

Coventry University Repository for the Virtual Environment
(CURVE)

Gary Hall

We can know it for you: the secret life of metadata

Author final draft

Original citation:

Hall, G. (2010) 'We can know it for you: the secret life of metadata.' [Catalogue essay from the *How We Became Metadata* Exhibition, Institute for Modern and Contemporary Culture, University of Westminster, May to July, 2010] University of Westminster: London.

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the author's final draft of a catalogue essay for the *How We Became Metadata* Exhibition held at the Institute for Modern and Contemporary Culture, University of Westminster, May to July, 2010. Some differences between the printed catalogue version and this version may remain and you are advised to consult the printed catalogue if you wish to cite from it.

This version is available from: <http://curve.coventry.ac.uk/open>

Available in the CURVE Research Collection: April 2010

Gary Hall

We Can Know It for You: The Secret Life of Metadata

Thirty years ago, the philosopher Jean-François Lyotard argued that science, lacking the resources to legitimate *itself* as true, had come to rely for its legitimacy on precisely the kind of knowledge it did not even consider to *be knowledge*: namely, non-scientific narrative knowledge. Specifically, it was philosophy's role to produce a discourse of legitimation for science in the form of narratives such as those of the Enlightenment, progress, modernity and the creation of wealth. Lyotard's intention was not to position philosophy as ultimately being able to tell us more about science than science itself, but to emphasize that, in a process of transformation which had been taking place since at least the end of the 1950s, such long-standing metanarratives had become obsolete.

So what does science do if the narratives that legitimate it are no longer credible? Lyotard's answer was that science increases its connection to society by optimizing the relationship 'between input and output'. For Lyotard, writing in 1979, technological transformations in research and the transmission of knowledge, including the widespread use of computers and databases, were already in the process of exteriorizing knowledge in relation to the 'knower'. Lyotard saw this as producing a major alteration in the status and nature of knowledge: *away* from questions of what is just, right and scientifically true, and *toward* a concern merely with 'optimizing the system's performance'.¹

Scroll down thirty years and we do indeed find many discourses in the sciences today being taken up with exteriorizing knowledge and information in order to maximise the system's performance by eliminating delays and inefficiencies, and solving technical problems. Witness those studies arguing that the 'open access' academic publishing model championed by the sciences, whereby peer-reviewed scholarly research is made available online for free, is actually the most cost-effective mechanism for scholarly publishing. Indeed, there is a rapidly expanding body of literature detailing the various increases 'open access' makes possible in the amount of research material that can be published, searched and stored, the number of people who can access it, the impact of that material and the range of its distribution. Even the data created in the course of scientific research is being made openly available for others to use, analyse and build upon. Known as Open Data, this initiative is presented as bestowing data with a 'vastly increased utility': openly published digital datasets are 'easily passed around'; they are 'more easily reused'; and they contain more 'opportunities for educational and commercial exploitation'.²

Yet this alteration in the status and nature of knowledge is not confined to science - or even the academy. The current global financial crisis has only served to add greater emphasis to the belief of many in the UK that the government should relinquish its copyright on data gathered with taxpayers' money - most notoriously that relating to MPs expenses - and make it openly available to the public online. From a liberal perspective, freeing government funded and collected data helps society perform more efficiently by leading to an increase in citizen participation in democracy, as access to information is no longer restricted either to the state or to those corporations, organizations and individuals who have sufficient money and power to acquire it for themselves. But neoliberals likewise perceive making the data freely and openly available to the public as a means of getting more out of the system for less. From this viewpoint such communicative transparency is held as ensuring greater value

for money, enabling costs to be distributed more effectively, reducing bureaucracy and paperwork while also increasing choice, innovation, enterprise, creativity and competitiveness.

In fact, to have participated in the shift away from questions of what is just and right, and toward a concern merely with optimizing the relation between input and output, you don't need to have actively contributed to the movements for 'open access' or free data at all. If you are one of the 1.3 million plus people who have purchased a Kindle e-book reader, then you've already signed a license agreement allowing Amazon - 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 23 million in the UK and 350 million worldwide who use Facebook, then you're already labouring for free, not only to help the owners of this social network make a reputed \$1 billion a year from demographically targeted advertising, but to supply law enforcement agencies with profile data about yourself, your family, friends and colleagues they can use in investigations.³ Even if you have done neither, you will in all probability have provided Google with a host of free data 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 via Google Street View, Google Book Search and YouTube, which Google also owns.

