"MEDIA LIBRARIES AND ARCHIVES FOR THE 21ST CENTURY" NODE

Distributed archives, content from the past, content for the future

Alessandro Ludovico

Editor in chief of *Neural* magazine

Abstract

The web (paradoxically similarly to a human being) is particularly fond of its own age and much less so regarding what happened before its birth. Historical events, news, documents and cultural artefact at large since 1994 to the present are way easier to find than the ones antecedent to the middle nineties; the earlier the worse, and the later the better. It seems that the web reflects more and more a structure like a river flow, always heading to its mouth, more than remembering when and where its own water started to flow or even before that. So even if the late 1990s and 2000s are quite well covered, there's a tangible historical vacuum in the chaos of online archived information. And this is evident when you need to do serious research beyond the culture classics with a substantial lack of information freely available. Probably because online giants are aware of this cultural gap, they think that the final establishment of the web's reputation as a universally trusted medium goes through the migration in the online form of traditional media, and this means to let people access what they used to trust more: printed. Now its digitalization seems to add a distinctive quality to (trusted) printed media: being universally accessible. But digitalizing printed sources is a big task, a massive effort in trying to archive printed content and make them readable online. In the endless debate about the future of print, it can be felt as the final passage from the printed to the digital form: when even the most obscure printed titles will be available in a digital form, will there be anybody left who will still need print? And digitalizing means also to make a copy of the original printed material and possibly storing it online. But actually we should ask ourselves: can this be properly defined as "archiving"?

Keywords

archiving, printed media, digitalization, universal access

Archivos distribuidos, contenido del pasado, contenido para el futuro

Resumen

Por paradójico que resulte, la web está encantada con su propia época como si de un ser humano se tratara y le preocupa mucho menos lo que sucedió antes de su nacimiento. Los sucesos históricos, noticias, documentos y artefactos culturales en conjunto desde 1994 a la actualidad resultan mucho más fáciles de encontrar que los que anteceden a mediados de la década de 1990; cuanto más antiguos peor y cuanto más recientes mejor. Parece que la web refleja cada vez más la estructura de la corriente de un río, que siempre se dirige hacia su desembocadura, en vez de recordar cuándo y dónde empezó a fluir su propia agua, o antes incluso. Así que, aunque la etapa de finales de los noventa y principios de este siglo quede bastante bien cubierta, existe un vacío histórico tangible en el maremágnum de información archivada en línea. Esta circunstancia resulta evidente cuando hay que hacer investigación rigurosa más allá de los clásicos de la cultura y se detecta una carencia sustancial de información disponible gratuitamente. Probablemente porque los gigantes en línea son conscientes de esta brecha cultural, piensan que la consagración definitiva de la reputación de la web como medio de confianza universal pasa por la migración en línea de los medios tradicionales, para que la gente pueda acceder a aquello en lo que más confiaban en el pasado: a lo impreso. Ahora la digitalización parece añadir un valor distintivo a medios impresos de (confianza): el de hacerlos accesibles universalmente. Pero digitalizar fuentes impresas constituye una tarea ingente, un esfuerzo masivo de intentar archivar contenido impreso y hacerlo legible en línea. En el debate inacabable sobre el futuro de lo impreso, parece que la fase final es el paso del formato impreso al digital: cuando incluso los títulos impresos más desconocidos estén disponibles en formato digital, ¿quedará alguien que aún necesite lo impreso? Y digitalizar también implica hacer una copia del material impreso original y probablemente almacenarla en línea. Pero en realidad deberíamos preguntarnos: ¿puede este proceso definirse correctamente como «archivo»?

Keywords

archivo, medios impresos, digitalización, acceso universal

The web (paradoxically similarly to a human being) is particularly fond of its own age and much less so regarding what happened before its birth. Historical events, news, documents and cultural artefact at large since 1994 to the present are way easier to find than the ones antecedent to the middle nineties; the earlier the worse, and the later the better. It seems that the web reflects more and more a structure like a river flow, always heading to its mouth, more than remembering when and where its own water started to flow or even before that. So even if the late 1990s and 2000s are quite well covered, there's a tangible historical vacuum in the chaos of online archived information. And this is evident when you need to do serious research beyond the culture classics with a substantial lack of information freely available. Probably because online giants are aware of this cultural gap, they think that the final establishment of the web's reputation as a universally trusted medium goes through the migration in the online form of traditional media, and this means to let people access what they used to trust more: printed. That happened because printed material used to be checked and proofread before being printed, thus controlled and tested before being embodied in an unchangeable product. Now its digitalization seems to add a distinctive quality to (trusted) printed media: being universally accessible. But digitalizing printed sources is a big task, a massive effort in trying to archive printed content and make them readable online. In the endless debate about the future of print, it can be felt as the final passage from the printed to the digital form: when even the most obscure printed titles will be available in a digital form, will there be anybody left who will still need print? And digitalizing means also to make a copy of the original printed material and possibly storing it online. But actually we should ask ourselves: can this be properly defined as "archiving"?

