

Trust in Techno-images: Early Media Collections as Precursors of Big Data

Abstract

This article proposes a consideration of today's discourses on 'big data' from a media archaeological point of view, confronting such discourses with those surrounding projects for large-scale image archives in the nineteenth and early twentieth centuries. Collections of photographs, stereographs and films were thought of as trustworthy and unbiased documents, that allowed for the production of new forms of knowledge. The expectations as to the impact of such new media that circulated at the time are not unlike those formulated today with respect to 'big data'. It is only by scrutinizing those discourses, and specifically the role attributed to media technologies, that we can understand the processes that govern the production of each medium's bias.

KEYWORDS: media archaeology; big data; image archives; data collections; trust

Introduction

The popular narrative of big data contains a claim to efficiency and accuracy.¹ The alleged scale of data available for analysis is supposed to compensate for any bias, thus producing accurate, objective and truthful results. Arguably, there is a prehistory to such ideas – that is, to suggestions that the availability of large numbers of records make possible a better, perhaps even more objective, understanding of the world. In this contribution, we would like to explore this prehistory. By reviewing promises, voiced by commentators, inventors and users of technology from early photography and film to today's databases and data dashboards, this article deconstructs the narrative and the promises shaping popular understandings of media technologies.

Utopian views of future uses and possibilities offered by new media discursively shape public understandings of technologies and their benefits for societies. Critical analysis of promotional and popular, but sometimes also scholarly, discourses, whether they address the 'technological imagination' in general terms, or more specifically the 'technological imaginary' of emerging cinema or the internet, or the 'myths' surrounding mobile communication technologies, allows us to deconstruct the utopian perspectives they paint.² A recent study by Taina Bucher applies a similar approach in an attempt to explore the 'algorithmic imaginary' by reviewing users' comments on the Facebook algorithm.³ Also, Erkki Huhtamo's media

archaeological explorations of discursive or iconographic ‘topoi’ involve such analyses, all of which aim at assessing the role that discursive constructions play in the way in which media and their characteristics are perceived and understood.⁴

One common denominator of most of the above-mentioned technologies – photography, film, computers and internet applications – is that the processing of records is or becomes increasingly automatised. On the one hand, this yields promises and perhaps even utopian expectations as to the way in which they can produce reliable, trustworthy, accurate representations of the real that will make possible a number of seemingly revolutionary new practices. On the other hand, the emergence of such technologies is accompanied both by corresponding negative reactions based on dystopian fears, pointing towards the threats such technologies represent, and (perhaps more importantly) by a scepticism that quickly leads to an interrogation of the utopian claims. The debates that ensue foreground certain functions of a medium and try to negotiate the conditions under which a media *dispositif* of trust can be established.⁵ We are thinking here of the kind of everyday functioning of a medium that requires a certain amount of trust in the configuration of technological, institutional, and textual practices that we take to be ‘the medium’, and that allows us to use it without having to continually question whether or not we fall prey to its flaws.

In what follows, we look at a number of examples of indiscriminate ‘collecting’ (of photographic, cinematographic, and other records) that allow us to draw analogies between historical and contemporary forms of ‘big data’. In doing so, we focus on the discourses that informed such practices, allowing us to identify some of the central issues in the accompanying debates, as alluded to above. Arguably, it is the negotiation of those debates that, in the end, lead to what we might call ‘media literacy’ in its most basic form: an understanding of the medium not as a black box, but as a process of translation in the course of which ‘input’ is *processed* in order to *produce* an ‘output’. Exploring said debates requires that we begin with a closer look at early manifestations of the aforementioned belief in the fundamental objectivity of techno-images.

‘An Enormous Collection of Forms’: Stereoscopy, Truth, and the Image Archive

Photographic images are part of what media theorist Vilém Flusser has called ‘techno-images’. According to Flusser, science and technology since the nineteenth century have increasingly delegated the process of picture-making to machines, because of the superior quality of reproduction that can be attained through them.⁶ Techno-images, indeed, seem to be seductively convincing in their promise of rendering an accurate depiction of the world. What is often neglected here is the fact that there is an apparatus between the world and the user. The reason seems to be that the inner working mechanisms of the apparatus remain opaque, and that therefore, the machine operates as a black box. Perhaps, it is more accurate to say then that the existence of the apparatus is in fact not so much neglected, but rather generates additional trust in the objectivity and accuracy of the images it produces.

This trust fuelled scientist François Arago’s enthusiasm when he gave his report on the daguerreotype in the French parliament in 1839, in which he stressed, among other things, the

fidelity of photographic records. By way of example, Arago claimed that the Egyptian hieroglyphs could have been reproduced easily and without any errors by means of daguerreotypes, whereas the handmade copies compiled by draughtsmen during Napoleon's expedition have many flaws.⁷ In this comparison between automated and human labour, Arago argued along similar lines as Charles Babbage, who 'rhapsodized about the advantages of mechanical labor for tasks that required endless repetition, great force, or exquisite delicacy.'⁸ Here, Arago was less concerned with photography's trustworthiness as a witness than with photography's reliability as a mechanical copying device.

