

F A K T Y , P R O B L E M Y , P O L E M I K I

PIOTR TOCZYSKI

*Maria Grzegorzewska University in Warsaw, Poland***MEDIA ONLINE ARCHIVES REQUIRE MORE ATTENTION
FROM SOCIOLOGISTS**

Researchers and theorists in the humanities and social sciences have been interested in documenting certain facts, ideas and the atmosphere of the life of society as newspaper editors choose to select them. It is easy to check with the example of our current journal. Firstly, the pages of *Kultura i Społeczeństwo* / *Culture and Society* provide an instructive reminder on the meta-level that the press is important to sociologists for many reasons (Kotowska 2010). Undoubtedly, the archiving of media content is an important achievement of both technical and symbolic civilisation, and providing access to this content for those interested promotes the development of the social sciences. Secondly, a brief ad hoc review shows that writers for *Culture and Society* also, through access to the current and archived press, develop reflections they later share with the readers of their work. For example, some who publish in *Culture and Society* read one of the leading Polish dailies, *Gazeta Wyborcza*, for professional reasons, and later share their thoughts. I found out about this when I put a total of two phrases, “Gazeta Wyborcza” and “Kultura i Społeczeństwo” (the original Polish title for *Culture and Society*), into the Alphabet-owned search engine for academic texts, Google Scholar, yielding a set of results of ten per page. (Interested persons can reproduce the above procedure on their own, or use the link <https://scholar.google.pl/scholar?start=>

0&q="gazeta+wyborcza"+"kultura+i+społeczeństwo"&hl=pl.) Going through the first thirteen pages, I obtained more than one hundred and twenty PDF files with texts published in *Culture and Society* in which the authors used references to *Gazeta Wyborcza*. After glancing through some of them, it was easy to conclude that for sociologists, anthropologists, historians or researchers and theoreticians from the borderline of these disciplines who have published in *Culture and Society*, this very newspaper has for years been an important source of data, and perhaps even information, knowledge and wisdom, shaping their insights and opinions on the social and cultural reality.

In sociology, and more broadly in the social sciences, reviewing newspaper content and inferring societal life on that basis is therefore such an important norm in desk research procedure that it would be worth putting forward a few postulates. I think the social sciences should develop a position on the issue of press archives. To contextualise these postulates, we may also refer to the Code of Ethics, approved by the International Sociological Association Executive Committee in autumn 2001 (ISA 2001).

The Code says: "Sociologists who are being given access to records are expected to respect the privacy conditions under which the data were collected. They can, however, make use of data gathered in historical archives, both private and public, under the legal conditions laid down in the country concerned and usually accepted by the international scientific community, and subject to the rules of the archive". (ISA 2001)

More than twenty years after the introduction of the ISA code of ethics and its references to historical archives, the development of digital society also raises the question of media archives. How can sociologists respect the privacy conditions of media online archives? How can sociologists refer to the legal conditions laid down in the country concerned in the digital era? What is the standard accepted by the international scientific community? And finally, what actually are the rules of media online archives? Some tentative answers may be found in the current debates on free speech, data protection and privacy in the digital age, which extend at least into the transatlantic sphere, as shown in the examples of transnational proceedings. In order to build a global framework of reference for media online archives, there are several phenomena that require more attention. Some insights have already been collected (Kos 2019; Milosavljević, Poler & Čeferin 2020), but a good dozen or so questions still remain untouched. They are listed below and then

briefly discussed in order to inspire new research and directions of debate.

The list related to media online archives contains points such as: (1) Current underestimation of media online archives' potential; (2) Commitment to media online archiving: a right or the duty of publishers?; (3) The dilemma of being a library or medium; (4) Local, European, Western and global approaches to archiving; (5) Search engines and social media; (6) Publishers' undervaluation of archives; (7) Unequal online reputation management; (8) Self-regulatory possibilities; (9) Distribution and spreadability as a challenge; (10) Designing a standard of contextualising archival content. The above list of ten discussion points is not final, and could be further expanded in order to elaborate a more coherent Euro-American and global digital culture of responsible free expression.

