

Using digital technologies to reconstitute the provenance of musical instruments: an overview of new methodological perspectives

The Musée de la musique considers provenance research as a central priority in its current missions. This commitment, outlined in its scientific and cultural project (2020-2025)¹, is reflected through several actions. A comprehensive study of the provenance of the musical instruments in its collection has been initiated, aiming to shed light on their history and question the legitimacy of their presence in the collections. The first phase of this study focused on a corpus of 2,343 objects acquired after 1933 and manufactured before 1945, and successfully traced the provenance of 652 instruments, without revealing any cases of Nazi-era spoliation.² The museum is currently continuing this study by expanding its corpus to include other works in the collection and future acquisitions. It also transparently showcases the results of this research on the collections platform³ as well as in the museum's permanent exhibition. Additionally, the institution is actively involved in making available the archives of instrument makers from its collection, to enhance access to these essential sources for provenance research.⁴ The museum's dedicated archive website offers several digitized collections, accessible freely online.⁵ Furthermore, the Musée de la musique plays an active role in the provenance research network, while fulfilling its advisory role to public missions and institutions regarding musical instruments. In this context, the symposium *The Spoliation of Musical Instruments in Europe. 1933-1945*, organised on April 7, 8, and 9, 2022, by the Cité de la Musique – Philharmonie de Paris⁶, played a key role in fostering collective momentum among researchers, curators, and art market professionals, while providing an initial overview of current methodologies and research. The museum likewise acknowledges the importance of strengthening collaborations with other stakeholders and institutions. The study "Documenting the Violin Trade in Paris: the Archives of Albert Caressa and Émile Français, 1930-1945", conducted in collaboration with Carla Shapreau and published in the reference work *Collecting and Provenance. A Multidisciplinary Approach*, is a good example of this.⁷ Additionally, the Musée de la musique has aimed to solidify this commitment over the long term by creating a permanent position for a provenance research officer.

In this context, the museum aims to contribute to the development of innovative research methodologies and has chosen to leverage digital technologies, positioning itself within the field of digital humanities. Over the coming years, it plans to strengthen this ambitious approach to facilitate the research process and improve the quality of results. Digital tools indeed offer an innovative response to the challenges of provenance research, especially for musical instruments, where traceability relies on the meticulous analysis of scattered, multilingual, and sometimes incomplete resources. This demanding activity, in terms of time and human resources, can be greatly optimized through emerging technologies.

The aim of this presentation is to showcase the potential of digital tools for provenance research. After defining digital humanities, the aim will be to demonstrate that the potential of digital technologies remains underutilized in the field of provenance research. An overview of their contributions will then be provided, followed by a case study conducted at the museum to illustrate these benefits.

I. Digital humanities: an underutilized approach in provenance research methodologies

a. The rise of a transdisciplinary approach with great potential...

The digital era, characterized by the widespread use of computers, the internet, smartphones, along with the explosion of digital data and the development of artificial intelligence, impacts all sectors, especially research in the humanities and social sciences, which are experiencing a deep transformation in their practices. "The age of digital humanities has arrived. It promises not to be brief," states Olivier Le Deuff in his book *Le temps des humanités digitales. La mutation des sciences humaines et sociales*.⁸

The beginnings of this transformation actually date back to the 1950s with the work of Roberto Busa (1913-2011). With the rise of early computers, this Italian Jesuit was the first to use computing tools on textual corpora with the aim of "improving the quality, depth, and scope of [his] research."⁹ This approach, blending digital technologies and the humanities, gradually developed throughout the second half of the 20th century. However, it was only in the early 21st century, with the emergence of the web and artificial intelligence in particular, that this transdisciplinary approach truly expanded. Thanks to these technological advancements, digital humanities have become a broader field, drawing on a variety of technologies (digitization, mapping, data visualization, natural language processing ...) applied to many humanities disciplines (history, art history, literature, geography, archaeology, sociology). The term "digital humanities" emerged alongside these developments and was defined in 2010 as an approach that:

