



HAL
open science

Neglect, Stammering, Focus: Processes of an Archival Experience of Archive Collections and Audiovisual Projects at AMU Posted on the Web Over the Past 30 Years

Véronique Ginouvès

► To cite this version:

Véronique Ginouvès. Neglect, Stammering, Focus: Processes of an Archival Experience of Archive Collections and Audiovisual Projects at AMU Posted on the Web Over the Past 30 Years. RESAW2025 – 6th Conference : The Datafied Web, University of Siegen; Collaborative Research Center "Media of Cooperation", Jun 2025, Siegen, Germany. <halshs-05113368>

HAL Id: halshs-05113368

<https://shs.hal.science/halshs-05113368v1>

Submitted on 15 Jun 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Neglect, Stammering, Focus: Processes of an Archival Experience of Archive Collections and Audiovisual Projects at AMU Posted on the Web Over the Past 30 Years

Véronique Ginouvès (IRHC CNRS, UAR3125)

Part of

RESAW2025 – 6th Conference : *The Datafied Web*, Siegen, 5-6 June 2025.

Session 6B: Panel: The Challenge of Archival Practices' Context for a Better Understanding of Data Web Archives at Aix-Marseille University
Location: Obergraben Pink Room (first floor) 11:00
Mara Bertelsen, Sophie Gebeil, Véronique Ginouvès, Christine Mussard

Thank you for this opportunity to present our work at the 10th anniversary edition of the RESAW conference. I'm going to share with you a personal and institutional journey through 30 years of digital work with sound and audiovisual archives at the University of Aix-Marseille. A journey shaped by three words — *neglect*, *stammering*, and *focus* — which reflect how our relationship with web archiving has developed, with all its hesitations and realizations along the way.

Part 1 - Neglect

I have been an archivist at the CNRS since 1995¹. In those days, I was the only person in TELEMMe², my research unit, sending emails on behalf of researchers using my university address. Unfortunately, I never thought to preserve them, and I never printed out those messages.

When I joined the project to create the media library at the MMSH, I was responsible for the sound archives. I took charge of a sound archives database built with a software package called *Texto-Logotel*, developed by the company

¹ I have already written a post on the digital technologies I have come across during my professional experiences (2018), but here I'd rather look at the technological disappearances, and specifically the websites I have been involved with.

² Maryline Crivello (dir.), 2014.

*Chemdata*³, created in 1975. One of its key strengths was its ability to organize structured data and export it easily in ASCII⁴ format. The software was widely used within the French CNRS community⁵ and even had some international reach. As early as 1977, major CNRS bibliographic databases like *Pascal* and *Francis*⁶ were already using it.

Meanwhile, books and journals were catalogued by various laboratories, all of which used *Texto* for their bibliographic records. At its time, *Texto* was promoted as offering “modern documentation management”⁷. In the professional literature⁸, one could read: “The adoption of documentary software like *Texto* (...) offers, in our view, a better guarantee of data sustainability than the proprietary systems commonly used in libraries.”

In France, the most cutting-edge journal in the field of information science was *Le Médiéviste et l'ordinateur*, created in 1979. In its 1985 issue n°14, reviews of *Texto* were enthusiastic: the software was seen as efficient, flexible, and delivering “positive results.” It was also praised for its portability and ease of configuration. Most importantly, it allowed for easy export in plain text or structured ASCII files.

At the time, I felt reassured. When Internet arrived, I believed the software would evolve with the times, and that our data and workflows continue to be supported. But that didn't happen — the evolution of digital tools was too drastic. However, *Texto* was acquired in 1998 by a Company named CINCOM System, and it simply vanished.

Before this evaporation, the Web already existed — but *Texto* offered no tools to interface with it. So, I improvised a website, based on MySQL exports of our sound archive database⁹. In the early days of the web, when information was still scarce, the catalog attracted attention. But with the acquisition by CINCOM's far less capable product, I had to find new software and rebuild everything. At the time, I was just relieved not to lose any data. But I couldn't

³ <https://www.chemdata.fr> : la première capture du site de la Wayback Machine date du 29 décembre 1996, puis à partir de l'url est redirigée vers la société CINCOM puis une page générique présent CINCOM.

⁴ American Standard Code for Information Interchange.

⁵ Thérèse Tréfeu, 1991.

⁶ <https://pascal-francis.inist.fr>

⁷ Thérèse Tréfeu, 1991.

⁸ Hervé le Crosnier, 1990.