Not much work has been done so far towards responding to the above issues. One such attempt was the report "Media online archives" by Dorota Głowacka, Joanna Smętek and Zuzanna Warso (Głowacka et al. 2015), focusing on three Central European digital archiving regimes. It covers Czech, Hungarian and Polish archives of the printed press, online editions, and horizontal portals. All three were first of all desk-researched from the perspective of certain publishers' and editors' practice, and then analysed from the viewpoint of emerging legal standards. Different domestic laws as well as international human rights systems — the latter common for the three above-mentioned countries and dozens of others in Europe — were taken into consideration. Furthermore, current challenges with the Polish digital media archives were additionally tackled by the survey (conducted in the second half of 2015, among some 23 publishers). Such sampling could be contested in terms of completeness, but it is sufficient for achieving the general aim of just exploring new socio-legal phenomena and forming some recommendations. The authors offer classifications of current approaches chosen by the courts, and transform them into a checklist enabling analysis of future cases.

A byproduct of the above-mentioned pioneering research is an insight into the variance of society, law and media cultures as marked by different attitudes towards archival content in the countries analysed. Apart from its modest scope, probably resulting from a modest budget, there is nothing to dislike about the report. It is an intellectually rewarding work that provides room for following up with further questions, suggestions and discussion. The report's conclusions and recommendations are

certainly applicable beyond the three countries from which declarative data was drawn and some state-of-the-art analysis has been carried out. Surprisingly or not, even such a modest sample of countries and publishers turns out to be sufficient for discussing the internet policy implications of archiving challenges faced by digital age media. Some of the remarks from the current paper should be read as an encouragement to read the report with a curious mind, in search of inspiring insights in at least ten areas. The report contributes to the current debate by raising and inspiring questions about archiving as such, the extent of media and non-media actors' responsibility for keeping content online, and the procedures that should be applied to archive management.

A discussion of the points mentioned above follows, leading to even more questions and a list of facilitating factors, followed by some concluding remarks.

CHALLENGE 1 AND FACILITATING FACTORS:
CURRENTLY THE MEDIA ONLINE ARCHIVES' POTENTIAL
IS UNDERESTIMATED

In the future, media content archiving will probably be seen as one of our civilisation's greatest achievements. It enables the accumulation of knowledge by those interested, through accessing data, information, and artistic and intellectual resources. Archives of the media are a log of civilisation, and given the ongoing development in new methods of interacting with them (such as text data mining), they could probably become one of the important resources for public debate and for pursuing private intellectual interests. The internet, being a high civilisational achievement of its own, is potentially a booster for the media archives, and its limit for this has not yet been tested. Interestingly, the value of archiving is currently still underestimated. This is due to the shortage of convincing business models for archiving, and the lack of motivation and competencies for using archives among the majority of users. Thus it can be expected that sooner or later, when people realise the value of archives, problems with media online archives will intensify.

What can a sociologist do? Studying the future may be relevant, and the sociological imagination may be useful for building a possible framework for media archives' future shape. As archives are a tool for public intellectuals and scholars of the history of ideas, these sociologists may be especially interested in seeing to the relevant infrastructure for their future work.

CHALLENGE 2 AND FACILITATING FACTORS:
 COMMITMENT TO MEDIA ONLINE ARCHIVING
 MAY BE A RIGHT OR THE DUTY OF PUBLISHERS

If archiving is a commitment of the media and their duty, then the issues of pricing access to the archives, licensing archival content, usability of the archives and the intellectual privacy of archives searches should be rethought. One such question is whether archive searches should be logged, and who should be allowed to analyse those logs? In other words: to what extent are media content archives a common good, and to what extent are they private property? Currently the archival content of entire media outlets is disappearing because of newspaper publishers' lack of diligence. The closure of an online medium, e.g. because of bankruptcy, results in the erasure of its archives. Should it be the responsibility of national libraries or national archives to archive this content (as should probably be the case with the archiving of intellectuals' and public figures' social media activity)? The relationship between public and private actors of digital archiving and the media should be rethought in order to answer this question.

From a single researcher's perspective, access and intellectual privacy are particularly important. However, other issues relate to broader social systems, issues of sustainability and even social order.

