

Living Books About Life

# Digitize Me Visualize Me Search Me

Open Science and its Discontents

Edited by Gary Hall

# Digitize Me, Visualize Me, Search Me

Open Science and its Discontents

By [Gary Hall](#)

## White Noise: On the Limits of Openness (Living Book Mix)

One of the aims of the Living Books About Life series is to provide a 'bridge' or point of connection, translation, even interrogation and contestation, between the humanities and the sciences. Accordingly, this introduction to *Digitize Me, Visualize Me, Search Me* takes as its starting point the so-called 'computational turn' to data-intensive scholarship in the humanities.

The phrase '[the computational turn](#)' has been adopted to refer to the process whereby techniques and methodologies drawn from (in this case) *computer science* and related fields – including science visualization, interactive information visualization, image processing, network analysis, statistical data analysis, and the management, manipulation and mining of data – are being used to produce new ways of approaching and understanding texts in the humanities; what is sometimes thought of as 'the digital humanities'. The concern in the main has been with either digitizing 'born analog' humanities texts

and artifacts (e.g. making annotated editions of the art and writing of [William Blake](#) available to scholars and researchers online), or gathering together ‘born digital’ humanities texts and artifacts (videos, websites, games, photography, sound recordings, 3D data), and then taking complex and often extremely large-scale data analysis techniques from computing science and related fields and applying them to these humanities texts and artifacts - to this ‘big data’, as it has been called. Witness Lev Manovich and the Software Studies Initiative’s use of ‘[digital image analysis and new visualization techniques](#)’ to study ‘20,000 pages of Science and Popular Science magazines... published between 1872-1922, 780 paintings by van Gogh, 4535 covers of Time magazine (1923-2009) and one million manga pages’ (Manovich, 2011), or Dan Cohen and Fred Gibbs’s text mining of ‘[the 1,681,161 books that were published in English in the UK in the long nineteenth century](#)’ (Cohen, 2010).

*What Digitize Me, Visualize Me, Search Me* endeavours to show is that such data-focused transformations in research can be seen as part of a major alteration in the status and nature of knowledge. It is an alteration that, according to the philosopher Jean-François Lyotard, has been taking place since at least the 1950s, and involves nothing less than a shift away from a concern with questions of what is right and just, and toward a concern with legitimating power by optimizing the social system’s performance in instrumental, functional terms. This shift has significant consequences for our idea of knowledge. Indeed, for Lyotard:

The nature of knowledge cannot survive unchanged within this context of general transformation. It can fit into the new channels, and become operational, only if learning is translated into quantities of information. We

can predict that anything in the constituted body of knowledge that is not translatable in this way will be abandoned and that the direction of new research will be dictated by the possibility of its eventual results being translatable into computer language. The 'producers' and users of knowledge must now, and will have to, possess the means of translating into these language whatever they want to invent or learn. Research on translating machines is already well advanced. Along with the hegemony of computers comes a certain logic, and therefore a certain set of prescriptions determining which statements are accepted as 'knowledge' statements. (1986: 4)

In particular, *Digitize Me, Visualize Me, Search Me* suggests that the turn in the humanities toward data-driven scholarship, science visualization, statistical data analysis, etc. can be placed alongside all those discourses that are being put forward at the moment - in both the academy and society - in the name of greater openness, transparency, efficiency and accountability.

## **Open Access**

The open access movement very much provides a case in point. Witness [John Houghton's](#) 2009 comparison of the benefits of OA for the United Kingdom, Netherlands and Denmark, which claims to show that the open access academic publishing model, in which peer reviewed scholarly research and publications are made available for free online to all those who are able to access the Internet, is actually the most cost effective mechanism for scholarly publishing. Others meanwhile have detailed the increases open access publishing enables in the amount of material that can be published, searched and stored, in the number of people who can access it, in the impact of that material, the

range of its distribution, and in the speed and ease of reporting and information retrieval. The following announcement, posted on the BOAI (Budapest Open Access Initiative) list in March 2010, is fairly typical in this respect:

Today PLoS released Pubget links across its journal sites. Now, when users are browsing thousands of reference citations on PLoS journals they will be able to get to the full text article faster than ever before.

Specifically, when readers encounter citations to articles as recorded by CrossRef (which are accessed via the 'CrossRef' link in the 'Cited in' section of any article's Metrics tab), a PDF icon will also appear if it is freely available via Pubget. Clicking on the icon will take you directly to the PDF.

On launching this new functionality, Pete Binfield, Publisher of PLoS ONE and the Community Journals said: 'Any service, like Pubget, that makes it easier for authors to quickly find the information they need is a welcome addition to our articles. We like how Pubget helps to break down content walls in science, letting users get instantly to the article-level detail that they seek.' (Pubget, 2010)

## **Open Data**

Yet it is not just the research literature that is positioned as being rendered more accessible by scientists. Even the data created in the course of scientific research is promoted as being made freely and openly available for others to use, analyse and build upon. This includes data sets that are too large to be included in any resulting peer-reviewed publications. Known as open data, or data-sharing, this initiative is motivated by the idea that publishing data online on an

open basis bestows it with a 'vastly increased utility'. Digital data sets are said to be 'easily passed around'; they are seemingly 'more easily reused', reanalysed and checked for accuracy and validity; and they supposedly contain more 'opportunities for educational and commercial exploitation' (Swan, 2009).

Interestingly, certain academic publishers are already viewing the linking of their journals to the underlying data as another of the 'value-added' services they can offer, to set alongside automatic alerting and sophisticated citation, indexing, searching and linking facilities (and to no doubt help ward off the threat of disintermediation posed by the development of digital technology, which enables academics to take over the means of dissemination and publish their work for and by themselves cheaply and easily). Significantly, a 2009 JISC report also identified 'open-ness, predictive science based on massive data volumes and citizen involvement as [all] being important features of tomorrow's research practice'.

In a further move in this direction, all Public Library of Science (PLoS) journals are now providing a broad range of article-level metrics and indicators relating to usage data on an open basis. No longer withheld as trade secrets, these metrics reveal which articles are attracting the most views, citations from the scholarly literature, social bookmarks, coverage in the media, comments, responses, 'star' ratings, blog coverage, and so on. PLoS has positioned this programme as enabling science scholars to assess 'research articles on their own merits rather than on the basis of the journal (and its impact factor) where the work happens to be published', and they encourage readers to carry out their own analyses of this open data (Patterson, 2009).

Yet it is difficult not to perceive such article-level metrics and management tools as also being part of the wider process of transforming knowledge and learning into ‘quantities of information’ (Lyotard, 1986: 4); quantities, furthermore, that are produced to be exchanged, marketed and sold (1986: 4) – for example, by individual academics to their departments, institutions, funders and governments in the form of indicators of ‘quality’ and ‘impact’ (1986: 5).

## **From Open Science to Open Government**

Such developments around open access and open data are themselves part of the larger trend or phenomenon that is coming to be known as ‘open science’. As Murray et al put it:

Open science is emerging as a collaborative and transparent approach to research. It is the idea that all data (both published and unpublished) should be freely available, and that private interests should not stymie its use by means of copyright, intellectual property rights and patents. It also embraces open access publishing and open source software... (Murray et al., 2008)

One of the most interesting and well known examples of how such open science may work is provided by the Open Notebook Science of the organic chemist Jean-Claude Bradley. ‘[I]n the interests of openness’, Bradley is making the ‘details of every experiment done in his lab freely available on the web’. This ‘includes all the data generated from these experiments too, even the failed experiments’. What is more, he is doing so in ‘real time’, ‘within hours of production, not after the months or years involved in peer review’ (Poynder, 2010). Again, we can see how emphasis is being placed on the amount of research that can be shared, and the

speed with which this can be achieved. This openness on Bradley's part is also positioned as a means of achieving usefulness and impact, as is evident from the very title of one of his Open Notebook Science projects, [UsefulChem](#).

To be fair, however, such discourses around openness, transparency, efficiency and utility are not confined to the sciences – or even the university, for that matter. There are also wider political initiatives, dubbed 'Open Government', or 'Government 2.0'. Both the Labour and the Conservative/Liberal Democrat coalition administrations in the UK have made a great display of freeing government information. The Labour government implemented the Freedom of Information (FOI) Act in 2000, and then proceeded to launch a [website](#) expressly dedicated to the release of governmental data sets in January 2010. It is a website that the current Conservative/Liberal Democrat coalition government continues to make extensive use of. In a similar vein, the [Guardian](#) newspaper has campaigned for the UK government to relinquish its copyright on all local, regional and national data collected with taxpayers' money and to make such data freely and openly available to the public by publishing it online, where it can be collectively and collaboratively scrutinized, searched, mined, mapped, graphed, cross-tabulated, visualized, audited and interpreted using software tools.