⁹ In the field of ethnology, a group of archivists has been working on structuring sound data using TEXTO software since the 1990s. A first manual was published in 1997, distributed by the Fédération des associations de musiques et danses traditionnelles (FAMDT). Between 2001 and 2007, new versions of the manual have been published, still focusing on structuring information to describe sound, but independent of any specific computer system. It's fascinating to realise that this type of questioning no longer arises today.

recover the first version of the site I had created. It's only because I remember people contacting me and talking to me about the contents of the database that I remember it existed.

As I was writing this text, I was reminded of a site I co-created in 2003 called 'Bibenligne'. In a crowded web space, we wanted to create a rarity by writing a reasoned selection of catalogs of libraries and documentation centers in the Humanities and Social Sciences that could be searched online in the field of Mediterranean studies. The site ran on ASP-Nuke (v1.2), an open software. Unfortunately, at the time, the MMSH's Microsoft-evangelist IT strongly dislike open-source tools. To be free, we transferred the records to the Revues.org platform in 2005, and once again I forgot to archive the site. Thankfully, *Internet Archives* had harvested it, and I found *Bibenligne* while preparing this conference paper — a small relief.

At the MMSH, several researchers who had built entire projects with *Texto* saw them disappear. The database was migrated to a proprietary software called LORIS from EVER Company, in UNIMARC format. Unfortunately, the research laboratory IREMAM whose library had best documented its articles and works with long abstracts did not preserve a digital copy of its original catalog, nor was it ever put online. The *Bibliothèque du monde arabe / The Library of the Arabic World*, was never not exported or preserved. As a result, 20 years of documentary work disappeared in the blink of an eye, reduced to a simplified UNIMARC¹⁰ format with a software that didn't support Arabic script. Yet this database had been forward-thinking, publishing printed catalogues of its entries every year. Today only these printed catalogs remain¹¹, but readers struggle to understand what these correspond to

SO --- What remains from that period? A few user memories, a few pages on *Internet Archives*, printed internship reports, printed catalogs, handwritten or annotated data entry forms, and improvised exports in formats that are not always readable or understandable. What has disappeared, for the IREMAM laboratory is years of electronic documentary work — and, more broadly, the ability to understand how research was structured and conceptualized at the time. As for the rest, as archivist, I thought that simply transferring data from one application to another without losing information would be enough to

¹⁰ UNiversal MACHine Readable Cataloguing.

¹¹ However, these publications are not so easy to find through a bibliographic search under a number of titles: *Travaux et documents de l'Iremam*, *Cahiers de l'IREMAM* and *Annuaire d'Afrique du Nord*.

archive it, but I didn't realize that the interfaces — and the conceptual frameworks behind them — would vanish too. As an archivist, that's a shame.

Part 2 - Stammering

This lack of anticipation was not the result of deliberate negligence, but rather a form of collective blindness: we assumed these tools would still be there tomorrow—and the day after.

Working on large-scale platforms, such as migrating the *Bibenligne* website to *Revues.org* or launching a blog on *Hypotheses.org*, gradually made us aware of the importance of archiving our online content. It was only around 2005 that we began to archive our materials intentionally and systematically using the *Wayback Machine*. At that time, I discovered—like with *Bibenligne*—that Archive.org had already archived a lot of our sites, even without our intervention. That meant these sites were significant, didn't they? At least that's how I understood it back then, and I was rather proud of it.

But I had to learn to work with loss and disappearance. Example, in 1995, we established a space for critical engagement with digital practices in the humanities and social sciences: the “Pôle image-sons, pratiques du numérique.” It was a place for discussion, training, and experimental digital audiovisual projects — including a pioneering electronic journal *Le Recueil électronique*. Most of these projects were developed using *Adobe Flash*. At the time, the technology offered great creative freedom—interactive animations, dynamic interfaces, immersive experiences. But this creative freedom eventually came out to be a technological trap. Discussions about *Flash* were already contentious back then. When official support for *Flash* ended in 2020, many of our digital publications suddenly became inaccessible.

This time, we realized in time. Thanks to tools such as *Conifer* (formerly *Webrecorder*) and the *Wayback Machine*, we are now able to emulate several of these projects in their original environments. This experience underscores the importance of continuous technical monitoring and of designing system for long-term preservation in mind from the outset: choosing open formats, planning dynamic captures, and documenting software dependencies. *Flash* is not an isolated case—it serves as a reminder that today's dominant technologies can quickly become digital graveyards if we don't anticipate their obsolescence.