CHALLENGE 3 AND FACILITATING FACTORS:
 THE IDENTITY DILEMMA OF BEING A LIBRARY OR A MEDIUM
 IS CRUCIAL FOR SOCIAL PRACTICES OF ARCHIVING

What is the parallel of libraries in the digital media world? As we know, hardly anyone could sue a library for keeping newspapers' archives, even if featured in an unwanted context. Is a newspaper archive more like a library, or more like a current edition of a newspaper or magazine? When does the life of an online article end, and what should be the principles of republishing articles whose timeline has ended? Should online content have by default an "expiry date"? Should the archives be searchable from the level of search engines or social media, or should they be kept unindexed, as separate databases available when conducting a more in-depth search? Even in today's not yet developed digital world, such a choice determines whether the so-called "internet publication rule" or "single publication rule" is applied to online archives. Whatever the choice and future standard, given the current online content status for historical research, the responsibility of the media (e.g. for keeping their

archives well maintained, accurate and accessible) should probably be audited by impartial bodies. A more advanced issue worth considering is whether the archives themselves should not be maintained separately from publishers' websites.

Again, from an average researcher's perspective, the above issues may be important mostly in relation to the usability of archives, but scholars of media sociology will also find some of the topics mentioned above relevant as research topics on their own.

CHALLENGE 4 AND FACILITATING FACTORS:
LOCAL, EUROPEAN, WESTERN AND OTHER GLOBAL APPROACHES
TO ARCHIVING IN TERMS OF FREE SPEECH AND PRIVACY

Online reputation (and its management) lies on two continua: firstly, on the continuum between free speech and privacy, and secondly on the continuum between justified free speech and unjustified (defamatory or hateful) free speech. An important context would be the knowledge of how these two aspects have been approached around the world. Contexts from Western Europe and from outside Europe would enable referring to re-occurrences of the traditional American tension between the First and the Fourth Amendment, in regard to archiving in digital space. The transatlantic perspective is the most appropriate framework for situating other European approaches (such as Central European), but Australasian situations could also be informative and inspiring for understanding the Western standard and comparing it with the global approaches emerging in other parts of the world.

Researchers may find it fascinating to get a deeper insight into the issue of media online archives as an expression of global social and political processes, such as Europeanisation, Americanisation, Westernisation or Easternisation.

CHALLENGE 5 AND FACILITATING FACTORS:
SEARCH ENGINES AND SOCIAL MEDIA ARE NOT HELPING
WITH THE RIGHT TO HAVE CONTENT ADJUSTED

Online archives of newspapers are currently being redefined by two internet developments: search engines, and social media. With new patterns of individual access to online content within online spaces such as Google and Facebook, archives can obtain "new life". So far it seems that googling the archives causes more problems and makes us rethink archiving more than social media. However, due to the popularity of

social networking, social news, sharing and peer-reproduction practices, social media should be expected to bring new challenges for media online archives. To what extent should the right to be forgotten be implemented by both the media and technology industries, and by each of those “frenemies” separately? It would be hard for traditional print media to execute the right to be forgotten, whereas information technology companies are expected to implement it fully. Given the right to be forgotten and (even better justified) the emerging right to have content amended, new roles will emerge for companies offering search engine optimisation and search engine marketing, as well as for companies offering social media marketing. When rethinking digital rights it would seem reasonable to include the above market actors.

Change in social awareness and in business is resulting in the emergence of new roles and new social practices. Public content available to researchers differs from the content that is hidden. The need for differentiating adjusted content from the originally published content results in the need for new research strategies. For instance, more research contacts with gatekeepers may be needed.

CHALLENGE 6 AND FACILITATING FACTORS: PUBLISHERS' UNDERVALUATION OF ARCHIVES IS THE NORM

Will there actually be any archives in the future, or will resources that we nowadays see as archives just be parts of what is always available online? With strategies of web and print already differing from web-only media, it is clear that business-led media will sooner or later need to reinvent their approach to their archival resources. In the meantime, their understanding of the relationship between web and print is unclear, resulting in a lack of strategy and incomprehension of the archives' mission, resulting in its business value being understated. Organisations are not internalising norms for archival content. In effect, when confronted with requests for content to be unpublished, media organisations seem defenceless, even in terms of their intellectual capacity to rethink the request properly and find proper arguments for or against it. There is a lack of policy, resulting from insufficient creativity and low self-consciousness in both digitized and born-digital media — in both business and editorial terms. The professionalisation of archives management is a skill for publishers, and they most often simply do not have it.