According to media theorist Pasi Väliäho, 'self-recording' devices such as photography and cinematography apparently produced an epistemological tension between phenomenological and automatised perception:

It is crucial to note, regarding this epistemological problem, that self-recording and simulation machines embody a degree zero of perception, a kind of 'zeroness' of perception, which takes place *prior* to the emergence of the human observer. Consequently, these machines can be understood as a type of non-human observer that generates the very possibility of visual knowledge (...). As producers of non-sensed sensibilia, self-recording machines operate as partial observers that embody the affections and perceptions without which scientific functions and propositions would remain unintelligible. They create the sensibilia that scientific functions suppose.⁹

Lorraine Daston and Peter Galison, focusing on the role of photography as a scientific tool in the second half of the nineteenth century, observe that scientists were well aware of the fact that photographs were anything but a direct and unfiltered product of 'the pencil of nature'.¹⁰ But at the same time, the general trust in the fundamental objectivity of the photographic image was not questioned in any radical way. So, despite widely available knowledge about the manipulability of photographs, the truth claim of the photographic medium as such could be upheld, even though the authenticity of individual photographs might be in dispute. This is what the French film critic André Bazin, in his 1945 essay on "The Ontology of the Photographic Image", referred to as the 'essentially objective character of photography' (in a phrase alluding also to the fact that in French, the lens of the camera is called *objectif*; or, as in German: *Objektiv*).¹¹

The presumed 'essentially objective character' of the medium, then, could at the same time be disputed and asserted as early as the second half of the nineteenth century. The 1864 book *Soundings from the Atlantic* by Oliver Wendell Holmes, a professor of anatomy and physiology at Harvard University and also a prolific writer, comprised a collection of essays, three of which were dedicated to photography and stereoscopy (and originally published in the journal *The Atlantic Monthly* in the years between 1859 and 1863). Holmes was fascinated by the new medium of photography, and even more so by the possibility of creating stereoscopic images. He even constructed a handheld viewer for the latter, which he decided not to patent so that it could be widely used. Being well aware of the complicated process of producing a photograph and the various manipulations that occurred throughout the process, Holmes was by no means

naive regarding the medium's trustworthiness. However, as he did with the stereograph, he credited it with greater reliability than a simple photograph, precisely because it was more complex:

Another point in which the stereograph differs from every other delineation is in the character of its evidence. A simple photographic picture may be tampered with. (...) But try to mend a stereograph and you will soon find the difference. Your marks and patches float above the picture and never identify themselves with it. The impossibility of the stereograph's perjuring itself is a curious illustration of the law of evidence.¹²

For Holmes, its evidentiary character made the stereograph an ideal tool for documentation – even for replacing the actual, original object. This is evident from a declaration reminiscent of the proclamations of certain digital enthusiasts some 140 years later:

Form is henceforth divorced from matter. In fact, matter as a visible object is of no great use any longer, except as the mould on which form is shaped. Give us a few negatives of a thing worth seeing, taken from different points of view, and that is all we want of it. Pull it down or burn it up, if you please.¹³

According to Holmes, the possibility of producing faithful stereographic representations in an apparent three-dimensionality of all that is visible would lead to a massive 'stereographisation'. The future imagined here was not unlike that projected by large-scale digitisation initiatives today: commercial ones such as those of Google Books or Google Maps, or funded heritage projects such as the Dutch film digitisation scheme 'Images for the Future' ('Beelden voor de Toekomst') or the international collaboration that led to Europeana Collections, and so on. At the time, Holmes had a certain expectation:

The consequence of this will soon be such an enormous collection of forms that they will have to be classified and arranged in vast libraries, as books are now. The time will come when a man who wishes to see any object, natural or artificial, will go to the Imperial, National, or City Stereographic Library, and call for its skin or form, as he would for a book in any common library.¹⁴

For this reason, he advocated 'the creation of a comprehensive and systematic stereographic library, where all men can find the specific forms they particularly desire to see as artists, or as scholars, or as mechanics, or in any other capacity.'¹⁵

However, Holmes also paid attention to the given that the iconic substrates of objects, their 'forms', need to be formatted in some way to make working with them possible. Therefore, the library should not collect just any stereographic reproduction of objects, but preferably those produced according to a certain protocol. He suggested that to 'render comparison of similar objects, or any that we may wish to see side by side, easy, they should be taken, so far as possible, at the same distance, and viewed through stereoscopic lenses of the same pattern. In this way

the eye is enabled to form the most rapid and exact conclusions.’¹⁶ Such formatting would concern both the recording and the viewing apparatus, in order to avoid distortions and misinterpretations by the user, while at the same time enhancing the objectivity, and thus trustworthiness, of the stereographic records.

Early Cinematographic Collections and the Quest for Exhaustive Knowledge

In 1898, about three and a half decades after Holmes’ plans for a stereographic library, the Polish photographer and cinematographer Boleslas Matuszewski published a little brochure entitled *Une nouvelle source de l’Histoire (Création d’un dépôt de cinématographie historique)*, followed that same year by a book entitled *La Photographie animée, ce qu’elle est, ce qu’elle doit être*. In both publications, Matuszewski promoted animated photography, i.e. moving pictures, as an important source for the production of historical documents and as a scientific tool. In *Une nouvelle source de l’Histoire*, he stressed in particular the advantage of cinematography vis-à-vis photography, claiming that the sheer amount of individual photographic records on a filmstrip protects animated pictures against attempts at manipulation:

Perhaps the cinematograph does not give history in its entirety, but at least what it does deliver is incontestable and of an absolute truth. Ordinary photography admits of retouching, to the point of transformation. But try to retouch, in an identical way for each figure, these thousand or twelve hundred, almost microscopic negatives...! One could say that animated photography has a character of authenticity, accuracy and precision that belongs to it alone. It is the ocular evidence that is truthful and infallible *par excellence*.¹⁷

The argument here echoes the one made by Holmes concerning the trustworthiness of stereographs: once again, the medium’s higher complexity is seen as a safeguard against manipulation.