Obviously, 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 effort to identify patterns in your life – what's called life-tracking. Nevertheless, for most people, refusing to take part in this transformation of knowledge and information into quantities of data is not really an option. 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'.⁴ Besides, such notions of refusal and resistance have their basis in a conception of the autonomous, rational, self-identical humanist subject that these developments in digital media culture and technology may be in the process of reconfiguring. As a result, they risk overlooking the manner in which computers, databases, archives, servers, blogs, microblogs, RSS feeds, image and video-sharing, social networking and 'the cloud' are not just being used to change the status and nature of knowledge; they may be involved in the constitution of a new form of human subject.⁵

Indeed, could the move toward supplying ever more research, information and data online for free be part of the development of what Deleuze dubbed a society of control? Here we are no longer subject primarily to those closed, disciplinary modes of power Michel Foucault described in books such as *Discipline and Punish*. Increases in computer processing capacity and the associated availability of large data sets have instead enabled a degree of data mining and pattern recognition to be achieved that makes it possible to automatically anticipate and predict – and thus control, albeit in a relatively open way – actions on the part of the subject *before* they actually take place. Think of the way Google News aggregates 'headlines from news sources worldwide, groups similar stories together and displays them according to each reader's personalized interests'.⁶

From this perspective it's crucial to study such technology platforms, media tools and spaces, and to understand how they work and are used. But it's also crucial to investigate, analyze and interrogate them critically. It's important, for example, to be aware that just 'three companies - Google, Yahoo! and Microsoft - handle 95 percent of all search queries'; that 'for searches containing the name of a specific political organisation, Yahoo! and Google

agree on the top result 90 percent of the time’;⁷ and that conventional search engines actually reach only a tiny fraction of the total number of available web pages.

So it’s not enough to simply ‘Free Our Data’,⁸ or to operate on the basis that ‘information wants to be free’. We can put more and more research and data online, we can make it available under open access and open data 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 it’s going to be found*. Ideas of this kind ignore the fact that all information and data is ordered, structured, selected and framed. This is what metadata is for, after all. Metadata is *data about data*. It’s information or data that describes, links to, or is otherwise used to control, order, structure, find, sort, select, filter, access, 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 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. 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’s not *just* a question of free and open access to the research and data. It’s also a question of who (and what) makes decisions regarding the metadata, and on what basis such decisions are made. To paraphrase Lyotard, ‘Who decides what metadata is, and who knows what needs to be decided?’ Will the ‘ruling class’ – ‘corporate leaders, high-level administrators’ and so on – continue to be ‘the class of decision makers’ with regard to having ‘access to the information these machines have in storage’, but also in terms of creating and controlling the metadata?⁹ Or does the fact these social relations are the result not of objective and immutable historical, scientific or technological processes, but of contingent, pragmatic yet temporary decisions involving power, conflict and violence, mean that they can be disarticulated and transformed, and that new sets of social relations can be established?

To this end, is it possible to generate a plurality of *different*, and at times conflicting and even incommensurable possibilities for the creation, organization, presentation and control of data and metadata? Rather than just opening the already existing ‘memory and databanks’ to the people – the line Lyotard suggested we follow – could there not in the future also be a multitude of online, open, ‘counter-institutional’ platforms, tools, databases and other media experiments capable of maintaining a much needed level of opacity, noise, error, feedback, delay, antagonism and dissensus within the system?

Is this exhibition ‘How We Became Metadata’ a move in that direction?

¹ Jean-François Lyotard, *The Postmodern Condition: A Report on Knowledge* (Manchester: Manchester University Press, 1986) p.11, 4, p.xxiv.

² Alma Swan, ‘Open Access and Open Data’, 2nd NERC Data Management Workshop, Oxford, 17-18 February 2009. Available at

[http://eprints.ecs.soton.ac.uk/17424/1/Swan - NERC_09.pptx](http://eprints.ecs.soton.ac.uk/17424/1/Swan_-_NERC_09.pptx).

[http://209.85.229.132/search?q=cache:fxpeg9SnIQYJ:eprints.ecs.soton.ac.uk/17424/1/Swan - NERC_09.pptx+open+data+journals&cd=7&hl=en&ct=clnk&gl=uk&client=firefox-a](http://209.85.229.132/search?q=cache:fxpeg9SnIQYJ:eprints.ecs.soton.ac.uk/17424/1/Swan_-_NERC_09.pptx+open+data+journals&cd=7&hl=en&ct=clnk&gl=uk&client=firefox-a)

³ Marcia Hoffman, 'EFF Posts Documents Detailing Law Enforcement Collection of Data From Social Media Sites', *Electronic Frontier Foundation*, 16 March, 2010. Available at <http://www.eff.org/deeplinks/2010/03/eff-posts-documents-detailing-law-enforcement>.

⁴ Gilles Deleuze and Felix Guattari, *A Thousand Plateaus: Capitalism and Schizophrenia* (London: Athlone, 1988) p.492.

⁶ Google News, 'About Google News', http://news.google.com/intl/en_us/about_google_news.html . Accessed 18 March, 2010.

⁷ Matthew Hindman *The Myth of Digital Democracy* (Princeton, NJ and Oxford: Princeton University Press, 2009) p.59, 79.

⁸ This is a reference to *The Guardian* newspaper's 'Free Our Data' campaign. See Charles Arthur and Michael Cross, 'Give Us Back Our Crown Jewels', *The Guardian*, 9 March, 2006. Available at: <http://www.guardian.co.uk/technology/2006/mar/09/education.epublic>

⁹ Lyotard, *The Postmodern Condition*, p. 9, 14, 14.