Critical sociologists may be interested in adding some best practice suggestions and seeing media boards improve their self-awareness and

decision-making processes, so that they perceive archives maintenance and related regulations as a socially valuable mission bearing a certain level of social responsibility.

CHALLENGE 7 AND FACILITATING FACTORS:
UNEQUAL ONLINE REPUTATION MANAGEMENT, BRINGING FORWARD
THE ISSUE OF DIGITAL AND SOCIAL INEQUALITIES

Inequality in access to online reputation management services is at least worth noting. Unequal access to “optimisation of information retention” resulting from how influential the stakeholder is constitutes a new aspect of internet inequalities. The market of online reputation management will grow; the question is whether it should be left to the free market, with its capital-driven inequalities. We do not know what the current level is of unreported “silent removal” of content from archives, carried out for business reasons such as commercial relationships between publishers and advertisers, as well as between publishers and their stakeholders (such as shareholders or shareholders’ dependencies). Or perhaps private relations play a role in access to such removal? Content removal may be a process worth exploring.

Clearly research could be putting more attention, both in terms of methodology and basic research, on understanding and explaining the social processes of content removal from media online archives.

CHALLENGE 8 AND FACILITATING FACTORS:
SELF-REGULATORY POSSIBILITIES SHOULD BE EXPLORED
IN A MORE INFORMED WAY

Another issue worth rethinking is who should be dealing with requests for unpublishing or amending, and at what pace. Should it be left to the publishers themselves? One option could be external bodies such as “joint industry committees”, or internal bodies involving ethicists, social researchers and philosophers. Such review boards could deal with requests as and when made, taking into consideration both the broader context and details of the case, and balancing values such as freedom of expression and privacy. Media and policy-makers might like to rethink what role media online archives should have: closed (a source for social researchers and business intelligence), or accessible (opinion-forming, be it either the website or syndicated spreadable content)? How to manage its potential threat to privacy and reputation, and in effect social bonds and social life? Discussing such complex issues should be approached with an

interdisciplinary range of specialists such as lawyers, philosophers, human rights scholars, media researchers and librarians. The problem is not bi-polar, and has many shades. What content should by definition disappear from public access, and what should remain? What are the consequences of permanent deletion of content, whereas it may still exist in individuals' private archives? Given the level of media's comprehension of complexity, their self-regulation may fail and other ways of regulation may be more effective.

Public intellectuals from sociology circles need to be present in joint industry committees, considering all the dilemmas related to the ever-pertinent question of who should be watching the watchmen.

CHALLENGE 9 AND FACILITATING FACTORS: DISTRIBUTION AND SPREADABILITY AS A CHALLENGE

The media value chain is dominated by new forms of digital distribution, and as such they may need new areas of (self-)regulation. Relationships between media, their websites and their current, past and distant content presence in both search engines and social media should be rethought. Could the media be expected to monitor the afterlife of their archival content in digital social networks and tools such as search engines (which they actually already do to some extent, using automated algorithmic tools for licensing reasons)? Archival content is often distributed — which is reflecting the predominant usage pattern of the internet through its gates such as Google and Facebook. Social media also constitute a social archive. Thus the technology of distributing appendices and adjustments should be developed for both search engines and social media. Any piece of content included within search results or social media feed would be updated with its amendment whenever such an amendment is published in the archive. Consequently, amendments should sometimes be included in the headlines and distributed with them. Search engines, social media and other technologies of algorithmic distribution should be equipped with such a capability. The question is whether media publishers can develop such media innovations themselves, or whether they need other actors (public, business or non-profit). Who is capable of developing new tools of this kind and mechanisms for the real-time syndicating of corrections and amendments? As media distribute their content to entire ecosystems, and since spreadability is the measure of success, the right to have content adjusted should follow. Such a solution should be cost-effective, but it should also maximise the likelihood of reaching the

same target group. A message admitting the mistakes should follow the preceding message.