Both long-term and sudden historical events—such as the war in Syria (2010–2024), or the October 17 uprising in Beirut—have compelled us to learn how to archive events as they unfold. Jean-Christophe Peyssard, Head of the MMSH Mediathèque has worked on this with our team. He couldn't join us today but Sophie Gebeil will tell you more about the WebLab we are currently developing at AMU. On the side of our archive catalogs, in 2020, when the database software we had been using was acquired, we had to perform a full metadata retroconversion.

Today, our data are structured in EAD — Encoded Archival Description — an international XML standard used to describe archival collections. They have been integrated into *Calames*, the national platform managed by the French Ministry of Higher Education and Research. Our catalog is now harvested by several platforms, France Archives, the European Archives Portal, the Arab Phonograph Search project, or Isidore.

Over time, we've realized that ubiquity is a preservation strategy in itself — spreading content across systems and platforms helps ensure its survival. In a way, disappearance has become a constant companion — compelling us to question our practices and the lifespan of what we produce.

We are currently in the process of officially archiving our collections through the Centre Informatique National de l'Enseignement Supérieur - CINES, France's national government-funded infrastructure under the Ministry of Higher Education and Research, responsible for the long-term preservation of scientific and cultural heritage data.

Part 3 - Focus

But I still had to learn and lose. Between February 2014 and February 2017, I took part in the *Europeana Sounds programme* led by the *British Library*, involving 24 libraries across Europe. It was a wonderful experience — joyful, collective, and rigorous — that taught us a great deal, particularly in preparation for what would later be called, a few months after the project began, the FAIR principles¹².

The main goal was to make our data interoperable by “aggregating” it through the Europeana Data Model – EDM, into the *Europeana* platform, and to make it reusable by working throughout the project on ethical and legal aspects related to access and dissemination. A significant effort was made to engage

¹² Wilkinson and al., 2016.

audiences, either by encouraging participatory enrichment of the data or by experimenting with various multimedia tools for sharing it.

Throughout the project, we regularly submitted blog posts, along with audio files for which we had secured distribution rights, to the communications workpackage managed by the French National Library BnF. In reality, we lost track of how many posts we contributed — we assumed they would naturally be preserved by Europeana and its programmes, or by the BnF. But they weren't preserved.

At the end of the project, the blog was taken offline. Its entire content replaced by a single static information page. Our posts vanished from the web.

After a moment of stunned disbelief, we turned to the *Wayback Machine* to try and retrieve them — but it had not crawled the site consistently. Many of the posts were no longer there: 5 snapshots in 2014, 11 in 2015, 17 in 2016, and just 3 in early 2017, up to March, the end of the project. After that, only the generic information page remained.

Once again, we had to reconstruct the content from our drafts¹³.

But do we ever truly learn that websites disappear?

Today, the MMSH media library team is trained. Last year, we even invited Niels Brügger¹⁴ to lead a training session for colleagues from Aix-Marseille University who are interested in this issue. We continue to offer training through the Weblab created by Sophie Gebeil and Jean-Christophe Peyssard. And above all, we know — something that should be obvious to any archivist — that archiving requires constant vigilance.

But that vigilance is not self-evident. It demands an effort of memory and organisation in a digital environment where everything appears permanent... until it suddenly vanishes.

While I was drafting the abstract of this presentation, the *Wayback Machine* temporarily went offline due to a cyberattack — a genuine shock. I train researchers to cite websites using *Wayback Machine* URLs. Researchers are

¹³ The disappearance of this series of blog posts led the team of archivists at the MMSH Media Library to find all the drafts of the posts and their associated sound files. They were published in an Advent calendar dedicated to the question of disappearance : <https://phonotheque.hypotheses.org/tag/avent2020>

¹⁴ *Collecting, Using and Preserving the Web for Education and Research*, March 11, Aix-en-Provence <https://imageson.hypotheses.org/4038>

already finding it hard to follow this advice, so if the URL disappears too, how am I going to convince them to be vigilant?

Today, we have set up a process. We have exported our former database into Conifer. We cite *Wayback Machine* URLs when linking to our previous, still-functional online database. Fortunately, our research blog is hosted on *Hypotheses.org* and preserved by CINES. We also rely on the legal deposit of the web by INA and the BnF.

From having no strategy at all, we — as archivists — have integrated the “web tab” to our archival mindset.

Now, we systematically include web archiving in every digital project. We raise awareness among teams, diversify our preservation platforms, but forgetting still lurks in the background. In fact, while writing this presentation, I realised that telling the story of what was forgotten is itself an archival act.

Thank you for allowing me to share it.