However, as well as proposing a purely quantitative argument to advocate on behalf of cinematography’s trustworthiness, Matuszewski also attributed this quality to the technological specificity of animated photography (which, more than sixty years later, would be famously restated in the following terms by a character in Jean-Luc Godard’s *Le Petit Soldat* of 1963: ‘Photography is truth, the cinema is truth 24 times per second.’). This way, he highlighted both the technological advantage of cinematography over photography and the fact that what is recorded by animated pictures cannot be doubted: ‘It can verify oral tradition, and if human witnesses contradict each other on some matter, it can bring them into accord, shutting the mouth of whoever would dispute it.’¹⁸ In a footnote to the text, a reference is made to an alleged diplomatic incident that occurred during the visit of the French President Félix Faure to St Petersburg. Somewhat triumphantly, Matuszewski declared that the footage shown during the projection of one of his own films, recorded on this occasion, ‘was found indisputably to refute the false assertions from abroad.’¹⁹

There is one aspect to cinematographic records that Matuszewski considered to decrease their value as historical documents: the fact that many of them were produced for entertainment purposes (which, in his opinion, included the entire production of the Lumière brothers²⁰). Therefore, he argued, all ‘animated photographs’, before being admitted to the repository, should be evaluated. He demanded that ‘A competent committee [would] accept or discard the proposed documents after having appraised their historic value.’²¹ In his second publication, *La Photographie animée*, Matuszewski again stressed the task of such a committee, which would first entail eliminating ‘everything that is pure amusement and does not represent [a] character of utility.’²² So, experts were needed to assess the specific quality of the record as document. It is not clear, however, as to which competences would facilitate such a committee. Matuszewski did not offer any details here, but one might guess that the experts would have to be able to both judge the adequacy of the representations and have an understanding of the modes of production of the images – as they would have to assess their scientific and historical value and eliminate pictures made to simply entertain the general audience.

This attitude is quite different from the one articulated about a decade later by the film producer Charles Urban. Like Matuszewski, Urban advocated for the creation of an archive, but unlike him, he proposed to collect records indiscriminately:

Animated pictures of almost daily happenings, which possess no more than a passing interest now, will rank as matters of national importance to future students, and it behoves our public authorities, and the heads of museums and universities, to see that the institutions under their control become possessed of these important moving records of present events.²³

Urban believed that it was up to future experts to discover, or identify, the value of the pictures preserved in the archive; or in other words (one might infer), to understand in what way they can ‘rank as matters of national importance’.

In spite of those differences, Matuszewski and Urban did reason along similar lines at least in some respects, as neither of them included staged scenes in their respective archival projects. Matuszewski, obviously, presumed that views taken to entertain the general audience were in conflict with the scientific nature of the documents which the repository was supposed to preserve. Urban, in contrast, argued that an animated (documentary) picture, regardless of its original purpose, could turn into a valuable document when viewed in an appropriate context or from a relevant perspective. Both authors also stressed the importance of the indexical qualities of the cinematographic image, and furthermore considered it a first guarantee for its status as a record. For Matuszewski, however, this was not a sufficient argument for inclusion (as he did not consider animated photographs made for entertainment purposes to be potentially trustworthy historical documents). Therefore, a competent – or, in somewhat anachronistic terms: media literate – committee was needed to select the views worthy of being kept in the archive. Urban instead left it to the competent or media literate *future* viewer to read an animated picture in such a way that its documentary value could be revealed.

It follows from the above that in spite of their diverging approaches, both men – much like Holmes before them – required protocols that governed the ways in which images were to be chosen for preservation. The technology itself, however, remained ‘black-boxed’ – for Urban as well as Matuszewski. The cinematograph, they reasoned, allowed the capturing of events truthfully, and therefore it was considered a reliable tool to gather visual records for an archive that complemented, perhaps even surpassed, the existing archives of written documents.

Matuszewski’s and Urban’s film collections were conceived from the start as being for future generations’ use. Containing ‘trustworthy’ documents that could show the world and events of the past, they constituted essential primary resources for historians. In Matuszewski’s projection, the panel of experts that decided on the in- or exclusion of specific images was instrumental in shaping future knowledge about the past; in contrast, Urban’s proposal tended towards a potentially infinite collection of data, similar to the stereographic collections advocated by Holmes.