The role of social researchers in designing policies and best practices could be to ensure that decisions are better informed. Their experience in using media content as research material could be enlightening.

CHALLENGE 10 AND FACILITATING FACTORS: DESIGNING A STANDARD OF CONTEXTUALIZING ARCHIVAL CONTENT

One of the additional research tasks to be tackled is comparing the quality of current archives from the point of view of clarity (for the reader) of amendments and adjustments. It can be assumed that archival content — especially parts that are disputable or contested — needs to be contextualised. It is not reasonable to rely on the readers and their motivation to contextualise information presented in media content, or on their information skills and time to make additional searches or gather information consciously. Given the broad accessibility of archival content, connecting it with the information adjusting it is crucial. But should this not be done using square brackets or footnotes rather than just amendments? How about pictograms? Given the current state of knowledge of readers' (mis)perceptions, designing effectively working solutions is a challenge.

Well established methods of researching user perception and ethical design exist, and they should be consulted and taken into consideration. The standard for the style and content of explanatory notes should also be (re)invented. Some examples of dilemmas are: should notes be mutually agreed upon when there are multiple stakeholders often described or depicted within the content? Or should there even be several notes accompanying disputable archival content? A symposium on such a topic would be beneficial to all stakeholders.

CONCLUDING REMARKS

The ISA (2001) code mentions sociologists, access to records, respecting the privacy conditions of data collection, using data gathered in archives, legal conditions, country laws, standards of the international scientific community, and good archiving practices. To this I would add the following: that whatever the future standard emerging from the current chaotic stage of internet development disputes and reflection is, it will probably take a number of aspects into account. These would include:

quick reaction, logging content changes, distributed reaction involving all ecosystems in which particular pieces of content turn up, and balancing between public interest advocacy and reputation management. The struggle to find solutions reconciling free speech and privacy protection — two important principles of Western civilisation — has to take into consideration that they themselves are both essentially contested concepts. And this struggle is a good illustration of the vagueness of digital rights and their current intersections with business and public policies.

The role of sociologists and scholars of related disciplines seems to be important from the viewpoint of the future of media archives. I have attempted to show several directions of thought, currently more or less informed, in which the role of social scientists as the proponents of best practices concerning media archives could be well received. However, awareness of the level of complexity in gathering data may be equally important.

REFERENCES

- Głowacka Dorota, Smętek Joanna, Warso Zuzanna, 2015, *Media Online Archives — A Source for Historical Research or a Threat To Privacy?*, Helsinki Foundation for Human Rights, Warsaw.
- Kos Maja Ovčak, 2019, *The Right to Be Forgotten and the Media*, “LeXonomica”, 11(2): 195–212.
- Milosavljević Marko, Poler Melita, Čeferin Rok, 2020, *In the Name of the Right to Be Forgotten: New Legal and Policy Issues and Practices Regarding Unpublishing Requests in Slovenian Online News Media*, “Digital Journalism”, 8(6): 780–796.
- Kotowska Agata, 2010, *Prasa a internet: konkurencja czy komplementarność?*, “Kultura i Społeczeństwo”, 54(2): 93–108.



ISA — International Sociological Association Executive Committee, 2001, Code of Ethics (<https://www.isa-sociology.org/en/about-isa/code-of-ethics>).

Abstract

Based on pioneering media online archives research, the International Sociological Association’s Code of Ethics, current debates on the right to be forgotten and some anecdotal evidence from *Kultura i Społeczeństwo*, the article’s author suggests several areas of interest for sociologists as public intellectuals and members of joint industry committees. The role of sociologists and scholars of related disciplines is important from the viewpoint of designing the future of media archives. In a world where unequal online reputation management brings forward

the issue of digital and social inequalities, the role of social researchers in designing policies and best practices, and standards of contextualising archival content, may be crucial. They could mediate between the interests of publishers, audiences and regulators in regard to social values and the progress of civilisation.

key words: online archives, social scientists as public intellectuals, digital inequalities