National Heritage Science Forum

Keynote speech at the NHSF Members & Friends event by Alastair McCapra, Chairman, 16 December 2015

Cash for Data?

Why are we talking about data for cash?

Because science is expensive to do, and cash is often short. We always need to be thinking about whether we are working our assets as hard as possible to ensure we have the resources to develop the science in ways which benefit society.

Although the Chancellor's recent autumn statement did not signal the radical reduction in science funding which had been predicted in some quarters, the overall shortage of science funding in the UK is widely recognised. Last month, just before the autumn statement, Professor Brian Cox spoke out about under-investment in science, and the Commons Science and Technology Committee warned that Britain was in danger of losing its status as a 'science superpower' as investment dropped well below the average for developed nations – just 1.7% of GDP compared with an average of 2.4% for OECD countries.[1]

So perhaps monetising our data can help generate new funds for heritage science?

If we have a look first at the world beyond heritage science research, it is clear that cash for data is a concept which is gaining widespread acceptance.

For example, there are commercial organisations which invite you to sign up and sell your personal data to them - including firms such as Handshake and Datacoup.

There are also examples of selling data for the advancement of science. For example, in 1998 there was an agreement between the government of Iceland and a firm called DeCode genetics. The purpose was to allow DeCode to aggregate health records of the entire Icelandic population and match it with genealogical records and genetic data on a single database. Under Iceland's Health Sector Database Act, DeCode agreed to pay for the cost of the database, to pay the government a

---

[1] http://www.telegraph.co.uk/news/science/science-news/11982679/Britain-risks-losing-science-superpower-status-due-to-lack-of-cash-warn-MPs.html?utm_medium=Social&utm_source=Twitter&utm_campaign=Echobox&utm_term=Autofeed#link_time=1447028198

licence fee of 70m kronur (6m euro) per year to operate the database plus 6% of any profit it made from using the database. Expected annual income to the Icelandic state from the deal was 6-12m euro per year.[2]

So do examples like these suggest that there are good prospects for generating income through the sale of heritage science data?

Not necessarily. From the point of view of the seller, cash for personal data may represent a good deal if people have nothing else to sell. However a data broker like Datacoup generally pays less than $10 a month for a person's personal data[3] and soon, so much data will be readily available through the internet of things anyway that it will become harder and harder for an individual to monetise their own data in any effective way.

Interestingly, from the point of view of the purchaser, data acquired in this way may be of limited value because it is partial, and some of the most valuable data will not readily be captured. There is a possible parallel here with the commercial sale of heritage data. Institutions in the most pressing financial circumstances will sell first; those in a more robust state or with more to sell will hold out for better terms or decline to sell at all, because they will be able to develop different kinds of partnership and make use of their data in proprietary ways on their own. The result - a partial, fragmentary data landscape with many of the most valuable pieces missing, will be of limited appeal to any commercial buyer.

Similarly, the deal between the government of Iceland and DeCode genetics did not go well. The law permitting the health sector database was later found to be unconstitutional and the company never built the controversial database it allowed for. Instead of scraping data from existing health records as planned, DeCode steadily built its own database of genetic data from scratch. DeCode never made a profit so there was never a windfall for the Icelandic people, but despite this it was bought by Amgen in 2012 for $415m. It has subsequently been sold on again, albeit for a lesser sum.[4] The government of Iceland got a minimal initial licence fee, but whatever value there might be in the intellectual property is now owned by a Chinese firm, Wuxi Pharmatech. Despite their personal data being collected and used in these commercial exchanges, the people of Iceland never received any further financial reward from the venture.[5] There are perhaps some interesting general points of reflection here on how apparently attractive deals can end up not quite as expected.

---

[2] 'Changing Properties of Property', Franz von Benda-Beckmann, Keebet von Benda-Beckmann, Melanie Wiber Berghahn Books, 2009 p.320
[3] http://www.theguardian.com/news/datablog/2014/apr/22/how-much-is-personal-data-worth
[4] https://en.wikipedia.org/wiki/DeCODE_genetics
[5] http://www.els.net/WileyCDA/ElsArticle/refId-a0005180.html

Coming closer to home now and thinking about the world of heritage - while the notion of commercialising data may be relatively new, the imperative to generate commercial income is far from new. It was in the 1980s that universities, museums, galleries and libraries first came under pressure to work in a more business-like way. Long before it was practical to consider monetising raw data, universities were becoming very focused on developing and managing their IP, and on creating spinoff ventures from their research activities. Museums and galleries became particularly focused on managing images of their collections.