Nor is this phenomenon confined to the UK. In the United States Barack Obama promised throughout his election campaign to make government more open. He followed this up by issuing a memorandum on transparency the very first day after he became President, vowing to make openness one of [the](#)



touchstones of this presidency” (Obama, cited in Stolberg, 2009): ‘My Administration is committed to creating an unprecedented level of openness in Government. We will work together to ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government’ (The White House, 2009).

## **The Politics of Openness**

Now, the connection I am making here between the movements for open access, open data, open science and open government is one that has to a certain extent already been pointed to by Michael Gurstein in his reflections on the experience of attending the 2011 conference of the Open Knowledge Foundation. For Gurstein:

the ‘open data/open government’ movement begins from a profoundly political perspective that government is largely ineffective and inefficient (and possibly corrupt) and that it hides that ineffectiveness and inefficiency (and possible corruption) from public scrutiny through lack of transparency in its operations and particularly in denying to the public access to information (data) about its operations. And further that this access once available would give citizens the means to hold bureaucrats (and their political masters) accountable for their actions. In doing so it would give these self-same citizens a platform on which to undertake (or at least collaborate with) these bureaucrats in certain key and significant activities—planning, analyzing, budgeting that sort of thing. Moreover through the implementation of processes of crowdsourcing this would also provide the bureaucrats with the overwhelming benefits of having access to and input from the knowledge and wisdom of the broader interested public.

Put in somewhat different terms but with essentially the same meaning—it's the taxpayer's money and they have the right to participate in overseeing how it is spent. Having "open" access to government's data/information gives citizens the tools to exercise that right. (Gurstein, 2011)

Interestingly, for Gurstein, a much clearer understanding is needed than has been displayed by many open data/open government advocates to date of what exactly is meant by openness, and of where arguments in favour of open access, open information and open data are likely to lead us in the not too distant future. With this in mind, perhaps we could endeavour to put some flesh on the bones of Gurstein's sketch of the politics of openness and suggest that, from a liberal perspective, freeing publicly funded and acquired information and data – whether it is gathered directly in the process of census collection, or indirectly as part of other activities (crime, healthcare, transport, schools and accident statistics) – is seen as helping society to perform more efficiently. For liberals, openness is said to play a key role in increasing citizen trust, participation and involvement in democracy, and indeed government, as access to information – such as that needed to intervene in public policy – is no longer restricted either to the state or to those corporations, institutions, organizations and individuals who have sufficient money and power to acquire it for themselves.

Such liberal beliefs find support in the idea that making information and data freely and transparently available goes along with Article 19 of The Universal Declaration of Human Rights. The latter states that everyone has the right 'to seek, receive and impart information and ideas through any media and regardless of frontiers'. Hillary Clinton, the United

States Secretary of State, put forward a similar vision when, at the beginning of 2010, she said of her country that 'We stand for a single internet where all of humanity has equal access to knowledge and ideas', and against the authoritarian censorship and suppression of free speech and online search facilities like Google in countries such as China and Iran. Clinton declared:

Even in authoritarian countries, information networks are helping people discover new facts and making governments more accountable.

During his visit to China in November [2009], President Obama held a town hall meeting with an online component to highlight the importance of the internet. In response to a question that was sent in over the internet, he defended the right of people to freely access information, and said that the more freely information flows, the stronger societies become. He spoke about how access to information helps citizens to hold their governments accountable, generates new ideas, and encourages creativity. The United States' belief in that truth is what brings me here today. (Clinton, 2010)

This political sentiment was shared by Jeff Jarvis, author of *What Would Google Do?*, when, in support of Google's decision to stop self-filtering search results in China, he argued in March 2010 for a bill of rights for cyberspace: 'to claim and secure our freedom to connect, speak, assemble, and act online; to each control our identities and data; to speak our languages; to protect what is public and private; and to assure openness' (Jarvis, 2010: 4). Yet are Clinton and Jarvis not both guilty here of overlooking (or should that be conveniently forgetting or even denying) the way liberal ideas of freedom and openness (and, indeed, of the human) have long been used in the service of

colonialism and neoliberal globalisation? Does freedom for the latter not primarily mean economic freedom, i.e., freedom of the market, freedom of the consumer to choose what to consume – not only in terms of goods, but also lifestyles and ways of being?

It is interesting that ‘fifteen years after the Freedom of Information Act law was passed’ in the US in 1966, ‘the General Accounting Office reported that 82 percent of requests [for information] came from business, nine percent from the press, and only 1 percent from individuals or public interest groups’, even if all this was before the widespread use of networked computers (Fung et al, 2007: 27-28). Certainly, in the UK today, the ‘truth is that the [UK] FOI Act [2000] isn’t used, for the most part, by “the people”, as Tony Blair acknowledged in his recent memoir. ‘It’s used by journalists’ (Blair, 2010) – and by businesses, one might add. In view of this, it is no surprise to find that neoliberals also support the making of government data freely and openly available to businesses and the public. They do so on the grounds that it provides a means of achieving the best possible ‘input/output ratio’ for society (Lyotard, 1986: 54). This way of thinking is of a piece with the emphasis placed by neoliberalism’s audit culture on accountability, transparency, evaluation, measurement and centralised data management: for example, in the context of UK higher education, it is evident in the emphasis placed on measuring the impact of research on society and the economy, teaching standards, contact hours, as well as student drop-out rates, future employment destinations and earning prospects. From this perspective, such openness and communicative transparency is perceived as ensuring greater value for (taxpayers’) money, supposedly helping to eliminate corruption, enabling

costs to be distributed more effectively, and increasing choice, innovation, enterprise, creativity, competitiveness and accountability.

Meanwhile, some libertarians have gone so far as to argue that there is no need to make difficult policy decisions about what data and what information it is right to publish online and what to keep secret at all. Instead, we should work toward the kind of situation the science-fiction writer Bruce Sterling proposes. In *Shaping Things*, his non-fiction book on the future of design, Sterling advocates retaining all data and information, 'the known, the unknown known, and the unknown unknown', in large archives and databases equipped with the necessary bandwidth, processing speed and storage capacity, and simply devising search tools and metadata that are accurate, fast and powerful enough to find and access it (Sterling, 2005: 47).

Yet to have participated in the shift away from questions of truth, justice and especially what, in *The Inhuman*, Lyotard places under the headings of 'heterogeneity, dissensus, event...the unharmonizable' (1991: 4), and *toward* a concern with performativity, measurement and optimising the relation between input and output, one does not need to be a practicing [data journalist](#), or to have actively contributed to the movements for open access, open data, open science or open government. If you are one of the 1.3 million plus people who have purchased a Kindle, and helped the sale of digital books outpace those of hardbacks on Amazon's US website, then you have already signed a license agreement allowing the online book retailer - but not academic researchers or the public - to collect, store, mine, analyse and extract economic value from data concerning your personal reading habits for free.

Similarly, if you are one of the over 687 million worldwide who use the Facebook social network, then you are already voluntarily giving your time and labour for free, not only to help its owners, their investors, and other companies make a reputed \$1 billion a year from demographically targeted advertising, but to supply law enforcement agencies with profile data relating to yourself, your family, friends, colleagues and peers that they can use in investigations (Hoffman, 2010). Even if you have done neither, you will in all probability have provided the Google technology company with a host of network data and digital traces it can both monetize and give to the police as a result of having mapped your home, digitized your book, or supplied you with free music videos to enjoy via Google Street View, Google Maps, Google Earth, Google Book Search and YouTube, which Google also owns. Lest this shift from open access to Google should seem somewhat farfetched, it is worth recalling that 'Google has moved to establish, embellish, or replace many core university services such as library databases, search interfaces, and e-mail servers' (Vaidhyathan, 2009: 65-66); and that academia in fact gave birth to Google, Google's PageRank algorithm being little more 'than an expansion of what is known as citation analysis' (Knouf, 2010).

Obviously, no matter how exciting and enjoyable such activities may be, you don't *have* to buy that e-book reader, join that social network or display your personal metrics online, from sexual activity to food consumption, in an attempt to identify patterns in your life – what is called life-tracking or self-tracking. (Although, actually, a lot of people are quite happy to keep contributing to the networked communities reached by Facebook and YouTube, even though they

realise they are being used as free labour and that, in the case of the former, much of what they do cannot be accessed by search engines and web browsers. They just see this as being part of the deal and a reasonable trade-off for the services and experiences that are provided by these companies.) Nevertheless, refusing to take part in this transformation of knowledge and learning into quantities of data, and shift *away* from critical questions of what is just and right *toward* a concern with optimizing the system's performance is not an option for most of us. It is not something that can be opted out of by simply declining to take out a Tesco Club Card or use cash-points, refusing to look for research using Google Scholar, or committing social networking '[suicide](#)' and reading print-on-paper books instead.

