William Uricchio

William Uricchio

on the Colonization of the Data-Imaginary in the Public Service Sector

KEYWORDS: data epistemologies; data imaginaries; data literacy; public service institutions; television history

Among media historians, William Uricchio, professor of Comparative Media Studies at MIT and emeritus professor at Utrecht University, is known primarily for his work on the histories of all kinds of (formerly) ‘new’ media. In recent years, he has been preoccupied, among many other topics, with the cultural use of algorithms, specifically in such settings as public archives or by such institutions as public broadcasters.

In a conversation on the topic, he talks about his concerns for the present and future of data use. In doing so, he also builds on some long-term interests: in imaginations of television present and past, and in forms of ‘Americanization’ in the media.

– Eef Masson and Karin van Es

William Uricchio.

EM/KvE: In your recent reflections on big data use by public-facing organizations, you have been critical of a so-called ‘failure of imagination’, arguing that it entails the threat of a ‘colonization of the data-imaginary’. Can you explain what this threat entails, and point to where we might have seen it before?

WU: By ‘imaginary’, I mean the range of possible conceptions, deployments, and meanings that something might have. This use of the term echoes Donald Norman’s notion of ‘affordances’, but goes beyond it in the sense of embracing the very conceptualization of something, as well as its semiotic constraints.¹ The semiotic is crucial, because language marks the conceptual borders that we take as givens, and in the case of ‘data’, they are slippery. Do we mean by ‘data’ literally only the information, in whatever form, that will be organized and processed? Do we include the larger epistemic system within which that information is given form? Do we also include its processing within a larger set of, say, algorithmic operations, using the term ‘data’ synecdochally? Or do we use ‘data’ metaphorically to characterize a particular claim? Only when we pin down what, precisely, we include in the term, can we move on to talk about possible conceptions and deployments.

From pluriform to uniform – that’s the developmental trajectory that we can see in most technologies, regardless of culture. Television offers a good example. From the late nineteenth century onwards, a technological ensemble capable of sending and receiving live image and sound was imagined in a wide number of ways. First and foremost, it was an extension of the telephone, a means of point-to-point connectivity. By the mid-1930s, it was deployed as large-screen cinema-style entertainment, as a domestic companion, as a means of aerial reconnaissance, as an element in a remote-control system, and indeed, as a means of point-to-point communication.² At least in Germany, all systems worked, all had backers and business models. Cut to the early 1950s, and the world was left with basically just one model of television. And wherever in the world television appeared, regardless of ideology, regardless of economic development, that single model dominated. The televisual imaginary was taken over, dominated, colonized by a particular notion of the medium.

When I refer to the ‘colonization of the data imaginary’, this is what I mean. Regardless of the institutional or cultural setting, data mean pretty much the same thing – in commercial broadcasting or public service broadcasting, advertising agencies or public archives. There are of course differences in use that come with the agendas of various organizations, and some academic programs in universities have certainly been asking critical questions (even if university administrations stick to ‘data as usual’!). And just as it was in television’s case, the reason for the overarching conformity is professionalism. Jérôme Bourdon has a wonderful essay on the ‘Americanization’ of European television, a process that he describes as ‘self-inflicted’.³ He makes a strong case for the medium’s colonization (my term – I don’t mean to saddle him with it) by pointing out the usual subjects – the exercise of post-war American hegemony, whether through ‘education’ programs or market domination. But the most insidious part of the process was indeed self-inflicted: television makers wanted to ‘do it right’, to have ‘the right’ gear, ‘the proper’ lighting arrangements, to be, in a word, professional. But the larger questions regarding what the medium might be, where it fit within the pluriform options that preceded the 1950s,

how public service conceptions of the medium could differ, and so on – those questions were not even asked in the scramble of television's proponents to professionalize. Indeed, the people involved in making television, whether on a technological or program level, could easily move from public service to commercial and back again: their notions of the medium were identical, and the only really significant element of career differentiation was the perception of their professionalism. And that's exactly what has happened to data in the era of 'big data'.

EM/KvE: Can you explain why we should be concerned about this failure of imagination? And, once again, what we can learn from the past here?

WU: The more powerful our tools, the more critical and creative we need to be about putting them to work. When their uses become routinized and taken for granted, we lose critical distance. Powerful tools – certainly in our era, but you could go back to the industrial era and make the same argument – usually serve the interests of powerful social and economic elites. Those elites sometimes fund the development of the tools, but more often, they simply use their financial and political power to acquire them. And since, once acquired, social orders tend to use such constructions to sustain themselves, the more powerful the tool, the more powerful the ensuing inequities. The problem is that these tools, these technologies, usually begin as disruptive but soon become tamed, taken-for-granted. And whatever social and economic inequities are hard-baked into this now-reified deployment take on the appearance of 'normality'. This 'ordinary', almost invisible expression of the technological status quo is precisely what critical academics need to call out and question.