In the 1990s there was a heady period between the advent of the internet and the dotcom bubble when Digital Rights Management was thought to be the magic pot of gold that would transform the fortunes of heritage institutions. In 1998 McGraw Hill published a book entitled 'Digital Property, Currency of the 21st Century.' However, fifteen years into the new millennium, most institutions are still searching for the elusive pot of gold. Digital property, as we now understand, does not generally work for research and heritage institutions in the same way it works for iTunes or Disney. Even the largest institutions find this digital property can be expensive to create and maintain as well as difficult to manage and monetise. Far from being the panacea imagined twenty years ago, digital rights management can often generate major problems of its own.

For example in 2006 the Smithsonian Institution planned an agreement with a commercial film distribution called Showtime to distribute the Smithsonian's film collection. The deal was exclusive in nature and although the Smithsonian is a public, not for profit institution, it would have allowed Showtime to charge a fee for the film it made available. This provoked protests from film makers, heritage professionals, the public and politicians. Such was the outcry that the deal was never concluded, and the Smithsonian was required to appear before a Subcommittee of the House of Representatives to answer questions about how it was managing its public collections and expending public funds.' [6] In fact, in May 2006 a House of Representatives panel recommended cutting the Smithsonian's public funding by $20m to punish it for this deal. [7]

While universities have generally been successful in creating commercial value out of many of their knowledge assets over the last thirty years, this is less usually the case for museums, galleries and other institutions. The broad picture about the value of IP is revealed in the reports of the key funding bodies. Earlier this year a report from the Arts Council on the economic impact of museums in England listed 16 forms of income generated by English museums. It showed that in 2012-13 DCMS-funded museums made 25% of their income from trading – but this includes all trading,

[6] http://www.wipo.int/edocs/pubdocs/en/copyright/1001/wipo_pub_1001.pdf
[7] http://www.npr.org/templates/story/story.php?storyId=5402899

including ticket sales, catering, publications, events, filming, and venue hire, not just commercialisation of its IP.[8] Whatever share of that 25% of income generated through trading may derive from rights income, it is too small to split out on its own.

Equally, if we look at the AHRC's most recent Impact Report[9], there is no discussion about the economic value of IP generation or exploitation, except in a single table. This shows an interesting and unexplained anomaly in 2013-14 when the instances of IP creation and exploitation suddenly surged, but the recorded instances in the five previous years were not much better than zero.

Overall, the track record of heritage institutions commercialising their assets is mixed. The most successful areas of new business have been around those institutions' core roles in bringing visitors to exhibitions and publishing material about their collections. Finding viable markets for their IP beyond this can be difficult, complex and sometimes less rewarding than anticipated. As the example of the Smithsonian illustrates, most institutions need to find a balance between income generation and their mission to make their collections accessible to the public.

So perhaps this experience in the commercialisation of other assets does not hold out much promise for the future of trading in data. Indeed, the commercialisation of data could in some respect actually be damaging to heritage science. The problems heritage science faces include a general weakness in scale, depth and longitudinal consistency. If institutions lock down their data now on the promise of some potential future commercial yield, the effect on heritage science could be crippling. Instead of scaling up, heritage science will shrink back as each institution becomes ever more defensive about its assets.

The last thing I'd like to note about commercial use of data is that while there are certainly many pressures to generate commercial income, there are countervailing pressures in play as well. Among the most important of these is the EU Directive on the Re-use of Public Sector Information (2003/98/EC), generally known as the PSI Directive. This has gone through several forms, and the current form is Directive 2013/37/EU which came into force two years ago. This Directive encourages public sector bodies to make as much information available for re-use as possible, and to limit charging, where this is allowed, to the defraying of the direct costs of making the information available. Since 2013 content held by museums, libraries and archives falls within the scope of application as well.[10] The Directive was transposed into UK law earlier this year. The current orders

---

[8] http://www.artscouncil.org.uk/media/uploads/The_Economic_Impact_of_Museums_in_England-FINAL_REPORT.pdf

[9] http://www.ahrc.ac.uk/documents/project-reports-and-reviews/the-impact-of-ahrc-research/2013-2014/

[10] http://ec.europa.eu/digital-agenda/en/european-legislation-reuse-public-sector-information

allow an exemption for museums, libraries and archives. However the 2013 Directive has closed this loophole and the UK is due to amend its own Regulations to reflect this fairly shortly. Once this is done, the scope for making commercial income from data will be very severely restricted, and probably just not be a viable activity.