For one thing, the process of capturing data by means not just of the internet, but a myriad of cameras, sensors and robotic devices, is now so ubiquitous and all pervasive it is impossible to avoid being caught up in it, no matter how rich, knowledgeable and technologically proficient you are. The latest research indicates there are approximately 1.85 million CCTV cameras in the UK – one for every 32 people. Yet no one really knows how many CCTV cameras are actually in operation in Britain today - and that's without even mentioning other means of gathering data that are reputed to be more intrusive still, such as mobile phone GPS location and automatic vehicle number plate recognition (ANPR).

For another, and as the example of CCTV illustrates, it's not necessarily a question of actively doing something in this respect: of positively contributing free labour to the likes of Flickr and YouTube, for instance,

or of refusing to do so. Nor is it merely a case of the separation between work and non-work being harder to maintain nowadays. (Is it work, leisure or play when you are writing a status update on Facebook, posting a photograph, ‘friending’ someone, interacting, detailing your ‘likes’ and ‘dislikes’ regarding the places you eat, the films you watch, the books you read?) As Gilles Deleuze and Felix Guattari pointed out some time ago, ‘surplus labor no longer requires labor... one may furnish surplus-value without doing any work’, or anything that even remotely resembles work for that matter, at least as it is most commonly understood:

In these new conditions, it remains true that all labour involves surplus labor; but surplus labor no longer requires labor. Surplus labor, capitalist organization in its entirety, operates less and less by the striation of space-time corresponding to the physicosocial concept of work. Rather, it is as though human alienation through surplus labor were replaced by a generalized ‘machinic enslavement’, such that one may furnish surplus-value without doing any work (children, the retired, the unemployed, television viewers, etc.). Not only does the user as such tend to become an employee, but capitalism operates less on a quantity of labor than by a complex qualitative process bringing into play modes of transportation, urban models, the media, the entertainment industries, ways of perceiving and feeling – every semiotic system. It is as though, at the outcome of the striation that capitalism was able to carry to an unequalled point of perfection, circulating capital necessarily recreated, reconstituted, a sort of smooth space in which the destiny of human beings is recast. (Deleuze and Guattari, 1988: 492)

## **Transparency?**

Before going any further, I should perhaps confess that I am a staunch advocate of open access in the



humanities. Nevertheless, there are a number of issues that need to be raised with regard to making research and data openly available online for free.

The first point to make in this respect is that, far from revealing any hitherto unknown, hidden or secret knowledge, such discourses of openness and transparency are themselves often not very open or transparent. Staying with the relationship between politics and science, let us take as an example the response of Ed Miliband, leader of the UK's Labour Party, to the '[Climategate](#)' controversy, in which climate skeptics alleged that emails hacked from the University of East Anglia's Climatic Research Unit revealed that scientists have tampered with the data in order to support the theory that global warming is man-made. Miliband's answer was to advocate '[maximum transparency](#)— let's get the data out there', he urged. 'The people who believe that climate change is happening and is man-made have nothing to fear from transparency' (Miliband, quoted in Westcott, 2009: 7; cited by Birchall, 2011b). Yet, actually, complete transparency is impossible. This is because, as Clare Birchall has shown, there is an aporia at the heart of any claim to transparency. 'For transparency to be known as transparency, there must be some agency (such as the media [or politicians, or government]) that legitimizes it as transparent, and because there is a legitimizing agent which does not itself have to be transparent, there is a limit to transparency' (Birchall, 2011a: 142). In fact, the more transparency is claimed, the more the violence of the mediating agency of this transparency is concealed, forgotten or obscured. Birchall offers the example of 'The Daily Telegraph and its exposure of MPs' expenses during the summer of 2009. While appearing to act on the side of

transparency, as a commercial enterprise the paper itself has in the past been subject to secret takeover bids and its former owner, Lord Conrad Black, convicted of fraud and obstructing justice' (Birchall, 2011a: 142). To paraphrase a question from Lyotard I am going to return to at more length: Who decides what transparency is, and who knows what needs to be transparent (1986: 9)?

Furthermore, merely making such information and data available to the public online will not in itself necessarily change anything. In fact, such processes have often been adopted precisely as a means of avoiding change. Aaron Swartz provides the example of Watergate: 'after Watergate, people were upset about politicians receiving millions of dollars from large corporations. But, on the other hand, corporations seem to like paying off politicians. So instead of banning the practice, Congress simply required that politicians keep track of everyone who gives them money and file a report on it for public inspection' (Swartz, 2010).

## **Openness?**

Much the same can be said for the idea that making research and data accessible to the public supposedly helps to make society more open and free. Take the belief we saw expressed above by Hilary Clinton: that people in the United States have free access to the internet while those in China and Iran do not. Those of us who live and work in the West do indeed have a certain freedom to publish and search online. Yet none of this rhetoric about freedom and transparency prevented the Obama government from condemning Wikileaks in November 2010 as 'reckless and dangerous', after it opened up access to hundreds of

thousands of classified State Department documents (Gibbs, 2010); nor from putting pressure on Amazon and other companies to stop hosting the whistle-blowing website, an action which had echoes of the dispute over censorship between Google and the Chinese government earlier in 2010. (Significantly, the Obama administration has also recently withdrawn the majority of funding from the United States open government website [www.data.gov](http://www.data.gov), which served as an influential precursor to the previously mentioned [www.data.gov.uk](http://www.data.gov.uk) website in the UK.) Furthermore, unless you are a large political or economic actor, or one of the lucky few, the statistics show that what you publish online is unlikely to receive much attention. Just ‘three companies – Google, Yahoo! and Microsoft – handle 95 percent of all search queries’; while ‘for searches containing the name of a specific political organisation, Yahoo! and Google agree on the top result 90 percent of the time’ (Hindman, 2009: 59, 79). Meanwhile, one company, Google, reportedly has 65 % of the world’s search market, ‘72 per cent share of the US search market, and almost 90 per cent in the UK’ – a degree of domination that has led the European Union to investigate Google for abusing its power to favour its own products while suppressing those of rivals (Arthur, 2010: 3).

But it is not just that Google’s algorithms are ranking some websites on the first page of its results and others on page 42 (which means, in effect, that the latter are rarely going to be accessed, since very few people read beyond the first page of Google’s results). It is that conventional search engines are reaching only an extremely small percentage of the total number of available web pages. Ten years ago Michael K. Bergman was already placing the figure at 0.03%, or

'one in 3,000', with 'public information on the deep Web' even then being '400 to 550 times larger than the commonly defined World Wide Web'. Consequently, while according to Bergman as much as 'ninety-five per cent of the deep Web' may be 'publicly accessible information – not subject to fees or subscriptions' – by far the vast majority of it is left untouched (Bergman, 2001). And that is before we even begin to address the issue of how the recent rise of the app, and use of the password protected Facebook for search purposes, may today be annihilating the very idea of the openly searchable Web.

We can therefore see that it is not enough simply to 'Free Our Data', as the Guardian has it; or to operate on the basis that 'information wants to be free' (Wark, 2004) (although doing so of course may be a start, especially in an era when notions of the open web and net neutrality are under severe threat). We can put ever more research and data online; we can make it freely available to both other researchers and the public under open access, open data, open science and open government conditions; we can even integrate, index and link it using the appropriate metadata to enable it to be searched and harvested with relative ease. But none of this means this research and data is going to be found. Ideas of this kind ignore the fact that all information and data is ordered, structured, selected and framed in a particular way. This is what metadata is for, after all. Metadata is information or data that describes, links to, or is otherwise used to control, find, select, filter, classify and present other data. One example would be the information provided at the front of a book detailing its publisher, date and place of publication, ISBN number, and so on. However, the term 'metadata' is most commonly associated with the

language of computing. There, metadata is what enables computers to access files and documents, not just in their own hard drives, but potentially across a range of different platforms, servers, websites and databases. Yet for all its associations with computer science, metadata is never neutral or objective. Although the term ‘data’ comes from the Latin word *datum*, meaning ‘something given’, data is not simply objectively out there in the world already provided for us. The specific ways in which metadata is created, organized and presented helps to produce (rather than merely passively reflect) what is classified as data and information – and what is not.