In this respect, both initiatives are distinctive from what is likely the best-known project in image archiving of the early twentieth century: Albert Kahn’s ‘Archives de la Planète’. The photographs, stereographs and films to be collected in Kahn’s archive were to be produced *specifically* for this purpose, and the camera operators received instructions to capture ‘the familiar type’ and the ‘everyday’.²⁴ In other words, the first image selection, here, occurred at the very moment of their production. So, Kahn’s collection was not meant to be all-encompassing in the manner of Holmes’, Matuszewski’s or Urban’s: the images in the Archives de la Planète were to allow a comparative view of how people lived all over the planet, in order to create (specific) new knowledge. As Kahn expert Paula Amad argues, ‘The idea behind this comparative thrust was that new truths (and new order) might emerge from old familiars and the old disorder.’²⁵

Whether or not completeness was at stake, such large-scale archives inevitably demanded a system that allowed for the retrieval of the information gathered in it. This task was taken up as early as 1906 by the Union internationale de photographie, which proposed using the Universal Decimal Classification conceived by the Belgian lawyers Paul Otlet and Henri La Fontaine.²⁶ Otlet and La Fontaine pursued such classification efforts on an even larger scale in 1910, with their Mundaneum – an attempt to register the world’s knowledge, based around collecting the catalogues of notable libraries. The Mundaneum, indeed, was a collection of information about information, or ‘meta-data’.²⁷ Benefiting from such efforts, the large-scale collecting of photographic, stereographic and cinematographic records deemed necessary in order to amass documents that together gave access to trustworthy information almost automatically led to initiatives in data management. Similar tendencies towards amassing records or data in order to improve accuracy can be traced in other fields, both at the time and more recently – specifically, as part of ventures in what has been referred to in both periods as ‘social physics’.²⁸

To conclude this section, we would like to reiterate that the media dispositif of trust constituted by photography and the cinematograph relied primarily on the indexical and automatised form of reproduction that was made possible by said technologies. The relation between records and indexicality, however, was not given but discursively *produced*, and it was informed by the expectations that commentators attached to such emerging media. As it became common

knowledge that images could also be manipulated, trust was renewed by reference to more 'complex' media (the stereograph, the cinematograph), whose very complexity was taken to render manipulation impossible. The archives that stored the records produced with such technologies were to become a source of future knowledge by virtue of the sheer quantity of the documents they contained – a quantity, as we pointed out, that had to be made manageable, and accessible, with the help of other ('meta-')data.



Figure 1. Collecting professions: Orange vendors in Paris and Cairo (Archives de la Planète, Albert Kahn). Top: Auguste Léon, Marchand d'oranges, Cairo, 5 February 1914 (inventory number A 3 621 X). Bottom: Auguste Léon, Les marchandes d'oranges faubourg St Antoine, Paris, 14 May 1918 (inventory number A 14 050 S).

Computed Truth and Software-generated Techno-images

Arguably, Vilém Flusser's notion of techno-images can be applied to any visual representation of the world produced by a technological device. This also includes an automatically-generated data visualisation (such as a Google Ngram view), for instance, or a computer simulation. All of these products carry the very same qualities that Flusser found problematic in techno-images: they are translations of the world into code, while the codes of a technological device cannot be understood by the image's users – or only insufficiently so – unless, of course, we learn how to master the code. If we do not, Flusser warns, 'we are condemned to endure a meaningless existence in a techno-imaginary codified world that has become meaningless.'²⁹

Contemporary knowledge economies both depend and thrive on software applications. Through their graphical user interfaces, the world appears to be manageable, calculable and predictable. From data dashboards to social network visualisations to weather simulations, these new techno-images are an attempt to translate the complexity of the world into a comprehensible form. They appear (and are used) as objective and scientifically calculated – or: objective, *because* scientifically calculated – representations of various aspects of our life-world.³⁰

Compared to the images contained in the collections hailed by Holmes, Matuszewski, Urban or Kahn in the nineteenth and early twentieth centuries, those new techno-images seem to present a number of important differences. Firstly, they are not translations of a 'real' in front of a camera, translated into a techno-image by means of a photochemical process, but rather generated out of collections of data previously compiled, by means of algorithmic codes. Secondly, it is not the techno-images that are collected to constitute an archive; rather, they are ways to access the archive – here, a data collection – in a visual form. Thirdly, while photographs and films are known these days to be manipulable in ways that their advocates of a century ago did not or would not imagine, such data-based images do indeed carry a renewed hope of reliability, objectivity and truthfulness.

In reality, of course, things are not that simple – as is evident to those engaged in the emerging field of critical data studies.³¹ Scholars here develop methods for critical inquiry into data practices and analysis tools. Using those methods, they seek to reveal the working logics hidden under opaque interfaces, which remain incomprehensible to most users – for instance, bias in Google search results, predictive policing applications, automated public services, recidivism risk assessment, or the Value Added Teacher Model for evaluating teachers.³² Key analysis tasks are increasingly delegated to software, and as a result, decisions concerning the livelihoods of persons, demographic groups or entire populations unfold beyond public checks and balances. Therefore, it is important to confront the widespread technological imaginary of 'big data' producing accurate and unbiased results – as a way of shaping a much-needed form of contemporary media literacy.³³

Software, much like photographic or cinematographic cameras, cannot be considered neutral, but rather as an agent active in shaping knowledge and producing facts, transforming the very things it is supposed to analyse and represent or process 'neutrally' (and in this sense, a non-human actor in the Latourian sense³⁴). This raises questions about the technical design of

software applications, the collection and selection process of data, the quality of the data, the definition of indicators and the models implemented, the visualisations they produce, and, finally, how these processes inform decision-making. In terms of Flusser’s understanding of techno-images, the data-analysis processes, computer simulations, data visualisations and so on are codes to represent and understand the world. Flusser explicitly speaks of images as something to be understood as ‘tools’ or ‘maps’ for understanding or navigating the world.³⁵ Much like the records collected in cinematographic or photographic archives, they promise to provide, or even produce, new insights, to deliver accurate results, and to contribute to decision-making.³⁶ But in doing so, they hide what happens underneath.