Learning Lessons from the Science and Heritage Programme

So if the prospects of cash for data are not very encouraging, what alternatives are there for the future of heritage science data? Well there are some valuable learning points from the Science and Heritage Research Programme which ended last year. One of those points is that the commercial imperatives may often not be paramount among parties seeking collaboration, even when one or more of them is a commercial entity. Another is that heritage science needs to scale up, become more intensively cross-disciplinary, and involve more people.

In their project '**Mind the Gap: Rigour and Relevance in Heritage Science Research**' Nancy Bell, Matija Strlič and their co-investigators reminded us that the motivations and the satisfactions of participants in collaborative heritage science research projects are not always directly commercial. Among their conclusions was this:

'The long-held belief that knowledge transfer will lead to innovation and business performance represents too simplistic a view. The opportunity for engagement across many disciplines and many users highlights other reasons for engaging e.g. adding another source of knowledge or connecting with others to enhance organisational performance. The notion that knowledge transfer is largely driven by commercial interests, as reported elsewhere, was not found in this study. Knowledge sharing, access to expertise and building relationships were considered key enablers of collaboration.'[11] In other words, commercial partners can often gain something valuable, even if intangible, from collaboration with non-profit, public institutions, and this can be a motivation, just as much as the prospect of immediate profit can be.

We also need to think of the wider vision for the field and what might enable us to realise it. In her closing address to the Science and Heritage Conference in 2013, Professor May Cassar said

'Before the Science and Heritage Programme, most research questions addressing issues of cultural heritage were… primarily focussed on an underlying technique or technology. Whilst such an

---

[11] Mind the Gap: Rigour and Relevance in Heritage Science Research, N. Bell, M. Strlič, K. Fuseki, P. Laurenson, A.. Thompson, C. Dillon 2014 p.6 http://www.nationalarchives.gov.uk/documents/mind-the-gap-report-jan-2014.pdf

individual approach did deliver some valuable research outcomes, it had huge weaknesses for the development of the hard and soft science skills that the sector needs. After all, it is very important for the applicability of the research, that researchers have a full understanding of the heritage sector and the many complex evolving issues of particular relevance to it. Furthermore, an individual approach offers no route for sustained or strategic engagement among institutions. The Science and Heritage Programme has been a game-changer in developing a highly skilled cross-disciplinary heritage science research community explicitly focussed on cultural heritage needs and experienced at working with a broad range of partners.' [12]

'My vision is that within the next decade, heritage science will become as ubiquitous a concept and a reality, as environmental science is today, yet broader in its embrace of subject disciplines…. Cultural heritage is community business. Without time and effort given freely by heritage volunteers, the effectiveness of conservation, education and promotion of cultural heritage would decline significantly. English Heritage estimated that 450,000 adult volunteers contributed to the heritage environment in the UK in 2010-11. Just as volunteers are the lifeblood of the heritage sector, so too can volunteers help to strengthen and spread heritage science.'

In 2006, before the House of Lords Science and Technology Committee, we argued that heritage science was too small and too easily fell into the cracks between established funding programmes. It needed to scale up, we said, to build a denser network of links between existing scientists, to engage more effectively with scientists in other fields and with the broader public. We argued, successfully, for a one-off dedicated research programme to prime the pump. Nearly a decade on, there has been significant progress, but there is much more to be done, particularly in terms of overcoming the limitations of an infrastructure which remains somewhat fragmented. Do we have at our disposal a means of overcoming these limitations, while opening the possibility of bringing in new funding?

Open Knowledge, Open Data, Open Science

If we think about how the world has changed since 2006, one of the most remarkable things to emerge is the open knowledge movement. This existed, to be sure, before the Science and Heritage Programme started, but it was in its infancy and it was still somewhat marginal. This movement was born, in part, out of a rejection of commercialisation – Wikipedia was founded in 2001 and its motto is 'it's free and it always will be' and the site will never accept advertising. There was a widespread

---

[12] http://www.ucl.ac.uk/silva/heritagescience/Final_event/May_Cassar_Closing_Address

revolt against the commercial publishing model for academic journals; against online pay-per-view; against complex regimes of licences and institutional barriers designed to protect IPR. New technology made it easier than ever for people to share what they knew, and the mainstream culture of the internet quickly coalesced around the belief that if it was online, everyone expected it to be free.

The Open Knowledge Foundation, as it was then known, was founded in 2004. In 2005, the US National Institutes of Health decided to make peer-reviewed, final manuscripts stemming from NIH-funded research available to the public free of charge on PubMed Central within 12 months after publication in a scientific journal. The 2005 policy strongly encouraged authors to post for public accessibility as soon as possible.