Clearly, then, it is not just a question of free and open access to the research and data; nor of providing support, education and training on how to understand, interpret, use and apply it effectively, as Gurstein has argued (2010). It is also a question of who (and what) makes decisions regarding the data and metadata, and thus gets to exercise control over it, and on what basis such decisions are made. To paraphrase Lyotard once more: who decides what data and metadata is, and who knows what needs to be decided? (1986: 9). Who gets to legislate? And who legitimates the legislators (1986: 8)? Will the ‘ruling class’ – top civil servants and consulting firms full of people with MBAs, ‘corporate leaders, high-level administrators, and the heads of the major professional, labor, political, and religious organizations’, including those behind Google, Apple, Facebook, Amazon, JISC, AHRC, OAI, SPARC, COASP – continue to operate as the class of interpreters, gatekeepers and ‘decision makers,’ not just with regard to having ‘access to the information these machines must have in storage to guarantee that the right decisions are made’, but with regard to creating and

controlling the data and metadata, too (1986: 14)?

## **On Data-Intensive Scholarship**

If, as demonstrated above, discourses of openness and transparency are themselves not very open or transparent at all, much of the current emphasis on making the research and data open and free is also lacking in self-reflectivity and meaningful critique. We can see this not just in those discourses associated with open access, open data, open science and open government that are explicitly emphasizing the importance of transparency, performativity and efficiency. This lack of criticality is apparent in much of what goes under the name of ‘digital humanities’, too, especially those elements associated with the ‘computational turn’.

We tend to think of the humanities as being self-reflexive per se, and as frequently asking questions capable of troubling culture and society. Yet after decades when humanities scholarship made active use of a variety of critical theories – Marxist, psychoanalytic, post-colonialist, post-Marxist – it seems somewhat surprising that many advocates of this current turn to data-intensive scholarship in the humanities find it difficult to understand computing and the digital as much more than tools, techniques and resources. As a result, much of the scholarship that is currently occurring under the ‘digital humanities’ agenda is uncritical, naive and at times even banal ([Liu](#), 2011; [Higgen](#), 2010).

Witness the current emphasis on making the data not only visible but also visual. Stefanie Posavec’s

frequently referred to [Literary Organism](#), which visualises the structure of Part One of Kerouac's *On the Road* as a tree, provides one example; those cited earlier courtesy of Lev Manovich and the Software Studies Initiative offer another. Now, there is a long history of critical engagement within the humanities with ideas of the visual, the image, the spectacle, the spectator and so on: not just in critical theory, but also in cultural studies, women's studies, media studies, film and television studies. Such a history of critical engagement stretches back to Guy Debord's influential 1967 work, *The Society of the Spectacle*, and beyond. For example, in his introduction to a 1995 book edited with Lynn Cooke, *Visual Display: Culture Beyond Appearances*, Peter Wollen writes that an excess of visual display within culture has 'the effect of concealing the truth of the society that produces it, providing the viewer with an unending stream of images that might best be understood, not simply detached from a real world of things, as Debord implied, but as effacing any trace of the symbolic, condemning the viewer to a world in which we can see everything but understand nothing—allowing us viewer-victims, in Debord's phrase, only "a random choice of ephemera"' (1995: 9). It can come as something of a surprise, then, to discover that this humanities tradition in which ideas of the visual are engaged critically appears to have had comparatively little impact on the current enthusiasm for data visualisation that is so prominent an aspect of the turn toward data-intensive scholarship.

Of course, this (at times explicit) repudiation of criticality could be precisely what makes certain aspects of the digital humanities so seductive for many at the moment. Exponents of the computational turn

can be said to be endeavouring to avoid conforming to accepted (and often moralistic) conceptions of politics that have been decided in advance, including those that see it only in terms of power, ideology, race, gender, class, sexuality, ecology, affect etc. Refusing to 'go through the motions of a critical avant-garde', to borrow the words of Bruno Latour (2004), they often position themselves as responding to what is perceived as a fundamentally new cultural situation, and to the challenge it represents to our traditional methods of studying culture, by avoiding conventional theoretical manoeuvres and by experimenting with the development of fresh methods and approaches for the humanities instead.

Manovich, for instance, sees the sheer scale and dynamics of the contemporary new media landscape as presenting the usually accepted means of studying culture that were dominant for so much of the 20th century – the kinds of theories, concepts and methods appropriate to producing close readings of a relatively small number of texts – with a significant practical and conceptual challenge. In the past, 'cultural theorists and historians could generate theories and histories based on small data sets (for instance, “classical Hollywood cinema”, “Italian Renaissance”, etc.) But how can we track “global digital cultures”, with their billions of cultural objects, and hundreds of millions of contributors’, he asks (Manovich, 2010)? Three years ago Manovich was already describing the ‘numbers of people participating in social networks, sharing media, and creating user-generated content’ as simply ‘astonishing’:

MySpace, for example, claims 300 million users. Cyworld, a Korean site similar to MySpace, claims 90 percent of South Koreans in their 20s and 25 percent of that



country's total population (as of 2006) use it. Hi5, a leading social media site in Central America has 100 million users and Facebook, 14 million photo uploads daily. The number of new videos uploaded to YouTube every twenty-four hours (as of July 2006): 65,000. (Manovich in Franklin & Rodriguez'G, 2008)

The solution Manovich proposes to this 'data deluge' is to turn to the very computers, databases, software and vast amounts of born-digital networked cultural content that are causing the problem in the first place, and to use them to help develop new methods and approaches adequate to the task at hand. This is where what he calls Cultural Analytics comes in. 'The key idea of Cultural Analytics is the use of computers to automatically analyze cultural artefacts in visual media, extracting large numbers of features that characterize their structure and content' (Manovich in Kerssens & Dekker, 2009); and what is more, to do so not just with regard to the culture of the past, but also with that of the present. To this end, Manovich (not unlike the Google technology company) calls for as much of culture as possible to be made available in external, digital form: 'not only the exceptional but also the typical; not only the few cultural sentences spoken by a few "great man" [sic] but the patterns in all cultural sentences spoken by everybody else' (Manovich in Kerssens & Dekker, 2009).

In a series of posts on his Found History blog, Tom Scheinfeldt, managing director at the Center for History and New Media at George Mason University, positions such developments in terms of a shift from a concern with theory and ideology to a concern with methodology (2008). In this respect there may well be a degree of 'relief in having escaped the culture wars of the 1980s' – for those in the US especially – as a result

of this move ‘into the space of methodological work’ (Croxall, 2010) and what Scheinfeldt reportedly dubs ‘[the post-theoretical age](#)’ (cited in P. Cohen, 2010). The problem, though, is that without such reflexive critical thinking and theories many of those whose work forms part of this computational turn find it difficult to articulate exactly what the point of what they are doing is, as Scheinfeldt readily acknowledges (2010a).

Take one of the projects mentioned earlier: the attempt by [Dan Cohen and Fred Gibbs](#) to text-mine all the books published in English in the Victorian age (or at least those digitized by Google). Among other things, this allows Cohen and Gibbs to show that use of the word ‘revolution’ in book titles of the period spiked around ‘[the French Revolution and the revolutions of 1848](#)’ (D. Cohen, 2010). But what argument are they trying to make with this calculation? What is it we are able to learn as a result of this use of computational power on their part that we did not know already and could not have discovered without it ([Scheinfeldt](#), 2008)?

In an explicit response to Cohen and Gibbs’s project, Scheinfeldt suggests that the problem of theory, or the lack of it, may actually be a question of scale:

[It expects something of the scale of humanities scholarship](#) which I’m not sure is true anymore: that a single scholar—nay, every scholar—working alone will, over the course of his or her lifetime ... make a fundamental theoretical advance to the field.

Increasingly, this expectation is something peculiar to the humanities. ...it required the work of a generation of mathematicians and observational astronomers, gainfully employed, to enable the eventual “discovery” of Neptune... Since the scientific revolution, most theoretical advances

play out over generations, not single careers. (Scheinfeldt, 2010b)

Now, it is absolutely important that we as scholars experiment with the new tools, methods and materials that digital media technologies create and make possible, in order to bring into play new forms of Foucauldian *dispositifs*, or what Bernard Stiegler calls *hypomnemata*, or what I am trying to think in terms of [media gifts](#). I would include in this 'experimentation imperative' techniques and methodologies drawn from computer science and other related fields, such as information visualisation, data mining and so forth. Nevertheless, there is something troubling about this kind of deferral of critical and self-reflexive theoretical questions to an unknown point in time, still possibly a generation away. After all, the frequent suggestion is that now is not the right time to be making any such decision or judgement, since we cannot yet know how humanists will eventually come to use these tools and data, and thus what data-driven scholarship may or may not turn out to be capable of, critically, politically, theoretically. One of the consequences of this deferral, however, is that it makes it extremely difficult to judge whether this postponement is indeed acting as a responsible, political and ethical opening to the (heterogeneity and incalculability of the) future, including the future of the humanities; or whether it is serving as an alibi for a naive and rather superficial form of scholarship instead ([Meeks](#), 2010). A form of scholarship moreover that, in uncritically and un-self-reflexively adopting techniques and methodologies drawn from computer science, can be seen as part of the larger shift in contemporary society which Lyotard associates with the widespread use of computers and databases, and with the exteriorization of knowledge in

relation to the 'knower'. As we have seen, it is a movement away from a concern with ideals, with what is right and just and true, and toward a concern to legitimate power by optimizing the system's performance in instrumental, functional terms.