In the context of data visualisation specifically, a plethora of new images have emerged – from computer-generated calculations and spreadsheets via PowerPoint presentations to computer simulations. What exactly they represent, presumably, is not always clear. A case in point here is the now-infamous PowerPoint slide used by the U.S. armed forces that was intended to help map and understand the Afghan theatre of war (see Figure 2 below). As an involuntary counter-example to the functionality of images referred to by Flusser, it effectively failed to work as a ‘map’. Allegedly, the commander at the time, General McChrystal, reacted to it with the words, ‘When we understand that slide, we’ll have won the war.’³⁷

Critique of the persuasive objectivity of the results of computer-aided analysis processes, data visualisations and computer simulations has been amply voiced by, among others, such commentators as Evelyn Fox-Keller, Günter Küppers and Johannes Lenhard, as well as Sherry Turkle.³⁸ All of them acknowledge the value of those tools and processes for science, engineering, design and business, but they question the users’ ability to recognise

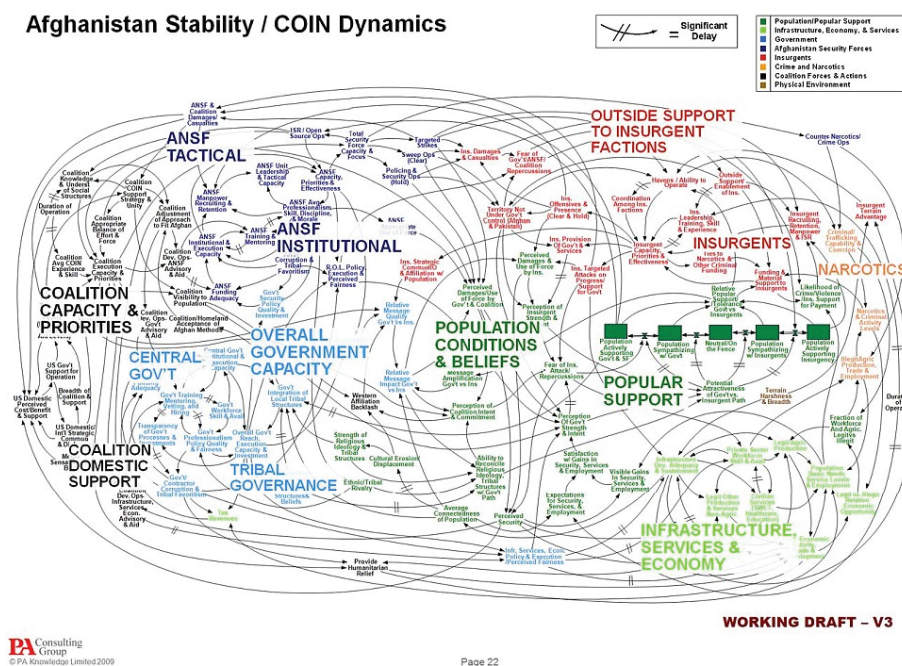


Figure 2. PowerPoint slide shown to US commanders meant to portray the complexity of American strategy in Afghanistan (2009). Source: Daily Mail online (28 April 2010), <http://www.dailymail.co.uk/news/article-1269463/Afghanistan-PowerPoint-slide-Generals-left-baffled-PowerPoint-slide.html>.

the technology's limitations, and to reckon with its opaqueness. Orrin H. Pilkey and Linda Pilkey-Jarvis, assessing the mathematical models used in predictive data analysis, have described their limitations in the field of environmental sciences as 'ordering complexity'.³⁹

The uncritical reading of analysis processes can have drastic consequences. For six days in April of 2010, air traffic was grounded in most European countries due to the potential distribution of the ash cloud caused by the Icelandic volcano Eyjafjallajökull. Regulators were very much guided by a simulation sketching the possible distribution of ash particles.

The image, a simulation compiled most likely from weather data and models informed by decades of meteorological records, did not depict the actual distribution of ash, but a possible one, and also did not take into account the density of particles at various altitudes or in different areas. The decision to ground air traffic on the basis of a computer simulation was subsequently criticised by computer scientist David Gelernter. He argued that blindly trusting knowledge technologies without questioning their limitations and understanding their logic can have dire consequences. In his assessment, this will eventually entail that 'Firstly we'll be covered in an ash cloud of anti-knowledge and secondly a moral and intellectual passivity will emerge that won't doubt or argue against the images.'⁴⁰

In their analysis of the data artwork *The Architecture of Radio*, an app that visualises digital radio signals and the presence of devices in the user's environment, Eef Masson and Karin van Es have shown that 'understandings of data as drawn directly from reality, and by implication as proof for truth claims about the world, wield their influence *even* in visualization explicitly positioned as speculative rather than evidentiary.'⁴¹ The persuasive power of data, the authors assert, continues to be effective even in the context of an art project, because, precisely, data continue 'to operate within an indexical paradigm.'⁴²