I'll try to answer to this question later, but I'd like to quote what the American economist Jeremy Rifkin wrote in his book *The age of access*:

The physical container becomes secondary to the unique services contained in it [...]. Books and journals on library shelves are giving precedence to access to services via the internet. (Rifkin, 2000, p. 76-93, 100)

Truth is that we still have unparalleled cultural resources in old media format. Billions of books and magazines for example are still readable, even if, in some cases, they are a few centuries old. But their access is most of the time more complicated. In fact, unless you're close to a copy of these books and magazines, or, even better, in the same physical place (that obviously can instead be on the other side of the world) you can't read them, you can't flip them, you can't search through them. If they are listed by some bibliographical catalogue, you can learn where the closest copy is, although, again, it can be very far away. On the other end, we have global networks, which are constantly indexed, hosting an astounding amount of knowledge and culture that can be searched anytime through private search engines. But as said above, they are mainly fond of the last two decades, and only a small fraction of what's physically available has been digitized and indexed online. Google has raised the status of the most visited search engine, and so, implicitly, one of the most prominent digital archives of human knowledge produced digitally and publicly available. That's probably why Google's founders claimed that the culture preserved in print that is missing online is a problem for humankind. So Nikesh Arora, president of Google's Global Sales Operations and Business Development, confessed that Google founders' dream is "the creation of a universal library."[1] That ideally sounds like a perfect extension of their dominating position in indexing online content: broadening the indexing to the previously print-based content, even providing the needed digitalization and hosting resources. Actually, more than a dream it sounds like another huge business opportunity. In fact such a massive amount of content to be freely enjoyed and navigated is another attractive online space that Google could exploit for selling ads. In fact, contextual ads, related to the content, would be displayed while reading the book of choice selected from an immense online library, in the classic benevolent and fatherly Google style. What Google is trying to do is to digitize (and possible get rights of) huge chunks of cultural printed matter. With five million dollars as initial investment and twenty thousand publishing partners, including major libraries, after a few years of work they can claim seven millions books scanned, with one million already available in full preview in their service Google Books. Just to have an idea of how it works: most of the books are scanned using a special industrial camera at a very fast rate of 1,000 pages per hour. Furthermore, in 2008 Google completed the purchase of twenty million digitized historical newspaper pages from PaperofRecord.com, a Canadian company.[2] And in 2010 they signed another agreement with a European national institution: the Italian Minister of Culture was happy to give to Google, for free, the right to scan and host. In fact, in two years another million books in public domain hosted by the Italian national libraries in Rome and Florence. This approach has two sensitive problems. First: the access to this enormous body of culture is controlled and regulated by Google. It's not UNESCO, it's Google. It's not a non-profit international cultural institution; it's a private global business. Second: because of its specific aim, Google tends to acquire the most "universal" type of culture in order to be as popular as possible. Then what about the rest?

1. Online archives strategies and success

I've tried to analyze the characteristics of digital archives containing printed material in my upcoming book titled *Post-Digital Print, the mutation of publishing since 1894*. I've tried to analyze independent digital archives, and there are already quite a few excellent efforts online to be noted. Ubuweb,[3] for example, is a curated one about "all forms of the avant-garde and beyond", but especially fond of the sixties and seventies. Curators are filtering the precious digitalized material submitted by a community of enthusiasts, making them searchable and freely available in standard and enjoyable formats. It embeds the virtues of being focused on a topic but also being perceived as an outstanding reference by a community that contributes actively, "donating" precious digitalisations of rare materials (as an excellent library should aim to be). Aaaaarg.org[4] was another one on the same wavelength: they're digitizing (and asking people to help to digitize) hundreds of books and papers mainly related to academic research in art, media and politics. All the files are searchable as text, so establishing a unique vertical search engine on these topics, self-built and free of charge (compared to the very expensive academic commercial services). All the files are then freely available for download, although a registration is required. Their definition is kind of imaginative:

AAAARG was created with the intention of developing critical discourse outside of an institutional framework. But rather than thinking of it like a new building, imagine scaffolding that attaches onto existing buildings and creates new architectures between them.

1. See: <http://online.wsj.com/article/SB10001424052748703701004575113511364939130.html>.
2. See: <http://blog.searchenginewatch.com/081202-100005>.
3. See: <http://www.ubuweb.com>.
4. See: <http://aaaaarg.org>.