"rely, [...], on the full range of paradigms, expertise, and knowledge specific to [the Humanities, Arts, and Letters], while also leveraging the unique tools and perspectives of the digital realm".¹⁰

In the digital humanities approach, digital technologies are not seen as an end in themselves, but rather as "mediating objects of historical knowledge."¹¹ They transform the *modus operandi* by changing the way sources are accessed, read, analyzed, and interpreted. They also expand the scope of research by allowing for the consideration of massive amounts of data, thus raising new questions. Digital humanities therefore offer new research perspectives for scholars in the humanities and social sciences.

b. ... but still underutilized in provenance research methodologies

Provenance research methodologies have been developing for several years now and are based on the examination of primary sources, material analysis, iconographic studies, and the collection of testimonies. These traditional approaches, often combined, allow researchers to reconstruct the provenance of artworks. However, they involve a research process that is often lengthy and fraught with obstacles, sometimes failing to yield conclusive results.

To facilitate these investigations, several initiatives using digital technologies have been implemented, primarily focusing on source accessibility. Indeed, institutions are increasingly conducting digitization campaigns and sharing the results online through image databases. For instance, a recent example is the digitization carried out by the Ministry of Foreign Affairs of the archives related to Artistic Recovery (209SUP collection).¹² Moreover, text databases have proliferated around provenance data (Lootedart, Répertoire des Biens Spoliés, Einsatzstab Reichsleiter Rosenberg archives, Répertoire des Acteurs du Marché de l'Art (RAMA), ...). Some of these offer transcriptions of sources, creating numerous queryable datasets. For example, the Getty Provenance provides a

manual transcription of the dealers' stock books from the Knoedler Gallery.¹³ These image and text databases are essential for researchers and provide a significant time-saving advantage.

However, although these new tools are based on digital technologies, they do not allow provenance research to be fully integrated into the field of digital humanities. The contribution of digital technologies is, in fact, underutilized, as the tools used by the provenance research community do not fundamentally change research practices. Yet, it is possible to take advantage of more advanced tools that would enable the development of new research methodologies.

II. Toward new perspectives: the potential of digital humanities applied to provenance research

In order to consider the following different directions, it is essential that the archives and research data be made available. It is also necessary for the tools used to be open source.

a. Automating the data conversion of sources

The "image" digitization of sources does not allow for the direct use of the content of the images. For example, it is impossible to conduct a keyword search within the sources to quickly access the desired information, which necessitates a complete reading of the archive. In contrast, manual transcription of the sources allows for content exploitation, but it is not an optimal solution and is rarely used.

To overcome these limitations, other technologies can be considered to ensure a faster, more usable, and higher-quality data conversion process. The use of image analysis technologies, such as Optical Character Recognition (OCR) or Handwritten Text Recognition (HTR), enables the automatic conversion of image content into usable data. The deployment of these technologies could significantly increase the number of sources available online.

b. Exploiting big data

Provenance researchers consult a large number of archives to trace the history of objects. These archives are often incomplete, necessitating meticulous cross-referencing of sources. Digital technologies have the advantage of optimizing this research by enabling massive data processing, based on a key principle: automation.

To exploit data on a large scale, technologies allow for the "remote" reading of sources.¹⁴ Several data mining methods exist for extensively utilizing sources, such as topic modeling, which highlights themes synthesizing the content of a set of archives, or named entity recognition, which automatically extracts names of people, organisations or companies, locations, quantities, distances, values, dates, etc., from sources.