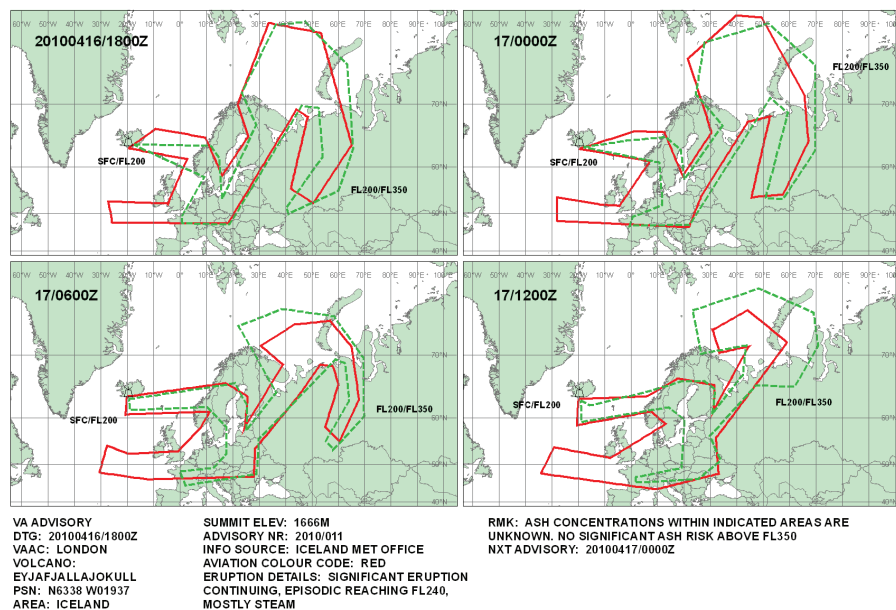


Figure 3. Simulation that predicted the volcanic ash cloud from Eyjafjallajökull, 2010. London Volcanic Ash Advisory Centre.

In sum, the contemporary use of data analysis, computer simulation and data visualisation techniques is informing our understanding of the world at a very large scale. In doing so, such techniques operate much like the kinds of ‘tools’ and ‘maps’ that Flusser talks about. In a sense, they are even more profoundly ‘techno-’images than the photographs, stereographs and films that we previously discussed, since they are generated from data rather than being recordings of the real, caught in front of a camera. Moreover, while archives of such records were meant to afford an accurate documentation of historic events, social conditions or accumulated knowledge, contemporary techno-images become active agents of interpretation processes and decision-making and are often credited with a predictive power. The media dispositif of trust that is constituted by these new forms of techno-images brings together indexicality (as the data from which visualisations are generated are supposed to be an unbiased, objective and accurate rendering of the real) with the almost immeasurable quantity of data, which is said to be capable of neutralising variance, and, in its effect, exclude error.

Conclusion

Even if they are separated by more than a century, emerging image practices such as photography, stereography and film, and computer visualisations, present similarities in terms of how they are taken to have an impact on the way in which knowledge is produced, stored and made accessible.⁴³ New media technologies always carry with them promises and threats, and media archaeological studies have shown that those more often than not materialise in similar kinds of metaphors or discourses.⁴⁴ As we have tried to show, the conceptualisations of archives collecting new forms of techno-images in the nineteenth and early twentieth centuries articulated hopes that are echoed by contemporary discourses on the possibilities offered by ‘big data’.

In order to understand the logics behind new technologies and to assess their effects, it is necessary to work with methods that are capable of opening up the black boxes in which their inherent biases are hidden – and that may in the process contribute to forms of media literacy that are particularly relevant today. For the technologies shaping today’s datafied society, critical data studies is a strong and promising field. But the discursive constructions of these technologies that shape their cultural reception and understanding are often much less innovative – while their effects are just as significant. And this, indeed, is why we would argue for a *historically informed* field of critical data studies that is even better equipped to help tackle the challenges that twenty-first-century media studies have to face.

Notes

1. See, for instance, Chris Anderson, “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete,” *Wired Magazine* 16, no. 7 (2008), <https://www.wired.com/2008/06/pb-theory/>.
2. See, respectively, Teresa De Lauretis, Andreas Huyssen and Kathleen Woodward, ed., *The Technological Imagination: Theories and Fictions* (Madison, Wis.: Coda Press, 1980); Michael Punt, “Early Cinema and the Technological Imaginary,” (PhD thesis, Universiteit van Amsterdam, 2000), and Patrice Flichy, *The Internet Imaginaire* (Cambridge, Mass., London: The MIT Press, 2008); Imar O. de Vries, *Tantalously*

