of it made more accessible to all.

***Digitised surrogate collections assembled from multiple repositories: the Valley of the Shadow***

The much admired Valley of the Shadow Project focuses in great detail on the experience of two communities, one Northern and one Southern, through the American Civil War, as an exemplar to give an understanding of the experience of the nation as a whole. It consists of a hypermedia archive of sources for Augusta County, Virginia, and Franklin County, Pennsylvania. A rich variety of materials has been assembled - newspapers, letters, diaries, photographs, maps, church records, population census, agricultural census, and military records. It encourages users to interact with materials rather than simply access them. http://www.iath.virginia.edu/vshadow2/

A collection with a quite different focus and ambition is the Great Britain Historical GIS Project (www.gbhgis.org), which aims to have systematic information on the history of every locality in Britain, using everything from Ordnance Survey maps to Victorian gazetteers and Defoe's *A Journey through the Whole Island of Britain*. It can be searched using postcodes and aims to allow everyone to access information relevant to their own area.

***Collections assembled specifically to be digitised***

The Aspect project (http://gdl.cdlr.strath.ac.uk/aspect/) was set up to create a digital archive of the ephemera - leaflets, flyers, postcards,

newsletters - produced by candidates and political parties for the first Scottish parliamentary election in May 1999. The archive is based on the collection of election ephemera held by the Andersonian Library at the University of Strathclyde, which is acknowledged to be an important and unique record of a key event in Scottish history. The creation of a digital archive will significantly improve the accessibility and usability of the information contained within the collection whilst conserving the original materials, which may be subject to deterioration through loss and damage. Thus a collection being built for use by future researchers is being made immediately available, using digitisation as a deliberate strategy in acquisition.

**Born Digital Resources**

The number and scale of these is growing from scholarly journals to new fiction, from datasets and satellite images to digital video and computer generated graphics. Many are being preserved. But examples of born digital collections are rare. It is arguable that these remain individual items rather than forming a coherently built collection. Perhaps the nearest to this is the various collections of learning objects being assembled in many universities. For example, Boezerooy[16] gives a comprehensive overview of the Australian experience which demonstrates that these exist but are not always created with library advice or assistance or indeed even with long term preservation in mind.

**Third Party Data Sources**

Most of the research activity on licensed data probably fals into the category of information policy. In the UK, JISC began its work of building the Distributed National Electronic Resource in 1990[17] and now has a hugely rich collection of resources licensed to the community. That

consortial licensing model has been widely followed. The International Coalition of Library Consortia (ICOLC) first met in 1997 and has grown to be a self-help group of some 150 consortia from all over the world, which considers issues of common concern, principally in the context of higher education and research. Without necessarily supporting it however, ICOLC (http://www.library.yale.edu/consortia/) in effect works within the present pattern of scholarly communication to make material as available as possible.

The electronic environment offers up new and as yet unexplored models of data acquisition, whether for a single institution or in consortia. The intention expressed by Singapore in its seminal planning for the Intelligent Island[18] (Chun Wei Choo, 1997) is to create an information entrepôt and hub for the region. It is easy to build on this concept to develop the concept of information arbitrage[19], the notion of buying and selling information around the world, taking advantage of time shift to buy data cheaply at off-peak times when little used in a country. Similar thinking has informed the development of 24x7 reference services.

**Mirroring and caching**

This is a somewhat neglected subset of third party licensing. An excellent early example of this is the Visible Human dataset. This was originally constructed in the United States with the support of the National Library of Medicine (NLM). It contains images of a 39-year old convicted murderer who, prior to his execution, donated his corpse to medical science. The dataset was subsequently expanded with the addition of the images of a female at greater resolution than for the male. The bodies have been "sliced" to create the images. NLM did not want to see copies of the dataset mounted outside the USA, quite properly fearing that issues such

as version control and quality assurance were not sufficiently settled in the mid-1990s to give comfort of proper data management. For the UK this proved a problem since this wonderful resource was heavily used in medical teaching and consumed great quantities of bandwidth as images were slowly downloaded. Mirroring was the obvious solution. Discussions began with NLM and after protracted discussions the final sticking point (according to folklore!) was the need for guarantees on what would happen to the data if the host institution disappeared. At that point, in 1997, JISC accepted an offer from the University of Glasgow to act as the host (http://vhp.gla.ac.uk/), not least on the grounds that it had already existed for half a century before Columbus sailed the ocean blue. Whether or not the tale is true, it does demonstrate that mirroring can be just as complicated an exercise as licensing commercial data. Certainly in the UK, as network charging begins to influence decisions, it seems reasonable to expect a greater interest in mirroring s a method of reducing traffic as much as improving accessibility.