Furthermore, the handling of numerous sources can be facilitated through advanced reading features within various distribution platforms. Indeed, collaborative platforms, interactive websites, and electronic editions can be enhanced with various reading functionalities: search engines, visualization tools, indexes, filters, and so on. All of these reading aids allow for quicker reading and make it easier to expand research. Other technologies related to the semantic web could be integrated into these platforms to create connections between provenance data and facilitate their cross-referencing.

c. Ensuring the longevity of data

Ensuring the longevity of provenance data is essential for researchers, yet it remains a marginal topic within the provenance research community. However, digital technologies offer various solutions to meet this challenge through different practices and tools. It is crucial to adopt these solutions to guarantee the accessibility of data for future generations. The use of open and standard data formats (such as CSV, XML, TEI, PDF/A) is essential to ensure the sharing and longevity of information. Moreover, metadata plays a key role by providing context, provenance, and structure to the data. Without metadata, information may lose its relevance over time. This metadata must also adhere to established standards, such as EAD for archives.

Furthermore, various tools and media (cloud, hard drive, remote server) are indispensable for ensuring regular data backups and preventing data loss. Digital archiving services, such as HAL, Zenodo, or Internet Archive, guarantee the preservation of data within appropriate infrastructures, with processes for updates and technological migration. Finally, it is crucial to document the data, tools, methods, and formats used to enable their reuse by future researchers.

III. A practical example of a computational method applied to dealers' stock books¹⁵

a. Defining a method for automatically deducing incomplete chronological data

The Musée de la musique wished to experiment with this approach to digital humanities by relying on a source particularly rich in provenance information: the dealers' stock books. These documents trace the purchase and sale transactions of objects, providing more or less complete details about the materiality of the instruments, the periods of ownership by the dealer, the names of previous and subsequent owners, as well as the terms of transaction. The dealers' stock books form a massive corpus to explore. This is why we deemed it relevant to employ digital technologies to establish a methodology for automated exploitation aimed at massively reconstructing the provenance of musical instruments.

To achieve this, we focused our work on a restricted corpus consisting of three purchase-sale registers of old instruments from a major violin-making workshop of the 19th and 20th centuries : Gand & Bernardel. Maintained between approximately 1840 and 1921, these three registers titled "N°1 Signalements", "N°2 Signalements" et "N°3 Signalements"¹⁶ reflect about a century of commercial operations. More than 2,950 entries of instruments are recorded, including multiple entries for the same instrument. However, some records exhibited gaps, particularly concerning temporal information, which could be indicated incompletely (for example, month and year or just the year) or completely absent from the registers. Of all the instrument entries, we were able to reconstruct only 2.03% of the periods of ownership by the workshop. We also noted that purchase dates were specified much less frequently than sale dates.

In light of this observation, we developed a computational method to estimate an approximate purchase period. This method allowed us to deduce earliest and latest possible purchase dates, aiming to reconstruct the period of ownership by the workshop for all instrument transactions.

The detail of this method is primarily based on a logic specific to the dealers' stock books: the general principles of deduction rely on known chronological data and the order of instruments entering the workshop. In addition to

these internal elements of the registers, this method requires that all chronological data be taken into account. Thus, digital technologies proved essential for implementing this estimative approach.

To apply this deduction based on all the data, a digital processing chain was established. A manual transcription of all chronological data was performed in a spreadsheet to obtain datasets usable by the machine. Next, the general principles for deducing purchase dates were translated into a series of algorithmic instructions using the Python programming language (version 3.10.9). These algorithms were finally applied to the produced datasets, resulting in new datasets containing enriched purchase dates.

b. Evaluation of the contribution of the method

This computational method has allowed for an estimated increase in the number of purchase dates for the violin section by over 95 points compared to the initial data, and by nearly 50 points for each of the other sections (violas, cellos, guitars). From these new data, it became possible to reconstruct an estimated 69.40% of the periods of ownership by the workshop, compared to just 2.03% initially.

In addition to this massive data deduction, the algorithms and documentation will be made available online to promote the reuse of this method on other dealer stock books.

In light of this presentation, it has been shown that digital humanities represent a relevant approach for provenance researchers. Digital technologies enable the optimization of the research process and expand the analysis to much larger data corpora. They thus provide solutions to the challenges faced in provenance research while opening up new perspectives for researchers.