So, one answer – the dismal one – is that the failure to imagine something more critical and creative defaults to the generally oppressive and annoying uses of data that abound today, whether as predictive policing or marketing or political manipulation. But another answer is that, particularly in an era where data generation is as abundant as processing capacity, the failure of imagination comes with an opportunity cost: how else might we envision data and their use? How might we better restore representational complexity and nuance? How might we more effectively see emergent patterns? These are questions that public-facing organizations – public service broadcasters, public archives, educational organizations, and governments on all levels – should be asking.

Over the long haul of human history, data have essentially been abstractions of a systematic body of information. This abstraction has manifested itself in notches on sticks, charts full of numbers, holes in punch cards, or our somewhat more invisible systems today. 'Raw Data is an Oxymoron', as Lisa Gitelman so aptly put it in the title of her volume.⁴ The selection and collection of some set of experiences, further reduced by the selection of measurable attributes of those experiences, together with the process of their abstraction into notches or numbers, can certainly lead to a traceable connection to the world. But the nature of that connection and its implications are highly variable. Like the alphabets we write with, these are deeply cultural processes. And just as alphanumeric literacy was not well-spread in the thirteenth century, thus imparting power to those who commanded it, so too 'data literacy' in our time. It is not well-spread, it bears the aura of objectivity to those ignorant of its workings, and thus imparts power to those who control it.

If data are constructed and to some extent arbitrary, there is also slippage in terms of what cultures actually do with them. The tally stick of the middle ages (with variants going back to the Palaeolithic age) both recorded and verified information.⁵ Proto-demographer John Graunt, circa 1663, assembled points of information as a way to diagnose a situation and to make predictions about its development. And Herman Hollerith tabulated and summarized massive amounts of census data on punch cards as early as 1880. Recording, predicting, summarizing 'data', whatever they may represent, have historically been put to work for an array of purposes.

The Hollerith electrical tabulating machine with card reader and sorting machine. Scientific American, 30 August 1890. Source: www.officemuseum.com.

I don't mean to give a superficial history of data and their uses; rather, I simply want to call attention to their flexibility and their undulations over time. We have a tendency to take the passage of time as inherently progressive, so perhaps these 'early' examples are steps on the road to today's advances. But I take a different lesson. Like words, data tell stories; data represent particular points of view; and as I just mentioned, data have a fundamental role in the maintenance of particular social and economic orders. Just as the place of the word changed around the mid-fifteenth century with the printing press, enabling a redistribution of knowledge and power, perhaps we can hope for more pervasive data literacy and with it, a more equitable distribution of resources. But until then, our notion of data is bound to a particular social order. And exposing that process and redistributing the power to generate and use data critically are the tasks of our public-facing organizations – and the good work that groups like the Utrecht Data School are doing!

EM/KvE: Can you expand a little on how professionalism and professionalization figure into the colonization of the data-imaginary in the public sector?

WU: One of my concerns regarding professionalization in this sector is that the public defers its understandings of what data are and how to work with them to a 'priesthood' of experts. Whether employed by a commercial company or a public service organization or even a radical political organization, data specialists will generally agree about what constitute appropriate data and how to work with them. Each group may seek to have its way by lobbying conflicting datasets at one another, but they will almost always agree about the nature of the data and the appropriateness of various deployments. Each organization's data specialists will have studied in the same array of academic programs, and indeed, they may move among radically divergent organizations, just like the computer specialists or accountants, since their main distinction is their degree of professionalization – not their philosophy of data or their critique of its various parameters.

I'm intrigued by how data have managed to enter our culture with the same allure, and to be enveloped in the same assumptions, as – say – mathematics. There are any number of plausible reasons for this, ranging from the authoritative institutional nature of most data sources, to the complex algorithmic formulas that are bound up in their processing, to the need for systematically processing many inputs that complex social systems bring with them. Whatever the motive, and barring a few outstanding exceptions, we as a culture seem to turn a blind eye to the interpretative process inherent in data definition and collection, as well as to the constructed nature of their processing, preferring instead to take them as given.

And, to reiterate the earlier point, this is what I mean by the colonization of the data imaginary. Data, which have been understood historically in radically different ways, which have reflected cultural difference, have finally been dominated by a particular order of expertise, with professionals as their agents. In the conflicts that emerge from differently aligned organizations – say, commercial versus public service broadcasters – we rarely see debates over the nature and meaning of data. The data imaginary – what we take data to be, how we understand their relation to the world, how they can be analysed and what they are capable of

showing – enjoys something approximating the certainty of mathematics. We have by and large stopped looking at data as representation, subject to point of view, arbitrariness, partiality, and even contingency; and instead, our social institutions have endowed them the certainty of facts.