- Close: An Archaeology of Communication. Desires in Discourses of Mobile Wireless Media* (Amsterdam: Amsterdam University Press, 2012).
3. Tania Bucher, "The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms," *Information, Communication & Society* 20, no. 1 (2017), 30–44.
 4. See Erkki Huhtamo, "Dismantling the Fairy Engine: Media Archaeology as Topos Study," in *Media Archaeology: Approaches, Applications and Implications*, ed. Erkki Huhtamo and Jussi Parikka (Berkeley, Los Angeles, London: University of California Press, 2011), 27–47.
 5. For the concept of *dispositif*, see Frank Kessler, "Notes on Dispositif," Frank Kessler home page, <http://www.frankkessler.nl/wp-content/uploads/2010/05/Dispositif-Notes.pdf>.
 6. Vilém Flusser, *Medienkultur* (Frankfurt am Main: Fischer Verlag, 1997), 13.
 7. Dominique-François Arago, *Rapport sur le Daguerriotype* [1839] (La Rochelle: Rumeur des Ages, 1995), 38. The 'fidelity' addressed by Arago has since the 1970s been discussed mainly in terms of 'indexicality' in a Peircian re-reading of André Bazin's reflections on the 'ontology of the photographic image'. See also footnote 11, below.
 8. Lorraine Daston and Peter Galison, *Objectivity* (New York: Zone Books, 2010), 139.
 9. Pasi Väliäho, *Mapping the Moving Image. Gesture, Thought and Cinema circa 1900* (Amsterdam: Amsterdam University Press), 42.
 10. Daston and Galison, *Objectivity*, 125–138.
 11. André Bazin, "The Ontology of the Photographic Image," *Film Quarterly* 13, no. 4 (1960), 4–9, here 7. Tom Gunning, in "What's the Point of an Index? Or, Faking Photographs," *Nordicom Review*, no. 1–2 (2004), 39–49, discusses Bazin's ideas and the way they have been related to Charles Sanders Peirce's concept of the 'index'. Gunning further elaborates on his reflections in his subsequent essay "Moving Away from the Index: Cinema and the Impression of Reality," *Differences: A Journal of Feminist Cultural Studies* 18, no. 1 (2007), 29–52, where he also asserts that, despite many claims to the contrary, digital photography does not address the viewer in ways fundamentally different from analogue photography. David Rodowick, *The Virtual Life of Film* (Cambridge, Mass.; London: Harvard University Press, 2007), takes a view opposite to Gunning's on this matter. For a discussion of the debate concerning analogue and digital photography's indexicality, see also Frank Kessler, "What You Get Is What You See: Digital Images and the Claim on the Real," in *Digital Material: Tracing New Media in Everyday Life and Technology*, ed. Marianne van den Boomen et al. (Amsterdam: Amsterdam University Press, 2009), 187–197.
 12. Oliver Wendell Holmes, *Soundings from the Atlantic* (Boston: Ticknor and Fields, 1864), 174–175.
 13. Holmes, *Soundings from the Atlantic*, 161 (emphasis by Holmes). For a modern equivalent of such a radical stance, see, for instance, Anderson, "The End of Theory," who claims: 'This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.'
 14. Holmes, *Soundings from the Atlantic*, 162.
 15. *Ibid.*, 162–163.
 16. *Ibid.*, 163–164. The anonymous reviewer of our article suggested that Holmes, being a scientist himself, followed standard scientific practice to ensure the possibility of subsequent scientific use for stereoscopic images. As previously suggested, we completely agree with this, and would like to add that Holmes' reflections indicate that he was well aware of the translation process that occurs when taking a photograph or, for that matter, a stereograph, rather than seeing the camera as a 'black box'.
 17. Boleslas Matuszewski, *Une nouvelle source de l'Histoire: Création d'un dépôt de cinématographie historique* (Paris: Imprimerie Noizette et Cie, 1898), 9. We quote this text after its English translation "A New Source of History [1898]," *Film History* 7, no. 3 (1995), 322–324, here 323.
 18. Matuszewski, "A New Source," 323.
 19. *Ibid.*, 324.
 20. Quite probably, his true reason for excluding their production was not so much their entertainment status, but rather that they were competitors, and that he did not want them to benefit from his initiative.
 21. Matuszewski, "A New Source".
 22. Boleslas Matuszewski, *La Photographie animée, ce qu'elle est, ce qu'elle doit être* (Paris: Imprimerie Noizette et Cie, 1898), 57.