All of this raises some rather significant and timely questions for the humanities. Is it merely a coincidence that such a turn toward science, computing and data-intensive research is gaining momentum at a time when the UK government is emphasizing the importance of the STEM subjects (science, technology, engineering and medicine) and withdrawing support and funding from the humanities? Or is one of the reasons all this is happening now due to the fact that the humanities, like the sciences themselves, are under pressure from government, business, management, industry and increasingly the media to prove they provide value for money in instrumental, functional, performative terms? Is the interest in computing a strategic decision on the part of some of those in the humanities? As the project of Cohen and Gibbs shows, one can get funding from the likes of Google (D. Cohen, 2010). In fact, in the summer of 2010 ['Google awarded \\$1 million to professors doing digital humanities research'](#) (P. Cohen, 2010). To what extent is the take-up of practical techniques and approaches from computing science providing some areas of the humanities with a means of defending (and refreshing) themselves in an era of global economic crisis and severe cuts to higher education, through the transformation of their knowledge and learning into quantities of information -- so-called 'deliverables'? Can we even position the 'computational turn' as an event created to justify such a move on the part of certain elements within the humanities ([Frabetti](#), 2010)?

Where does all this leave us as far as this Living Book on open science is concerned? As the argument above hopefully demonstrates, it is clearly not enough just to attempt to reveal or recover the scientific truth about, say, the environment, to counter the disinformation of others involved in the likes of the Climategate controversy. Nor is it enough merely to make the scientific research openly accessible to the public. Equally, it is not satisfactory simply to make the information, data, and associated tools, techniques and resources freely available to those in the humanities, so they can collectively and collaboratively search, mine, map, graph, model, visualize, analyse and interpret it in new ways – including some that may make it less abstract and easier for the majority of those in society to understand and follow – and, in doing so, help bridge the gap between the ‘two cultures’. It is not so much that there is a lack of information, or access to the right kind of information, or information presented in the right kind of way to ensure that the message of the scientific research and data comes across effectively and efficiently. It is rather that there is too much information, too much white noise, as ‘Bifo’ et al call it (2009: 141-142). As a [2010 Mintel report](#) showed – to stay with the example of climate change – most people in the UK already know what is happening to the environment. They are just suffering from Green Fatigue: they are bored with thinking about it and thus enacting a backlash against what they perceive as ‘extreme’ pressure from environmentalist groups. This is perhaps one reason why [‘the number of cars on UK roads has risen from just over 26million in 2005 to more than 31 million in 2009’](#) (Shields, 2010: 30).

As Gilles Deleuze and Felix Guattari put it, then: ‘We

do not lack communication. On the contrary, we have too much of it'. What we actually lack is creation. 'We lack resistance to the present' (1994: 108). It is therefore not simply a case of supplying more scientific research and data; nor of making the research and data that has otherwise been closed, hidden, denied or suppressed openly available for free – by opening the already existing memory and databanks to the people, for example (which is what Lyotard ended by suggesting we do). It is more a case of creating work around the research and data that does not simply go along with the shift in the status and nature of knowledge that is currently taking place. As we have seen, it is a shift toward STEM subjects and away from the humanities; toward a concern with optimizing the social system's performance in instrumental, functional terms, and away from a concern with questions of what is just and right; and toward an emphasis on openness, freedom and transparency, and away from what is capable of disrupting and disturbing society, and what, in remaining resistant to a culture of measurement and calculation, maintains a much needed element of inaccessibility, inefficiency, delay, error, antagonism, heterogeneity and dissensus within the system.

Can this Living Book on open science be considered one such a creation? And can this series of Living Books about Life be considered another? Are they instances of a resistance to the present? Or just more white noise?

(The above is based on a paper presented at the Data Landscapes, AHRC network event, held in conjunction with the British Antarctic Survey at the University of Westminster, London, December 15, 2010. An earlier version of some of the material provided above

appeared in Hall [2010])

## References

Arthur, C. (2010), 'Will Brussels Curb Google Guys', *The Guardian*, December 6.

'Bifo' Berardi, F., Jacquemet, M. and Vitali, G. (2009), *Ethereal Shadows: Communications and Power in Contemporary Italy*. Brooklyn, New York: Autonomedia.

Bergman, M. K. (2001), 'The Deep Web: Surfacing Hidden Value', *JEP: The Journal of Electronic Publishing*, vol.7, no.1, August.  
<http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0007.104.&nbsp;>

Birchall, C. (2011) 'There's Been Too Much Secrecy in this City': The False Choice Between Secrecy and Transparency in US Politics,' *Cultural Politics* 7(1), March: 133-156.

Birchall, C (2011b forthcoming) 'Transparency, Interrupted: Secrets of the Left', *Between Transparency and Secrecy*, Annual Review, *Theory, Culture and Society*, December.

Blair, T. (2010), *A Journey*. London: Hutchinson.

Clinton, H. (2010), 'Internet Freedom: The Prepared Text of U.S. of Secretary of State Hillary Rodham Clinton's speech, delivered at the Newseum in Washington, D.C., January 21.

[http://www.foreignpolicy.com/articles/2010/01/21/internet\\_freedom?page=full](http://www.foreignpolicy.com/articles/2010/01/21/internet_freedom?page=full)

Cohen, D. (2010), 'Searching for the Victorians', *Dan Cohen*, October 4.

[http://www.dancohen.org/2010/10/04/searching-for-the-victorians/?utm\\_source=feedburner&utm\\_medium=feed&utm\\_campaign=Feed%3A+DanCohen+%28Dan+Cohen%29&utm\\_content=Google+Reader](http://www.dancohen.org/2010/10/04/searching-for-the-victorians/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+DanCohen+%28Dan+Cohen%29&utm_content=Google+Reader).

Cohen, P. (2010) 'Digital Keys for Unlocking the Humanities' Riches', *The New York Times*, November 16.

[http://www.nytimes.com/2010/11/17/arts/17digital.html?\\_r=1&hp=&pagewanted=all](http://www.nytimes.com/2010/11/17/arts/17digital.html?_r=1&hp=&pagewanted=all).

Croxall, B. (2010) response to Tanner Higgen, 'Cultural Politics, Critique, and the Digital Humanities', *Gaming the System*. September 10.

<http://www.tannerhiggin.com/2010/05/cultural-politics-critique-and-the-digital-humanities/>.

Deleuze G. and Guattari, F. (1988) *A Thousand Plateaus: Capitalism and Schizophrenia*. London: Athlone.

Deleuze G. and Guattari, F. (1994), *What is Philosophy?*. New York: Columbia University Press.

Fung, A., Graham, M., Weil, D. (2007), *Full Disclosure: The Perils and Promise of Transparency*. Cambridge: Cambridge University Press.

Frabetti, F. (2010) 'Digital Again? The Humanities Between the Computational Turn and Ordinary Technicity', talk given to the Open Media Group,



Coventry School of Art and Design. November 9.  
<http://coventryuniversity.podbean.com/2010/11/09/open-software-and-digital-humanities-federica-frabetti/>.

Franklin, K. D. and Rodriguez'G, K. (2008) 'The Next Big Thing in Humanities, Arts and Social Science Computing: Cultural Analytics', *HPC Wire*. July 29.  
[http://www.hpcwire.com/features/The\\_Next\\_Big\\_Thing\\_in\\_Humanities\\_Arts\\_and\\_Social\\_Science\\_Computing\\_Cultural\\_Analytics.html](http://www.hpcwire.com/features/The_Next_Big_Thing_in_Humanities_Arts_and_Social_Science_Computing_Cultural_Analytics.html).

Gibbs, R. (2010), Presidential press secretary, cited in 'White House condemns WikiLeaks' release', *MCNBC.com News*, November 28.  
[http://www.msnbc.msn.com/id/40405589/ns/us\\_news-security](http://www.msnbc.msn.com/id/40405589/ns/us_news-security).

Gurstein, M. (2010a), 'Open Data: Empowering the Empowered or Effective Data Use for Everyone?', *Gurstein's Community Infomatics*, September, 2.  
<http://gurstein.wordpress.com/2010/09/02/open-data-empowering-the-empowered-or-effective-data-use-for-everyone/>.