EM/KvE: Do you have any suggestions for alternative uses of data in the public service sector?

WU: I do! One thing we might do is to track and learn from the ongoing debate about Artificial Intelligence (AI). While it seems to enjoy many of the same qualities as ‘big data’ (a priesthood of experts, a massive buzz, a slippery definition and pattern of deployment...), in fact it is far more extreme. Its buzz has led to a boom-and-bust pattern of development over the past half-century; its definitions are seriously fraught even among experts; but most importantly, the very foundations upon which the reigning concept is built seems to be shifting, and quickly. In fact, we might be close to what Thomas Kuhn called a ‘paradigm shift’ in our notion of AI.⁶ This is all to say that the story of AI also reveals some disruptive dynamics that we can learn from. When fault-lines appear, who should we watch – scientists? Investors? The press? Given the role of ‘big tech’ in both the AI and big data domains, we have much to learn about their triggers and how to effect conceptual change in areas that are presented to the public as givens.

A second and related effort regards expanding the public’s literacy in the data sector, and a key means to do this – and a goal in doing it – is to give the public more control over the data that they generate and a greater say in how data are used. Critical awareness of different data sources and uses can help to inform personal choices and public debate, of course; but they are also vital if people are to see added value in differentiations among data regimes. If the commercial sector generates some share of its profits by harvesting data, will a public service sector that supports data privacy, transparency, and literacy be seen as having any added value? And will it be sufficient value to ‘push back’ against the political and professional pressures to harvest and track data as a default? That’s as much the project of literacy as it is understanding the triggers that enable ‘big tech’ to change its ways. And it’s a great project for our public-facing organizations.

To be clear, although harkening back to the privacy of the good old days is one way to go, I think that there are some creative – and critical – uses of data to be had within the public sector. We need to have a much better data-based sense of groups that are under-served and under-represented in media, something that should emerge from worrying less about how to compete with commercial interests on their terms, and more about how to fulfil a public remit. Data offer ways to make members of the public aware of additional resources that might directly serve their interests, whether from government services, archives, or educational organizations. Imagine, if you opted in to such a service, having a trip to a museum use sites of interest (places that you might indicate or stop for a certain amount of time) as a basis to generate a personalized website of links to television and radio programs, other exhibits, and even archival holdings. We have so far failed to develop a data-based interface between the public and the many resources that we have inherited from public past.

Big data easily scratch the itch of Apophenia. Any large dataset yields patterns. The question is, are the patterns answers, or places to begin asking questions? Too many big data applications are concerned with the former; but I think the place for humanists to stake their ground is with the latter, with asking why? What does it mean? And what if...?

Berlin 'Fernsehtuben' (source unknown).

Notes

1. Donald Norman, *The Design of Everyday Things* [original title: *The Psychology of Everyday Things*] (1988; Cambridge: MIT Press, 1998).
2. William Uricchio, "Television's First Seventy-Five Years: The Interpretive Flexibility of a Medium in Transition," *The Oxford Handbook of Film and Media Studies*, ed. Robert Kolker (Oxford: Oxford University Press, 2008): 286–305.
3. Jérôme Bourdon, "Imperialism, Self-inflicted? On the Americanizations of Television in Europe," in *We Europeans? Media, Representations, Identities*, ed. William Uricchio (Chicago: University of Chicago Press, 2008): 93–108.
4. Lisa Gitelman, ed., *'Raw Data' Is an Oxymoron* (Cambridge: MIT Press, 2013).
5. For an example, see "Tally Sticks," Collection Items, British Library, accessed 20 August 2018, <https://www.bl.uk/collection-items/tally-sticks>.
6. Thomas S. Kuhn, *The Structure of Scientific Revolutions* (1962; Chicago: University of Chicago Press, 2012).

Biography

William Uricchio is professor of Comparative Media Studies at MIT. As founder and principal investigator of MIT's Open Documentary Lab, he focuses in his scholarly research on the interplay of media technologies and cultural practices in relation to representation, knowledge and publics. A specialist in old media when they were new, he explores such things as early nineteenth-century conjunctures between photography and telegraphy; the place of telephony in the development of television at the other end of the nineteenth century; and the work of algorithms in our contemporary cultural lives. William has held professorships in Utrecht, where he is emeritus professor, Sweden (Stockholm), Germany (FU Berlin, Marburg), Denmark (national DREAM professor) and China (China University of Science & Technology).