Bibliographie

Kahle Brewster, « Internet Archive Services Update », *Internet Archive Blogs*, 2024-10-17, <https://blog.archive.org/2024/10/18/internet-archive-services-update-2024-10-17>

Maryline Crivello (ed.). « Essai d’ego-histoire collective. Une histoire du laboratoire TELEMME (1994-2004) », *Rives méditerranéennes*. octobre 2014 n° 48. p. 264.

Hervé Le Crosnier, *Systèmes d’accès à des ressources documentaires : vers des anté-serveurs intelligents*, Aix-Marseille III. Université de droit, d’économie et des sciences. 1990. 372 p. <https://theses.hal.science/tel-00004654v1>

Véronique Ginouvès « Parcours professionnel : entre livres, archives et données », *Histoires de bibliothèques*, 1^{er} août 2018, <https://doi.org/10.58079/cjs6>

Véronique Ginouvès et Véronique Pérennou, *Guide d'analyse documentaire du son inédit pour la mise en place de banque de données*. Document dactylographié, Fédération des associations de musiques et danses traditionnelles. Commission Documentation. 1997. 102 p.

Alessandro Sturla, « Interaction entre transmission et diffusion de l'information », *Le médiéviste et l'ordinateur*. Automne 1985 n° 14. p. 14-15.

Thérèse Tréfeu, « La politique d'information scientifique et technique en France : l'informatisation dans les instituts scientifiques et les échanges Nord-Sud (1979-1989). Document de travail », Document dactylographié publié par le Ministère de la Coopération et du Développement, département de l'information et de la documentation, 1991, 150p.

Mark D Wilkinson, Michel Dumontier and Barend Mons « The FAIR Guiding Principles for scientific data management and stewardship », *Scientific Data*. 2016, vol.3 n° 160018. <https://www.nature.com/articles/sdata201618>

Journals

Annuaire d'Afrique du Nord :

https://web.archive.org/web/20250000000000*/https://cinumed.mmsch.univ-aix.fr/aan

Cahiers de l'IREMAM: <https://books.openedition.org/iremam/2632>

Travaux et documents de l'Iremam: <https://www.sudoc.fr/039938093>

Web sites

Arab Phonograph Search project (work in progress):

<https://agya.info/research/research-projects-by-year/apsearch-a-search-engine-for-historical-sound-recordings>

Bibliothèque nationale de France, legal deposit: <https://www.bnf.fr/fr/le-depot-legal>

Calames - Collections of manuscripts and archives of French higher education and research: <https://calames.abes.fr>

Carnets du secteur Archives de la recherche – Phonothèque de la Médiathèque de la MMSH: <https://phonotheque.hypotheses.org>

Catalogue des archives de France: <https://francearchives.gouv.fr>

CINES – French National Computer Center for Higher Education:

<https://www.cines.fr>

Europeana: <https://www.europeana.eu/fr>

Europeana Sounds: <https://www.eusounds.eu>

Pôle images, sons, pratiques du numérique en sciences humaines et sociales:

<https://imageson.hypotheses.org>

European Archives Portal: <https://www.archivesportaleurope.net>

Inathèque: <https://catalogue.ina.fr>

Isidore - search assistant in Humanities and Social Sciences:

<https://isidore.science>

URLs Wayback Machine

ASP Nuke

<https://web.archive.org/web/20040619151403/http://www.asp-nuke.com>

Bibenligne

https://web.archive.org/web/20040615000000*/http://bibenligne.mmsh.univ-aix.fr

Europeana Sounds

https://web.archive.org/web/20250000000000*/https://www.eusounds.eu

Recueil électronique du Pôle images, sons, pratiques du numérique en SHS

<https://web.archive.org/web/20140812045504/http://recueil.mmsh.univ-aix.fr/htmlaccueil/accueil.html>

Société Chemdata :

Premier moissonnage de la Wayback Machine :

https://web.archive.org/web/19961201000000*/http://www.chemdata.fr/

Page Chemdata.fr renvoyant vers Novacite :

<https://web.archive.org/web/19971022205643/http://www.chemdata.fr>

Page Chemdata société intégrée dans le troupe CINCOM :

<https://web.archive.org/web/19980213055551/http://www.chemdata.fr>

Erreur 404 :

<https://web.archive.org/web/20000409091744/http://www.chemdata.fr/>

Page générique CINCOM :

<https://web.archive.org/web/20100213175920/http://www.chemdata.fr>

Société Ever – Loris

<https://web.archive.org/web/20250607185738/https://www.societe.com/societe/loris-enr-493286355.html>