As the currents of public expectation shifted, professional bodies, which had previously reserved reading rights in their libraries for their own members, began publishing white papers, guidance notes and technical information for free. Many small journals ceased publishing. Universities went from posting brief taster seminars on their websites to delivering MOOCs to millions of learners around the world – many of them entirely free.

In 2008 Michael Heller published his book 'The Gridlock Economy: How Too Much Ownership Wrecks Markets' in which he introduced the world to the concept of 'The Tragedy of the AntiCommons'- a type of coordination breakdown, in which numerous rights-holders prevent each other from accessing and making use of a resource in order to protect their own stake in it, thereby frustrating what would be a socially desirable outcome. To promote those socially desirable outcomes and to prevent that tragedy from being enacted, Open Knowledge was soon joined by Open Data and Open Science - Figshare was set up in 2011; The Open Data Institute and Wikidata in 2012.

I don't recall that anyone who gave evidence to the Lords Science and Technology Committee in 2006 even mentioned open knowledge and open science. It did not seem to any of us, I think, to be a development which particularly relevant or useful, and we did not say anything about its potential impact on heritage science.

But since then, some other fields of scientific endeavour have embraced the open science mode with vigour. In astronomy, chemistry and the life sciences, data sharing is increasingly common; closer to home, open publication of data is very much the norm in archaeology.

What about Heritage Science? I'm going to suggest seven reasons why shared, open data will be a good idea for heritage science.

1. Shared data is safe data - and robust too as it is validated through sharing

2. Maximises collaboration

3. Maximises repurposing and re-use of datasets

4. Maximises public access and participation

5. Best way to rapidly scale up and generate lots of fundable activity

6. With declining funding, sharing data maximises the impact of every pound invested

7. Makes better science - A lot of heritage science is done on small datasets so may not be very sound – we need to scale it up but concatenating datasets from various sources (dirty data)

In other words, I believe that adopting the culture of open data will help us deliver the long term vision of heritage science which May set out in her speech last year – to scale up, to become more intensively cross-disciplinary, and to involve more people. It will also help secure new funding.

Science and Innovation Strategy

Last year the NHSF responded to a consultation on government proposals for long-term capital investment in science and research[13], and said

'The one output that is common to all heritage science research whether it is applied to the moveable or immoveable heritage is the generation of data. There needs to be an imperative for data reuse after its first use by the primary generators and heritage science could act as an exemplar through investment in capital infrastructure. It is inefficient and a potential net loss to the economy not to use this data to its full. Our proposal therefore has as its ultimate goal the effective use of data for the intellectual and commercial advancement of the UK.'[14]

'As a challenging area of applied science, with the appropriate level of investment in capital infrastructure, heritage science could be a vehicle for the delivery of two out of the eight Great Technologies in the Government's Industry Strategy, namely 'Big Data and Energy Efficient Computing' and 'Advanced Materials and Nanotechnology.' And because of the emphasis of this vision on public engagement, further investment could be drawn from the proposed 'Inspiring Science' Capital Investment Fund [as well as] Connecting Research and Communities.'

---

[13] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/321522/bis-14-757-consultation-on-proposals-for-long-term-capital-investment-in-science-and-research-v2.pdf
[14]

http://www.heritagescienceforum.org.uk/images/dynamicImages/file/SHP%20and%20NHSF_response%20to%20BIS%20consultation%20on%20Capital%20Infrastructure%20Investment.pdf

'The track record of researchers that the Science & Heritage Programme and the National Heritage Science Forum represent, in attracting leveraged funds and in kind support from collaboration is outstanding. For example, SHP-funded projects have together leveraged over £950,000 in additional funding. In the six months since the announcement of the CDT awards, SEAHA has raised £0.5M in hard cash and £6.9M in in kind support from heritage and industry partners.'

There is a wider context for the kind of approach suggested in that consultation response. Just six weeks ago, the UK was one of seventeen countries to sign the open data charter in Mexico City.[15] Our government is now talking in terms of 'Government as a Platform' (GaaP). Stephen Foreshew-Cain, the Executive Director of the Government Digital Service says that "we should value data as part of our national infrastructure."[16] The Cabinet Office has a Director of Open Data and Government Innovation, Paul Maltby.

We are now firmly committed to open government; but government is not alone in the sharing revolution. Commercial pressures are changing around us too – in recent weeks we have seen first Google and then Facebook open-source their artificial intelligence engines. Astra Zeneca and Sanofi have just agreed to swap more than 400,000 chemical compounds for medical research, reflecting an increasing acceptance of collaboration within their sector. Closer to home, the Centre for Doctoral Training in Science and Engineering in Arts Heritage and Archaeology has attracted investment from industry leaders such as Dyson and Philips Lighting who are attracted by the knowledge and transferrable skills they can gain through collaboration.