There are also online archives that are focused on magazines. One of the most celebrated magazine rebirths from the underground press was the *Radical Software* magazine.[5] All the eleven issues (printed in the 70s) were scanned and assembled in pdf format, and made freely available on their website, giving a significant contribution to media culture's researchers and scholars. The process was funded by The Langlois Foundation, focusing on this specific magazine, because it was considered as probably the first media art magazine ever published. Several other magazines, (the older the more considered) have been taken into consideration for being disembodied and then accessed online. To mention another seminal example, all *L=A=N=G=U=A=G=E*[6] issues, one of the most important experimental poetry magazines published in the 70s in USA, are available online in both pictures and pdf, but they are both not searchable as text. A different approach was taken by another magazine meant also as an art project: *PhotoStatic*.[7] It was a magazine that focused on xerography as the source of a peculiar visual language and art form, printed from 1983 to 1998. The editors slowly made pdf files of every issue, starting from the latest printed backwards; uploading them on their website and making them available for free download. The peculiarity of *PhotoStatic* is that where the original pdf files were not available they were "reconstructed", re-assembling the original layout, text and pictures as if they'd been made ready to be printed again. In this case the preserving effort was putting back in operation the original virtual "plates" that printed it, but in an abstract form that can eventually (and paradoxically) generate an even better quality printed product, that never existed in their own time. One of the essential questions is: are these processes really "archiving"? The long lasting printed copies are still in the libraries, and only some exceptional accident can delete them. The digital copies can be deleted (by accident or on purpose) in a second, although, their innate infinite duplication-induced spread easily their access. So having multiple copies of the scanned files, spread in potentially thousands or millions of different places and in different computers would help in the end their collective memory, and so maybe help in parallel and in a separate way their preservation. Furthermore, there's a hidden social and cultural aspect to be carefully considered in this respect. If you consider any "subculture" or literary, artistic or music movement, you can easily individuate a few key persons (journalists, historians, collectors, small institutions, obsessed fans) that have assembled over the years impressive collections that just lie in one or few rooms. These precious kinds of heritages are individually preserved but, in a word, they are invisible to the rest of the "scene" and to the world.

If their presence and at least essential references would be shared, then their context would become a public resource, a common that would have important consequences on both the online presence of the passionate collector and the general historical perception. If Wikipedia is the biggest effort in sharing knowledge from the most general and cultural perspective, then single but networked online archives of printed content should make a big difference in writing the history of subcultural, avant-garde and artistic movements.

In this perspective, there's artwork dealing with archiving and paper that is able to reverse our usual perspective. Tim Schwartz's *Card Catalog*[8] is an installation made of a more than 2 meter long drawer containing a catalogue of the author's 7,390 songs on his iPod, in the form of single cards for each song. Here the "data" is understandable (cards referring to a song), but it's not a representation of the data in itself (there are no scores, or music-related information), but rather of its complete index. Here the author is playing with space and its different qualities (when it's virtual and when it's real) so a paper catalogue index would be substantially bigger even than the whole amount of data hosted in a tiny mp3 player. Nevertheless, the installation is made in a 19th century style, entirely in classic wood with a brass knob, and that is not by accident. It reminds us that paper catalogues are still there, in old libraries, after centuries, still representing knowledge through paper in a universal form, while the invisibleness of hard disks is still arcane to most of us. This element is reinforced ordering the cards in a faithful chronology of the author's listening habits, which gives an even stricter correspondence between the invisible data and its tangible paper index.

2. Neural Archive and Distributed Archives

I'm the editor of *Neural*,[9] a magazine printed since 1993. Since then (and even before), we've started to develop our own archive of books, catalogues, magazines, posters, ephemerals and various other old and new media. Actually it contains approximately 2,000 items. Although it's totally unlikely that this collection will be publicly available in Bari, where *Neural* is based, we want to share it as much as possible. So we started a few tests on how to make online and offline strategies to publish online the first complete references to the things we collected. And the *Neural* archive is really far from being conceived as a solitary and self-referential effort. After compiling the reference of every single item, step 2 is trying to establish a model

5. See: <http://www.radicalsoftware.org/>.
6. See: <http://english.utah.edu/eclipse/projects/LANGUAGE/>.
7. See: <http://psrf.detritus.net/>.
8. See: <http://www.timschwartz.org/card-catalog/>.
9. See: <http://www.neural.it>.

of an online archive, sufficiently simple (even rough) and abstract, which could eventually be adopted by other people willing to put their collections online. This model would be the most valuable outcome of our archiving efforts and it should be shared as a sort of "manual" ("how to make your own online archive") including instructions on how to use some free software to build it (we're testing a Content Management System-based solution). Big institutions have already developed their own, like, for example, V2_ in Rotterdam,[10] which has an impressive one, but we want to make our work worthwhile for small institutions and personal collectors.