Despite the initiatives developed in this direction, the provenance research community has not yet fully exploited the potential of digital technologies. Limited to digitization and the dissemination of databases, digital tools can offer much more. The automatic conversion of sources into usable data, as well as the development of data exploitation tools, are, as we have seen, essential methods for enhancing research in this field. However, these digital tools should be seen as a valuable complement to traditional working methodologies, without replacing them. They provide effective support for researchers, but traditional approaches remain essential to ensure the rigor and depth of analyses.

¹ Cité de la musique – Philharmonie de Paris, *Projet scientifique et culturel du Musée de la musique (2020-2025)*, december 2019, <https://deneb.philharmoniedeparis.fr/uploads/documents/60cb0da5c6c21_projet-scientifique-culturel-musee-2020-pdf-accessible-web_1.pdf>.

² Caër Mathilde, Gaudin Fabienne, "Recherche en provenance sur les acquisitions du Musée de la musique depuis 1933 : état des travaux et premiers résultats", symposium *La spoliation des instruments de musique en Europe. 1933-1945*, april 2022, Paris, France.

³ Cité de la musique – Philharmonie de Paris, *Collections du musée de la musique*, n.d., <<https://collectionsdumusee.philharmoniedeparis.fr/>>.

⁴ Laloue Christine, "Les actions du Musée de la musique pour la conservation et la communication des archives de la facture instrumentale", symposium *La spoliation des instruments de musique en Europe. 1933-1945*, april 2022, Paris, France.

⁵ Cité de la musique – Philharmonie de Paris, *Archives du musée de la musique*, n.d., <<https://archivesmusee.philharmoniedeparis.fr/>>.

⁶ Conference available for streaming: <<https://pad.philharmoniedeparis.fr/colloque-the-spoliation-of-musical-instruments-in-europe-1933-1945.aspx>>.

⁷ Shapreau Carla, Échard Jean-Philippe, Laloue Christine, "Documenting the Violin Trade in Paris: the Archives of Albert Caressa and Émile Français, 1930-1945", in Milosch Jane C., Pearce Nick, *Collecting and Provenance: A Multidisciplinary Approach*, Rowman & Littlefield, 2019, pp.189-204.

⁸ Le Deuff Olivier, "Introduction", *Le temps des humanités digitales : la mutation des sciences humaines et sociales*, fyp, 2017, p. 11.

⁹ Ter Braake Serge, Fokkens Antske, Ockeloen Niels, Van Son Chantal, "Digital History: Towards New Methodologies", 2nd International Workshop on Computational History and Data-Driven Humanities (CHDDH), May 2016, Dublin, Ireland. pp. 23-32, <<https://inria.hal.science/hal-01616305>>.

¹⁰ Mounier Pierre, "Manifeste des Digital Humanities", *Journal des anthropologues*, 2010, pp. 122-123, <<http://journals.openedition.org/jda/3652>>.

¹¹ Kaplan Frédéric, Fournier Mélanie, Nuessli Marc-Antoine, "L'historien et l'algorithme", in *Le temps des humanités digitales*, op. cit., p. 50.

¹² The digitizations are available on the website of the digitized archives of the Ministry of Europe and Foreign Affairs, <<https://archivesdiplomatiques.diplomatie.gouv.fr/ark:/14366/xdm9rgwv5nt1>>.

¹³ Datasets available on the GitHub platform <<https://github.com/thegetty/provenance-index-csv>>.

¹⁴ Concept of "distant reading" invented by Franco Moretti (Italian literary critic).

¹⁵ For the details of this study, refer to: Échard Jean-Philippe, Lebreton Fanny, Laloue Christine, "Digital approach to reconstructing the provenance of old objects: a study of an automated method for exploiting chronological data from art trade archives", manuscript in preparation.

¹⁶ These three registers are kept at the Musée de la musique in Paris and are listed under inventory numbers E.981.8.51, E.981.8.52, and E.981.8.53.