Earlier this year, NESTA and the Open Data Institute ran an open data challenge. An analysis by PWC of the value in that programme indicated that for every £1 invested in making data open, £10 of economic value was generated. Another study, of the US Landsat dataset, comprising satellite imagery of the Earth's surface, showed the annual economic benefit of it being made openly available was $2.19bn in 2011 alone.[17]Research conducted this year by the Open Data Institute identified 270

---

[15] http://opendatacharter.net/seventeen-governments-adopt-the-new-international-open-data-charter/
[16] http://central-government.governmentcomputing.com/news/paul-maltby-takes-up-uk-data-leadership-role-4679475

[17] https://medium.com/@ODIHQ/the-economic-impact-of-open-data-what-do-we-already-know-1a119c1958a0#.ze9y5itp2

companies, large and small, that use, produce or invest in open data in the UK, with a combined annual turnover of over £92bn.[18] Open data means big business.

Increasingly, as commercial firms rely on open data as the basis for developing new products and services, they are going to have to pay to keep the flow of data fresh and relevant, because without it they will be stifled. They will rely more on their scale and expertise to bring developments to market based on that data, and less on IP rights. In other words we may well be looking at a future in which large commercial players regard it as a key part of their pipeline of new products and services to maintain a flourishing ecosystem of open research and open data, creating a superabundance of raw material from which they can draw what they need.

We are approaching a time when government will not commit to spending on any programmes which do not openly publish their data. The EU Research Frameworks, the UK Research Councils, the Lottery Funds, and other grant-makers will also increasingly expect open data from institutions they support. In July this year, RCUK agreed a new Open Data Concordat, which says:

'The concordat thus takes as its starting axiom that making research data openly available for use by others is an inherent good with many benefits. Within this new paradigm, the following expectations will be established:

- Researchers will make their research data open and usable within a short-well-defined period, which may vary by subject and disciplinary area

- Employers of Researchers will seek to provide appropriate access to infrastructure systems and services to enable their researchers to make research data open and usable.

- Funders of Research will support open research data through the provision of appropriate resources as an acknowledged research cost.[19]

So we see a strong and well-established trend in the direction of open publishing and open research as a requirement for funding. The culture is changing rapidly all around us. Paywalls, firewalls, stonewalls, wallets and walled gardens are all out. Open is in.

---

[18] https://medium.com/@ODIHQ/the-economic-impact-of-open-data-what-do-we-already-know-1a119c1958a0#.ze9y5itp2

[19] http://www.rcuk.ac.uk/RCUK-prod/assets/documents/documents/ConcordatOpenResearchData.pdf

Within this environment, those which share their data will be international leaders, able to form rewarding and remunerative partnerships. Those who hoard their data or who fail to meet the technical requirements of releasing it in usable formats will increasingly fall behind.

Conclusion: Cash for Data?

So my conclusion is that there is cash for data, but that it is much less likely, in future, to come as the result of a commercial transaction than was expected in the past. It is much more likely to come in the form of funding from mainstream sources which will increasingly expect, and indeed require, that data is shared publicly in some readily usable way from programmes they are supporting. If institutions are not equipped to share their data they will gradually find it harder to secure research funding.

The transition from the traditional business models to the new one of being funded and supported by both public and commercial partners to share data may not be straightforward, and where existing commercial activities are successfully generating income for their institutions these should certainly continue for as long as they bring value. But I think the overall direction of travel is clear, and we must be ready to step forward. We have now, so to speak, a chance to make a virtue of necessity. There are seven good reasons for heritage science to embrace the open data revolution wholeheartedly - unlocking cash is an important part of the picture, but it is only a part.

This new world of open data represents a huge opportunity for heritage science, and I believe the NHSF should help its members embrace the change and move forward with confidence.

That is what I wanted to say about cash for data. Thank you for your attention. I'd also like to thank a few other people as well, before I conclude:

- The outgoing co-chairs, Nancy Bell and May Cassar, for their leadership of the Forum, and for the encouragement they've given me to put myself forward for the role of chair.
- The trustees for choosing me as their new Chair
- I am very honoured indeed to be chosen for this role. Since giving evidence to the original House of Lords enquiry into Heritage Science nearly a decade ago it has been really wonderful to watch first how the Science and Heritage programme took shape, and then

how the Forum came to life. This is a tribute to the dedication, persistence and commitment of many people, not least of course of May Cassar.

- I very much look forward to working with you all.