The same is true of caching data. This is one of the black arts of computing but does have a significant impact on costs, traffic and availability. The UK National Cache has been studied in depth[20] in terms of performance and value for money and this is very informative in indicating the impact that an institutional caching strategy might have.

One major issue appears not to have been addressed so far. There is a bland assumption that there is an almost infinite supply of bandwidth and that issues of access and slow to load pages will disappear: that view is not necessarily shared by all. At the same time there is an equally unthinking assumption that resources are either good or bad. But there is a more sophisticated but so far neglected approach which asks whether

the Pareto Principle (the 80:20 rule) might also apply to on-line resources. It is typically assumed that access should be given to the best or most complete or most authoritative material, but these terms are never explored or defined. Networked environments add the complication of accessibility in a quite novel way. For example, in many parts of Europe, the quality of connectivity to the United States drops dramatically after the golden hours of the European morning, once American users wake up and begin to log on. So is a similar or smaller resource (but just as accurate) available twenty-four hours a day to be preferred to a larger resource effectively available for, say, only two-thirds of the day? There is a need for a much more sophisticated appraisal of all the factors surrounding internet gateway access than has perhaps been the case thus far.

## Conclusion

Thus we can see that there is a range of global and local research needs to address. In some cases local initiatives, or even personal crusades can contribute to the wider global agenda, while local research by definition addresses local needs. The model which best serves research even on a global topic such as information science is a distributed collegial one. It is clear that initiatives from all over the world are contributing to the common pool of knowledge. Perhaps the question posed in the title would then be better reversed to claim that local answers may well resolve universal questions. But there must be one absolutely critical rider to that point, and it is that local research must be disseminated widely and actively. We need to make sure that we not only invent search engines, but use them to share our findings.

---

## References

1. Moore, Nick. The International Framework of Information Policies in Elkin, J & Law, D. Managing Information in Higher Education Institutions. Milton Keynes: Open University, 2000
2. Garcia-Molina, Hector. Unpublished paper given at the JISC/NSF Project Review Meeting. Stanford: University of Stanford, 2003.
3. Kaushik, Radhika Spreading The Digital Word *Extreme Tech* 29 April 2003. http://www.extremetech.com/article2/0,3973,1047454,00.asp
4 . Garcia-Molina, Hector. *Op.cit.*
5 . Kautz, Henry *et al.* The Hidden Web. *AI Magazine* 18 (1997) pp27-36
6 . Kay, Nigel. Challenges and Opportunities. Unpublished presentation (2003)
7 . Nicholson, D et al. Whole Environment Research on distributed and collaborative digital and non-digital networked libraries in Scotland *Bibliotheck Forschung und Praxis* 26, (2002) pp113-123
8 . Cisler, Steve. Letter from San Francisco: the Internet Bookmobile. *First Monday* Issue 7. http://www.firstmonday.dk/issue7_10/cisler/
9 . Details may be found at http://www.greenstone.org/
10 . Details may be found at http://www.doaj.org/
11 . Five year information format trends. OCLC, March 2003 www.oclc.org/info/trends/
12 . Pinfield, Stephen (2003) Open Archives and UK Institutions *D-Lib Magazine* 9(3), 2003
13 . Guardian (2002) *The Education Guardian* Thursday May 9, 2002 [news item] http://education.guardian.co.uk/higher/humanities/story/0,9850,712877,00.html
14 . Winiata, Whatarangi (2002) Ka purea e ngā a hau a Tāwhirimātea: Ngā Wharepukapuka o Ngā Tau Ruamano. Keynote address, LIANZA Conference, Wellington, 2002. http://www.confer.co.nz/lianza2002/PDFS/Whatarangi%20Winiata.pdf
15 . Greenstein, Dan. Strategies for developing sustainable and scalable digital library collections. http://www.diglib.org/collections/collstrat.htm
16 . Boezerooy, Petra Keeping up with our neighbours: ICT developments in Australian Higher Education LTSN Generic Centre, [n.p.,2003]
17 . Law, D. The development of a national policy for dataset provision in the UK: a historical perspective. *Journal of Information Networking* 1(1994) pp103-116
18 . Chun Wei Choo. IT2000: Singapore's Vision of an Intelligent Island. Book chapter in Intelligent Environments, edited by Peter Droege Amsterdam, North-Holland, 1997
19 . Law, D. The Library in the Market: information arbitrage as the new face of an old service. *IATUL Proceedings* Vol 11 (New Series) 2001. Delft: Delft University of Technology, 2002.
20 . Sparks, Michael *et al.* An Initial Statistical Analysis of the Performance of the UK National JANET Cache http://wwwcache.ja.net/papers/initial_analysis/