Bibliography

Scientific and cultural project and publication

- Cité de la musique – Philharmonie de Paris, *Projet scientifique et culturel du Musée de la musique (2020-2025)*, december 2019, <https://deneb.philharmoniedeparis.fr/uploads/documents/60cb0da5c6c21_projet-scientifique-culturel-musee-2020-pdf-accessible-web_1.pdf>.
- Rygiel Philippe, *Historien à l'âge numérique*. Presses de l'enssib, 2017, <<https://books.openedition.org/pressesenssib/6303>>.

Articles in journals and collective works

- Bonfait Olivier, Courtin Antoine, Klammt Anne, "Humanités numériques et histoire de l'art en France : état de la recherche et perspectives", *Histoire de l'art*, 2021, 87, pp. 5-16, <<https://hal.science/hal-03352544>>.
- Clavert Frédéric, "Lecture des sources historiennes à l'ère numérique", *L'histoire contemporaine à l'ère numérique*, 2012, <<https://histnum.hypotheses.org/1061>>.
- Clavert Frédéric, Daniel Johanna, Fleckinger Hélène, Grandjean Martin et Idmhand Fatiha, "Histoire et humanités numériques : nouveaux terrains de dialogue entre les archives et la recherche", *La Gazette des Archives*, 2017, Meta/morphoses, Les archives, bouillons de culture numérique, 245 (1), pp.121-134, <<https://halshs.archives-ouvertes.fr/halshs-01521814/document>>.
- Grandi Elisa, Ruiz Émilien, "Ce que le numérique fait à l'historien.ne", *Diacronie*, OpenEdition Journals, N° 10, 2, 2012, <<http://journals.openedition.org/diacronie/2780>>.
- Heimburger Franziska, Ruiz Émilien, "Has the Historian's craft gone digital? Some observations from France", *Diacronie*, N° 10, 2, 2012, <<http://journals.openedition.org/diacronie/2795>>.
- Le Deuff Olivier, *Le temps des humanités digitales : la mutation des sciences humaines et sociales*, fyp, 2017.

-
- Mounier Pierre, "Manifeste des Digital Humanities", *Journal des anthropologues*, 2010, pp. 122-123, <<http://journals.openedition.org/jda/3652>>.
 - Mounier Pierre, *Read/Write Book 2*, OpenEdition Press, 2012, <<https://doi.org/10.4000/books.oep.226>>.
 - Shapreau Carla, Échard Jean-Philippe, Laloue Christine, "Documenting the Violin Trade in Paris: the Archives of Albert Caressa and Émile Français, 1930-1945", in Milosch Jane C., Pearce Nick, *Collecting and Provenance: A Multidisciplinary Approach*, Rowman & Littlefield, 2019, pp.189-204.
 - Sinatra Michaël E., Vitali-Rosati Marcello, "Histoire des humanités numériques", *Pratiques de l'édition numérique*, Montreal, 2014, (Parcours numérique), pp. 49-60, <<http://books.openedition.org/pum/317>>.
 - Ter Braake Serge, Fokkens Antske, Ockeloen Niels, Van Son Chantal, "Digital History: Towards New Methodologies", 2nd International Workshop on Computational History and Data-Driven Humanities (CHDDH), May 2016, Dublin, Ireland, pp. 23-32, <<https://inria.hal.science/hal-01616305>>.

Symposium

- Caër Mathilde, Gaudin Fabienne, "Recherche en provenance sur les acquisitions du Musée de la musique depuis 1933 : état des travaux et premiers résultats", symposium *La spoliation des instruments de musique en Europe. 1933-1945*, avril 2022, Paris, France.
- Laloue Christine, "Les actions du Musée de la musique pour la conservation et la communication des archives de la facture instrumentale", symposium *La spoliation des instruments de musique en Europe. 1933-1945*, avril 2022, Paris, France.

Article in preparation

- Échard Jean-Philippe, Lebreton Fanny, Laloue Christine, "Digital approach to reconstructing the provenance of old objects: a study of an automated method for exploiting chronological data from art trade archives", manuscript in preparation.