Gurstein, M. (2010b), 'Open Data (2): Effective Data Use', *Gurstein's Community Infomatics*, September, 9.  
<http://gurstein.wordpress.com/2010/09/09/open-data-2-effective-data-use/>

Gurstein, M. (2011), 'Are the Open Data Warriors Fighting for Robin Hood or the Sheriff?: Some Reflections on OKCon 2011 and the Emerging Data Divide', posting to the nettime mailing list, July, 5.

Hall, G. (2010), 'We Can Know It For You: The Secret Life of Metadata', *How We Became Metadata*. London:

Institute for Modern and Contemporary Culture,  
University of Westminster.

Higgen, T. (2010) 'Cultural Politics, Critique, and the Digital Humanities', *Gaming the System*. May 25.  
<http://www.tannerhiggin.com/2010/05/cultural-politics-critique-and-the-digital-humanities/>.

Hindman, M. (2009), *The Myth of Digital Democracy*.  
Princeton, NJ and Oxford: Princeton University Press.

Hoffman, M. (2010), 'EFF Posts Documents Detailing Law Enforcement Collection of Data From Social Media Sites', *Electronic Frontier Foundation*. March 16.  
<http://www.eff.org/deeplinks/2010/03/eff-posts-documents-detailing-law-enforcement>.

Houghton, J. (2009) 'Open Access - What are the Economic Benefits?: A Comparison of the United Kingdom, Netherlands and Denmark', Centre for Strategic Economic Studies, Victoria University, Melbourne. [http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fOA\\_What\\_are\\_the\\_economic\\_benefits\\_-\\_a\\_comparison\\_of\\_UK-NL-DK\\_FINAL\\_logos.pdf](http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fOA_What_are_the_economic_benefits_-_a_comparison_of_UK-NL-DK_FINAL_logos.pdf).

Jarvis, J. (2010), 'Time For Citizens of the Internet to Stand Up', *The Guardian: MediaGuardian*, March 29.

JISC (2009), 'Press Release: Open Science - the future for research?', posting to the BOAI list, November 16, 2009.

Kerssens, N. and Dekker A. (2009), 'Interview with Lev Manovich for Archive 2020', *Virtueel\_ Platform*.

<http://www.virtueelplatform.nl/#2595>.

Knouf, N. (2010), 'The JJPS Extension: Presenting Academic Performance Information', *Journal of Journal Performance Studies*, Vol 1, No 1. Available at <http://journalofjournalperformancestudies.org/journal/index.php/jjps/article/view/6/6>. Accessed 20 June, 2010.

Latour, B (2004), 'Why Has Critique Run Out of Steam? From Matters of Fact to Matters of Concern"', *Critical Inquiry*, Vol. 30, Number 2.  
<http://www.uchicago.edu/research/jnl-crit-inq/issues/v30/30n2.Latour.html>.

Liu, A. (2011) 'Where is Cultural Criticism in the Digital Humanities'. Paper presented at the panel on 'The History and Future of the Digital Humanities'" Modern Language Association convention, Los Angeles, January 7. <http://liu.english.ucsb.edu/where-is-cultural-criticism-in-the-digital-humanities>.

Lyotard, J-F. (1986), *The Postmodern Condition: A Report on Knowledge*. Manchester: Manchester University Press.

Lyotard, J.-F. (1991) *The Inhuman: Reflections on Time*. Cambridge: Polity.

Manovich, L. (2010a) 'Cultural Analytics Lectures by Manovich in UK (London and Swansea), March 8-9, 2010', *Software Studies Initiative*. March 8.  
<http://lab.softwarestudies.com/2010/03/cultural-analytics-lecture-by-manovich.html>.

Manovich, L. (2011) 'Trending: The Promises and the Challenges of Big Social Data', *Lev Manovich*, April 28:

[http://www.manovich.net/DOCS/Manovich\\_trending\\_paper.pdf](http://www.manovich.net/DOCS/Manovich_trending_paper.pdf).

Meeks, E. (2010), 'The Digital Humanities as Imagined Community', *Digital Humanities Specialist*. September 14. <https://dhs.stanford.edu/the-digital-humanities-as/the-digital-humanities-as-imagined-community/>.

Mintel report, 'Energy Efficiency in the Home - UK - July 2010'.

Murray, S. Choi, S., Hoey, J., Kendall, C., Maskalyk, J., and Palepu, A. (2008), 'Open Science, Open Access and Open Source Software at *Open Medicine*', *Open Medicine*, 2(1): e1–e3.

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3091592/?tool=pmcentrez>

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3091592/pdf/OpenMed-02-e1.pdf??tool=pmcentrez>

Patterson, M. (2009), 'Article-Level Metrics at PLoS – Addition of Usage Data', *PLoS: Public Library of Science*. September 16.

<http://blogs.plos.org/plos/2009/09/article-level-metrics-at-plos-addition-of-usage-data/>.

Pubget (2010), '[BOAI] PLoS Launches Fast (Open) PDF Access with Pubget', posted on the BOAI list by Peter Suber, March 8.

Poynder, R. (2010), 'Interview With Jean-Claude Bradley: The Impact of Open Notebook Science', *Information Today*, September.

<http://www.infotoday.com/IT/sep10/Poynder.shtml>.

Scheinfeldt, T. (2008), 'Sunset for Ideology, Sunrise for

Methodology?', *Found History*, March 13.  
<http://www.foundhistory.org/2008/03/13/sunset-for-ideology-sunrise-for-methodology/>

Scheinfeldt, T. (2010a) 'Where's the Beef?: Does Digital Humanities Have to Answer Questions?', *Found History*, March 13.  
<http://www.foundhistory.org/2010/05/12/wheres-the-beef-does-digital-humanities-have-to-answer-questions/>.

Scheinfeldt, T. (2010b) response to Dan Cohen, 'Searching for the Victorians', *Dan Cohen*. October 5.  
[http://www.dancohen.org/2010/10/04/searching-for-the-victorians/?utm\\_source=feedburner&utm\\_medium=feed&utm\\_campaign=Feed%3A+DanCohen+%28Dan+Cohen%29&utm\\_content=Google+Reader](http://www.dancohen.org/2010/10/04/searching-for-the-victorians/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+DanCohen+%28Dan+Cohen%29&utm_content=Google+Reader).

Shields, R. (2010), 'Green Fatigue Hits Campaign to Reduce Carbon Footprint', *The Independent*, October 10. <http://www.independent.co.uk/environment/climate-change/green-fatigue-hits-campaign-to-reduce-carbon-footprint-2102585.html>.

Sterling, B. (2005), *Shaping Things*. Massachusetts: MIT.

Stolberg, S. G. (2009), 'On First Day, Obama Quickly Sets a New Tone"', *The New York Times*. January 21.  
<http://www.nytimes.com/2009/01/22/us/politics/22obama.html>.

Swan, A. (2009), 'Open Access and Open Data', *2nd NERC Data Management Workshop*, Oxford. February 17-18. <http://eprints.ecs.soton.ac.uk/17424/>.

Swartz, A. (2010), 'When is Transparency Useful?'

Aaron Swartz's *Raw Thought* blog, February 11.  
<http://www.aaronsw.com/weblog/usefultransparency>.

Vaidhyathan, S. (2009), 'The Googlization of Universities', *The NEA 2009 Almanac of Higher Education*.  
[http://www.nea.org/assets/img/PubAlmanac/ALM\\_09\\_06.pdf](http://www.nea.org/assets/img/PubAlmanac/ALM_09_06.pdf)

Wark, M. (2004), *A Hacker Manifesto*. Harvard: Harvard University Press.

Westcott, S. (2009) 'Global Warming: Brits Deny Humans are to Blame,' *The Express*, December 7.  
<http://www.express.co.uk/posts/view/144551/Global-warming-Brits-deny-humans-are-to-blame>

The White House (2009), 'Memorandum for the Heads of Executive Departments and Agencies: Transparency and Open Government'. January 21.  
[http://www.whitehouse.gov/the\\_press\\_office/TransparencyandOpenGovernment/](http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment/)

Wollen, P. and Cooke, L. eds (1995), *Visual Display: Culture Beyond Appearances*. Seattle: Bay Press.