23. Charles Urban, *The Cinematograph in Science, Education and Matters of State* (London: Charles Urban Trading Co., 1907), 18–19.
24. Paula Amad, *Counter-Archive: Film, the Everyday, and Albert Kahn's Archives de la Planète* (New York: Columbia University Press, 2010), 78.
25. Amad, *Counter-Archive*.
26. See Luce Lebart, “L'internationale documentaire: Photographie, espéranto et documentation autour de 1900,” *Transbordeur photographie*, no. 1 (2017), 63–73.
27. See Frank Hartmann, “Von Karteikarten zum vernetzten Hypertext-System: Paul Otlet, Architekt des Weltwissens – Aus der Frühgeschichte der Informationsgesellschaft,” *Telepolis* (20 October 2008), <https://www.heise.de/tp/features/Von-Karteikarten-zum-vernetzten-Hypertext-System-3408411.html>.
28. A contemporary of Oliver Wendell Holmes, Belgian statistician Lambert Adolphe Jacques Quetelet saw large amounts of data as a means for generally limiting the bias of datasets (see Peter Atteslander, *Methoden der empirischen Sozialforschung* [Berlin, New York: De Gruyter, 1995], 19). Calling his research practice ‘social physics’, Quetelet formulated a research approach that involved the quantification of as many physical faculties of persons as possible, and that would provide insight into the moral state and development of society. See Adolphe Quetelet, *A Treatise on Man and the Development of His Faculties* (Edinburgh: William and Robert Chambers, 1842). In recent times, this denominator has been refashioned by Alex Pentland. Using large data repositories (e.g. 10 million transactions conducted by 1.6 million online traders), Pentland and his team try to understand how humans learn and share ideas, deeming such quantity essential to gaining insight into the working mechanisms of our social lives. As Pentland claims, ‘Social phenomena are really made up of billions of small transactions between individuals – people trading not only goods and money but also information, ideas, or just gossip. There are patterns in those individual transactions that drive phenomena such as financial crashes and Arab springs [sic].’ See Alex Pentland, *Social Physics* (New York: Penguin, 2014), 10.
29. Vilém Flusser, *Writings*, ed. Andreas Ströhl (Minneapolis, London: University of Minnesota Press, 2002), 41.
30. The more recent emphasis on ‘big data’ only amplifies those qualities. Analyses derived from so-called big data, at least in the popular understanding, are considered to be unbiased and accurate: ‘The new availability of huge amounts of data, along with the statistical tools to crunch these numbers, offers a whole new way of understanding the world. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all’ (Anderson, “The End of Theory”).
31. See the special issue of *Big Data & Society* 3, no. 3 (December 2016), edited by Andrew Illiades and Federica Russo.
32. See, respectively, Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, 2018); Serena Osterloo and Gerwin van Schie, “The Politics and Biases of the ‘Crime Anticipation System’ of the Dutch Police,” in *Proceedings of the Workshop on Bias in Information, Algorithms and Systems*, ed. Jo Bates et al. (Sheffield: CEUR Workshop Proceedings, 2018), 30–41; Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor* (New York: St. Martin's Press, 2018); Julia Angwin et al., “Machine Bias: There’s Software Used Across the Country to Predict Future Criminals. And It’s Biased Against Blacks,” *ProPublica* (23 May 2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>; Cathy O’Neil, “Don’t Grade Teachers with a Bad Algorithm,” *Bloomberg* (15 May 2017), <https://www.bloomberg.com/view/articles/2017-05-15/don-t-grade-teachers-with-a-bad-algorithm>
33. See Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, Mass., London: Harvard University Press, 2015).
34. See, for instance, Bruno Latour, “Technology is Society Made Durable,” in *A Sociology of Monsters: Essays on Power, Technology and Domination*, ed. John Law (London, New York: Routledge, 1991), 103–131.
35. Vilém Flusser, *Kommunikologie* (Frankfurt am Main: Fischer, 1998), 112.
36. For Flusser, the very first images produced by mankind were already functional, or instrumental. The cave paintings in Lascaux, according to him, were ‘good images if they lead to a successful hunt’. Flusser, *Kommunikologie*, 112.
37. Elisabeth Bumiler, “We Have Met the Enemy and He Is PowerPoint,” *The New York Times* (26 April 2010), <https://www.nytimes.com/2010/04/27/world/27powerpoint.html>.
38. Evelyn Fox-Keller, “Models, Simulation, and ‘Computer Experiments,’” in *The Philosophy of Scientific Experimentation*, ed. Hans Radder (Pittsburgh: University of Pittsburgh Press, 2003), 198–215; Günter Küppers and Johannes Lenhard, “The Controversial Status of Simulations,” *Proceedings of the 18th European Simulation Conference* (2004), 271–275; Sherry Turkle, *Simulations and Its Discontents* (Cambridge, MA: The MIT Press, 2009).

39. Orrin H. Pilkey and Linda Pilkey-Jarvis, *Useless Arithmetic: Why Environmental Scientists Can't Predict the Future* (New York: Columbia University Press, 2007), 32.
40. David Gelernter, "Gefahren der Softwaregläubigkeit: Die Aschewolke aus Antiwissen," *Frankfurter Allgemeine Zeitung* (26 April 2010) <http://www.faz.net/s/RubCEB3712D41B64C3094E31BDC1446D18E/Doc-E36DC93595654960A206495346999283~ATpl~Ecommon~Scontent.html>.
41. Eef Masson and Karin van Es, "Visualizing Connectivity: Data as Evidence in The Architecture of Radio," *First Monday* 22, no. 10 (2 October 2017), <http://firstmonday.org/ojs/index.php/fm/article/view/8039> (emphasis by Masson and van Es).
42. Masson and Van Es, "Visualizing Connectivity".
43. This list, of course, is anything but exhaustive: one could quite probably find similar discourses with respect to the electronic images of television, or video images on magnetic tape. These, however, are beyond the scope of our contribution.
44. See Huhtamo, "Dismantling the Fairy Engine".

Biographies

Frank Kessler is a professor of Media History at Utrecht University and currently the director of Utrecht University's Research Institute for Cultural Inquiry (ICON). His main research interests lie in the fields of early cinema and the history of film theory. He is a co-founder and co-editor of *KINtop: Jahrbuch zur Erforschung des frühen Films* and the KINtop-Schriften series. From 2003 to 2007 he was president of DOMITOR, an international association to promote research on early cinema. Together with Nanna Verhoeff, he edited *Networks of Entertainment: Early Film Distribution 1895-1915* (2007). He also published *Mise en scène* (2014).

Mirko Tobias Schäfer is an associate professor of New Media and Digital Culture at Utrecht University and principal investigator at Utrecht Data School. His research interests revolve around the socio-political impact of (media) technology. His publications cover user participation, datafication, and communication in social media. Mirko is co-editor and co-author of the volume *Digital Material: Tracing New Media in Everyday Life and Technology* (2009) and author of the book *Bastard Culture! How User Participation Transforms Cultural Production* (2011). Most recently, he edited, with Karin van Es, the volume *The Datafied Society: Studying Culture through Data* (2017).