Step 3 will be to make different archives, dealing with homogenous topics, to be searchable altogether. It would be accomplished through simple free indexing software that would periodically scan the content of each single archive and make an effortless dedicated search engine. The whole effort is aimed to build a method as simple as possible to make online archives that would reflect the physical ones.

The resulting scenario would then be to make "islands" of culture which are maintained by the respective responsible persons. If they'll reflect their respective physical archives, we'd have a reliable online representation, which would be searchable and would indeed publicly recognize the responsible efforts in making and maintaining printed material collections.

Step 4 is to help to build open source DIY book scanning machines, in order to start to digitize the respective productions and make them available online. Google uses industrial book scanners for its giant Google Books program (that seduces institutions convincing them to grant Google the access to their printed treasures in exchange for a free or cheap version of their digitized files, shared with Google, of course). But there's a small community dedicated to building similar scanners through a strictly open source process, thus sharing the whole methodology and the software part. It's a technical crucial passage, because once built and spread these machines would substantially speed up the process of digitizing entire series of (old and new) books or complete collections of independent magazines.

Neural has just started to discuss collaboration for a program with a few partners in the Balkan area. We want to build an archive or references for our entire collection, build a book scanner for each of us and start to digitize a part of our collection and make it available online. We're trying to get funding in order to go through all the steps (1 to 4), developing in parallel and in the end joining our respective archives.

In a way, both the processes (publishing references of items collected and even digitizing them) is a similar mechanism to the so called

seeding of p2p stuff. Seeding, in the p2p technical vocabulary, means that if you want to "own" something you have to share it, at least a part of it; that is exactly what we'd do with the opportunities of combining print and digital technologies. And this is something completely different from nostalgia, of course. Building these archives of references, and scanning small productions, according to publishers' and authors' will, would sometimes help to reconnect fragments that were lost over time, texts lent and never returned that would turn up again.

But, we'd ask ourselves: can we properly define these types of processes as "archiving"? I think we can't. All of the above is about "accessing", and not "archiving". Only in a few decades will we know whether jpg and pdf files will still be there. But starting to take the responsibility to make them and share them is the first constructive step we can take.

Conclusions

The resulting scenario could be of archived "islands" of culture slowly and independently emerging online, and then growing. They'd be made by people who share passion and just want to share valuable information and contribute to the access to important content that would be new for the web. It'd be a shared memory as peer-to-peer unquestionably proved to be. It'd be coupling the stability of our printed culture, and its being static, with the ephemerality of its digitalization, and even more its consequent dynamic characteristics. However, such a project requires permanent "seeders" to stick with the peer-to-peer parallel, or people who want to be responsible. That's why we at *Neural* have already started working on what we have thus far acquired, while simultaneously networking with others. If we'll be able to establish usable models, platforms and practices, we'll probably help a small portion of culture to survive its own future.

Reference

RIFKIN, Jeremy (2000) *The age of access*. New York: Tarcher/Putman. <http://74.125.77.132/search?q=cache:sPcBjxhOvdQJ:cf.hum. uva.nl/bai/home/eketelaar/Timemachine.doc+rifkin+The+physi cal+container+becomes+secondary&cd=1&hl=en&ct=clnk&cli ent=safari>

10. See: <http://archive.v2.nl/>.

Recommended Citation

CV

Alessandro Ludovico
Editor in chief of *Neural* magazine
a.ludovico@neural.it

Alessandro Ludovico is a media critic and editor in chief of *Neural* magazine since 1993. He is the author of several essays on digital culture, and has lectured worldwide. He's one of the founding contributors of the *Nettime* community, one of the founders of the Mag.Net (Electronic Cultural Publishers) organization and he co-edited the *Mag.Net Reader* book series. He teaches at the Academy of Art in Carrara and in Bari and has been a research fellow at the Willem de Kooning Academy. He also served as an advisor for the *Documenta 12's Magazine* Project. With Paolo Cirio and Ubermorgen he developed *Google Will Eat Itself* (Honorary Mention Prix Ars Electronica 2005, Rhizome Commission 2005, nomination Prix Transmediale 2006) and *Amazon Noir* (1st prize Stuttgarter Filmwinter 2007, Honorary Mention Share Prize 2007, 2nd prize Transmediale08) art projects.

UOC
Universitat Oberta de Catalunya
www.uoc.edu

laboral
Centro de Arte y
Creación Industrial