# Articles

## Open Science

**It's An Open (Science), Open (Access), Open (Source),  
Open (Notebook) World**

[Open Notebook Science](#)

Patrick O. Brown, Michael B. Eisen, Harold Varmus  
[Why PLoS Became a Publisher](#)

Sally Murray, Stephen Choi, John Hoey, Claire Kendall,  
James Maskalyk, and Anita Palepu  
[Open Science, Open Access and Open Source Software at  
\*Open Medicine\*](#)

## Community Science

[BioCurious: A Community Lab for Biotechnology](#)

Richard Stallman  
[Free Community Science and the Free Development of  
Science](#)

**'This Revolution Will Be Digitized': Online Tools for  
Open Science**

[Biogang](#)

Bill Hooker  
[The Future of Science is Open, Part 3: An Open Science  
World](#)

Chris Patil and Vivian Siegel  
[This Revolution Will Be Digitized: Online Tools for Radical  
Collaboration](#)

## **Open Science Publishing**

Philip E. Bourne

*What Do I Want from the Publisher of the Future?*

Cameron Neylon

*Science in the Open/or/How I Learned to Stop Worrying and Love My Blog*

## **Open Knowledge**

### **Access to Knowledge**

Open Knowledge Foundation

Gaëlle Krikorian and Amy Kapczynski, eds

*Access to Knowledge In the Age of Intellectual Property*

### **New Models for Open Sharing and Open Research**

Anne H. Margulies

*A New Model for Open Sharing*

Thomas B. Kepler, Marc A. Marti-Renom, Stephen M.

Maurer, Arti K. Rai, Ginger Taylor, Matthew H. Todd

*Open Source Research - The Power of Us*

### **Open Knowledge and its Discontents**

J.J. King

*The Packet Gang: Openness and its Discontents*

Michael Gurstein

*Are the Open Data Warriors Fighting for Robin Hood or the Sheriff?: Some Reflections on OKCon 2011 and the Emerging Data Divide*



## **Open Data**

### **Data-Intensive Science**

Vincent S. Smith

[Data Publication: Towards a Database of Everything](#)

Tony Hey, Stewart Tansley, Kristen Tolle, eds

[Scholarly Communication, \*The Fourth Paradigm: Data-Intensive Scientific Discovery\*](#)

### **World of Data**

#### [Free Our Data](#)

Simon Rogers

[How Canada Became an Open Data and Data Journalism Powerhouse](#)

### **We Can Know It For You**

Omer Tene

[What Google Knows: Privacy and Internet Search Engines](#)

Daniel Chandramohan, Kenji Shibuya, Philip Setel, Sandy Cairncross, Alan D. Lopez, Christopher J. L. Murray, Basia Żaba, Robert W. Snow, Fred Binka

[Should Data from Demographic Surveillance Systems Be Made More Widely Available to Researchers?](#)

## **Digitize Me**

### **Encode Me/Decode Me**

[Human Genome Project](#)

The ENCODE Project Consortium

[A User's Guide to the Encyclopaedia of DNA Elements](#)

(ENCODE)

deCODEme

## **Life-Tracking**

Quantified Self

Gary Wolf

The Data-Driven Life

Aiden R. Doherty and Alan F. Smeaton

Automatically Augmenting Lifelog Events Using  
Pervasively Generated Content from Millions of People

Jennifer S. Beaudin, Stephen S. Intille, and Margaret E.  
Morris

To Track or Not to Track: User Reactions to Concepts in  
Longitudinal Health Monitoring

**The Neurological Turn: or, 'How the Internet Gets  
Inside Us'**

Adam Gopnik

The Information: How the Internet Gets Inside Us

N. Katherine Hayles

Hyper and Deep Attention: The Generation Divide in  
Cognitive Modes

**Visualize Me**

**What is Visualization?**

Lev Manovich

What is Visualization?

Nathan Yau

Data Visualization Meets Game Design to Explore your

Digital Life

Bloom

## **Mood-mapping**

Celeste Biever

Twitter Mood Maps Reveal Emotional States of America

Twitter mood video

Moodscope

Mappiness

## **The Visualized Human (or, The Human As Spectacle)**

Nicholas Felton

The Annual Felton Report

Deb Roy

The Birth of a Word

Johanna Drucker

Humanistic Approaches to the Graphical Expression of Interpretation

## **Search Me**

### **Search-Engine Science**

Emily H. Chan, Vikram Sahai, Corrie Conrad, and John S. Brownstein

Using Web Search Query Data to Monitor Dengue Epidemics: A New Model for Neglected Tropical Disease Surveillance

Annie Y.S. Lau, Enrico Coiera, Tatjana Zrimec, and Paul

Compton

Clinician Search Behaviors May Be Influenced by Search Engine Design

## **The Science of Control**

Alession Signorini Alberto Maria Segre, Philip M. Polgreen  
The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. During the Influenza A H1N1 Pandemic

David Parry

*Surveillance*

Felix Stalder and Christine Mayer

The Second Index: Search Engines, Personalization and Surveillance (Deep Search)

## **Deep Search**

Michael K. Bergman

The Deep Web: Surfacing Hidden Value

Clare Birchall

The Invisible Web, *The In/Visible*

## **Media Gifts?**

Web 2.0 Suicide Machine

Freedom Box Foundation

Traceblog

The JJPS Firefox Extension

# Appendix

Karl Popper, [The Open Society and its Enemies](#)

# Attributions

## Open science/Attributions

Brown, P.O., Michael B. Eisen, M.B., Varmus, H. (2003), 'Why PLoS Became a Publisher', *PLoS Biology* 1(1): e36.

<http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.0000036> Copyright: © 2003 Public Library of Science. This is an open-access article distributed under the terms of the Public Library of Science Open-Access License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Murray, S., Choi, S., Hoey, J., Kendall, C., Maskalyk, J., and Palepu, A. (2008), 'Open Science, Open Access and Open Source Software at *Open Medicine*', *Open Medicine*, 2008; 2(1): e1–e3.

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3091592/?tool=pmcentrez> *Open Medicine* applies the Creative Commons Attribution Share Alike License, which means that anyone is able to freely copy, download, reprint, reuse, distribute, display or perform this work and that authors retain copyright of their work. Any derivative use of this work must be distributed only under a license identical to this one and must be attributed to the authors. Any of these conditions can be waived with permission from the copyright holder.

Stallman, R. (2005), Free Community Science and the Free Development of Science. *PLoS Medicine* 2(2): e47. <http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.0020047>

[1371%2Fjournal.pmed.0020047](#) Copyright: © 2005 Richard Stallman. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hooker, B. (2007), 'The Future of Science is Open, Part 3: An Open Science World', *3 Quarks Daily* [blog] January 22, 2007. Link to *3 Quarks Daily*: [http://3quarksdaily.blogspot.com/3quarksdaily/2007/01/the\\_future\\_of\\_s.html](http://3quarksdaily.blogspot.com/3quarksdaily/2007/01/the_future_of_s.html).

Patil, C. and Siegel, V. (2009), 'This Revolution Will Be Digitized: Online Tools for Radical Collaboration', *DMM: Disease Models and Mechanisms*, May-June 2(5-6): 201–205. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2675795/> From 2011, all articles published in *DMM: Disease Models and Mechanisms* are OPEN ACCESS: available immediately upon publication with unrestricted online access, published under a Creative Commons licence and deposited in PubMed Central for immediate release.

Bourne, P.E. (2010), 'What Do I Want from the Publisher of the Future?', *PLoS Computational Biology*, May 6(5): e1000787. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2877727/?tool=pmcentrez#pcbi.1000787-Hey1> Copyright Philip E. Bourne. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution,

and reproduction in any medium, provided the original author and source are credited.

Neylon C. (2008), 'Science in the Open/or/How I Learned to Stop Worrying and Love My Blog', *Collection Science in the 21st Century, Perimeter Institute Recorded Seminar Archive (PIRSA)*. Link to PIRSA: Perimeter Institute Recorded Seminar Archive: <http://pirsa.org/08090038/>

Krikorian, G. and Kapczynski, A. eds (2010), *Access to Knowledge In the Age of Intellectual Property*. New York: Zone Books.  
[http://www.soros.org/initiatives/information/focus/access/articles\\_publications/publications/age-of-intellectual-property-20101110/age-of-intellectual-property-20101110.pdf](http://www.soros.org/initiatives/information/focus/access/articles_publications/publications/age-of-intellectual-property-20101110/age-of-intellectual-property-20101110.pdf) This work is published subject to a Creative Commons by-nc-nd license: You may not alter, transform, or build upon this work.

Margulies, A.H. (2004), 'A New Model for Open Sharing: Massachusetts Institute of Technology's OpenCourseWare Initiative Makes a Difference'. *PLoS Biology* 2(8): e200.  
<http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.0020200> Copyright: © 2004 Anne H. Margulies. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Kepler, T.B., Marti-Renom, M.A., Maurer, S.M., Rai, A.K., Taylor, G, Todd, M.H. (2006) 'Open Source Research - The Power of Us', *Australian Journal of Chemistry*, 59:291–294.

[http://www.publish.csiro.au/view/journals/dsp\\_journal\\_fulltext.cfm?nid=51&f=CH06095](http://www.publish.csiro.au/view/journals/dsp_journal_fulltext.cfm?nid=51&f=CH06095) Authors in *Australian Journal of Chemistry* may choose to publish their papers Open Access. For all papers available Open Access, you are free to share - to copy, distribute and transmit the work - under the following conditions.

Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work). Non Commercial. You may not use this work for commercial purposes. No Derivative Works. You may not alter, transform, or build upon this work.

Note that: For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to

<http://creativecommons.org/licenses/by-nc-nd/3.0/>

King, J.J. (2004), 'The Packet Gang: Openness and its Discontents', *Mute: Culture and Politics After the Net*, January 12. <http://www.metamute.org/en/The-Packet-Gang> Made available to the public domain without copyright licence.

Gurstein, M. (2011), 'Are the Open Data Warriors Fighting for Robin Hood or the Sheriff?: Some Reflections on OKCon 2011 and the Emerging Data Divide', *Gurstein's Community Infomatics*, July 3. Link to *Gurstein's Community Infomatics*:

<http://gurstein.wordpress.com/2011/07/03/are-the-open->

[data-warriors-fighting-for-robin-hood-or-the-sheriff-some-reflections-on-okcon-2011-and-the-emerging-data-divide/](#)

Smith, V. S. (2009), 'Data Publication: Towards a Database of Everything', *BMC Research Notes*, 2: 113. <http://www.biomedcentral.com/1756-0500/2/113> © 2009 Smith, licensee BioMed Central Ltd. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hey, T., Tansley, S., Tolle, K., eds (2009) 'Part 4: Scholarly Communication', *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Seattle: Microsoft Research. Link to *The Fourth Paradigm: Data-Intensive Scientific Discovery*: [http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th\\_paradigm\\_book\\_part4\\_complete.pdf](http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part4_complete.pdf).

Rogers, S. (2010), 'How Canada Became an Open Data and Data Journalism Powerhouse', *The Guardian: Datablog*, November 9. Link to [www.guardian.co.uk](http://www.guardian.co.uk): <http://www.guardian.co.uk/news/datablog/2010/nov/09/canada-open-data>

Tene, O. (2008) 'What Google Knows: Privacy and Internet Search Engines', *Utah Law Review*, Vol. 2008, No 4.

<http://epubs.utah.edu/index.php/ulr/article/viewArticle/136> Copyright by the Utah Law Review Society (ISSN 0042-1448). In addition to the rights to copy that are granted by law, the owner of the copyright of each article published, except as expressly otherwise noted, grants permission for copies of that article to be made and used for educational purposes by nonprofit educational institutions, provided that the author and journal are identified, that proper notice of copyright is affixed to each copy of the article, and that copies are distributed at or below cost.

Chandramohan, D., Shibuya, K., Setel, P., Cairncross, S., Lopez, A.D., et al. (2008) 'Should Data from Demographic Surveillance Systems Be Made More Widely Available to Researchers?' *PLoS Medicine* 5(2): e57.

<http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.0050057> Copyright: © 2008 Chandramohan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The ENCODE Project Consortium (2011), 'A User's Guide to the Encyclopaedia of DNA Elements (ENCODE)', *PLoS Biology*, April; 9(4): e1001046.  
<http://pubmedcentralcanada.ca/articlerender.cgi?accid=PMC3079585&tool=pmcentrez> Copyright: The ENCODE Project Consortium. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any

medium, provided the original author and source are credited.

Wolf, G. (2010), 'The Data-Driven Life', *New York Times*, April 28. Link to [www.nytimes.com](http://www.nytimes.com):  
<http://xrl.us/bh3d4g>

Doherty, A. R. and Smeaton, A. F. (2010), 'Automatically Augmenting Lifelog Events Using Pervasively Generated Content From Millions of People', *Sensors*, 10 (3).  
<http://doras.dcu.ie/15300/1/Sensors-03-154-Doherty-ie-edited.pdf> Copyright: © 2010 by the authors; licensee Molecular Diversity Preservation International, Basel, Switzerland. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license  
<http://creativecommons.org/licenses/by/3.0/>.

Beaudin, J.S., Intille, S.S. , and Morris, M.E. (2006), 'To Track or Not to Track: User Reactions to Concepts in Longitudinal Health Monitoring', *Journal of Medical Internet Research*, October-December; 8(4): e29.  
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1794006/?tool=pmcentrez#ref1> Copyright: © Jennifer S. Beaudin, Stephen S. Intille, Margaret E. Morris. Originally published in the *Journal of Medical Internet Research* (<http://www.jmir.org>), 07.12.2006. Except where otherwise noted, articles published in the *Journal of Medical Internet Research* are distributed under the terms of the Creative Commons Attribution License (<http://www.creativecommons.org/licenses/by/2.0/>),

which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited, including full bibliographic details and the URL (see "please cite as" above), and this statement is included.

Gopnik, A. (2011), 'The Information: How the Internet Gets Inside Us', *The New Yorker*, February 14. Link to [www.newyorker.com](http://www.newyorker.com):  
[http://www.newyorker.com/arts/critics/atlarge/2011/02/14/110214crat\\_atlarge\\_gopnik](http://www.newyorker.com/arts/critics/atlarge/2011/02/14/110214crat_atlarge_gopnik)

Hayles, N. K. (2010), 'Hyper and Deep Attention: The Generation Divide in Cognitive Modes', *Posthuman Destinies*, November 24. Link to *Posthuman Destinies*:  
<http://www.sciy.org/2010/11/24/hyper-and-deep-attention-the-generational-divide-in-cognitive-modes-by-n-katherine-hayles/>

Manovich, L. (2010), 'What is Visualization?', *Poetess Archive Journal*, December, 2:1.  
<http://paj.muohio.edu/paj/index.php/paj/article/view/19/58> Copyright: © Lev Manovich 2010. While authors retain copyright of their essays, *Poetess Archive Journal* journal provides immediate open access to its content on the principle that making research freely available to the public supports a greater global exchange of knowledge.

Yau, N. (2011), 'Data Visualization Meets Game Design to Explore Your Digital Life', *FlowingData*, February 23. Copyright: © 2007-2011 by *FlowingData*.

All rights reserved. Made available here via link to *FlowingData*: <http://flowingdata.com/2011/02/23/data-visualization-meets-game-design-to-explore-your-digital-life/>

Biever, C. (2010) 'Twitter Mood Maps Reveal Emotional States of America', *The New Scientist*, July 21, 207(2771): 14. Copyright: © Copyright Reed Business Information Ltd.. Made available here via link to [www.newscientist.com: http://www.newscientist.com/article/dn19200-twitter-mood-maps-reveal-emotional-states-of-america.html](http://www.newscientist.com/article/dn19200-twitter-mood-maps-reveal-emotional-states-of-america.html).

Chan, E.H., Sahai, V., Conrad, C., and Brownstein, J.S. (2011), 'Using Web Search Query Data to Monitor Dengue Epidemics: A New Model for Neglected Tropical Disease Surveillance', *PLoS Neglected Tropical Diseases*, May; 5(5): e1206. <http://pubmedcentralcanada.ca/picrender.cgi?accid=PMC3104029&blobtype=pdf&tool=pmcentrez> Copyright: Chan 'et al. *This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

Lau, A.Y.S., Coiera, E., Zrimec, T., and Compton, P. (2010), 'Clinician Search Behaviors May Be Influenced by Search Engine Design', *Journal of Medical Internet Research*, April-June, 12(2): e25. <http://pubmedcentralcanada.ca/articlerender.cgi?accid=PMC2956236&tool=pmcentrez> Copyright: © Annie YS Lau, Enrico Coiera, Tatjana Zrimec, Paul Compton.

This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the *Journal of Medical Internet Research*, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

Signorini, A., Segre, A.M., Polgreen, P.M. (2011), 'The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic', *PLoS ONE* 6(5): e19467. <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0019467>. Copyright: © 2011 Signorini et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Stalder, F. and Mayer, C. (2009), 'The Second Index: Search Engines, Personalization and Surveillance', in Konrad Becker and Felix Stalder (eds.), *Deep Search: The Politics of Search Beyond Google*, Innsbruck/New Jersey: Studienverlag/Transaction Publishers, 98-116. Link to Felix Stalder's *Notes and Nodes* blog, February 10, 2010: <http://felix.openflows.com/node/113>

Bergman, M.K. (2001), 'The Deep Web: Surfacing Hidden Value', *JEP: The Journal of Electronic Publishing*, vol.7, no.1, August.

<http://quod.lib.umich.edu/cgi/t/text/text-idx?c=jep;view=text;rgn=main;idno=3336451.0007.104>.

Copyright: © 2010, Regents of the University of Michigan. *JEP: The Journal of Electronic Publishing* is a product of MPublishing. Except where otherwise noted, this work is subject to a Creative Commons Attribution 3.0 license.