

UNDER THE SURFACE: INTERPRETING TECHNICAL METADATA AS ARCHIVAL NARRATIVE

Gorini Adele

*University of Bologna
Italy*

adele.gorini2@unibo.it

ORCID <https://orcid.org/0009-0008-6137-7633>

Giagnolini Lucia

*University of Bologna
Italy*

Lucia.giagnolini2@unibo.it

ORCID <https://orcid.org/0000-0002-4876-2691>

Abstract – This study examines the role of technical metadata in the preservation and analysis of born-digital literary archives, focusing on the collections of contemporary Italian authors Silvia Avallone and Paolo Di Paolo housed at the Centro Manoscritti of the University of Pavia. Through the Digital Preservation Workflow (DPW) software developed within the Pavia Archivi Digitali (PAD) project, the research investigates how automatically extracted metadata can enhance understanding of digital documents across temporal, authorial, and textual dimensions.

The methodology combines forensic techniques with archival approaches, employing both Apache Tika and ExifTool to extract and interpret technical metadata. The comparative analysis reveals significant inconsistencies between file system metadata and embedded document properties, highlighting how creation dates, editing times, and authorship information require careful contextual interpretation.

Key findings include the discovery of systematic discrepancies in temporal metadata due to time zone normalization, the identification of complex authorial networks through creator fields, and insights into compositional practices through editing time and revision counts. The study also addresses the challenges in interpreting ambiguous metadata fields, particularly in legacy document formats.

This investigation establishes technical metadata as a valuable resource for both archival description and philological inquiry, capable of reconstructing creative processes and documenting authorial behaviors. However, it simultaneously emphasizes the necessity for interdisciplinary approaches that combine archival science, philology, and computational methods to effectively preserve and

interpret contemporary literary heritage in digital formats.

Keywords – Metadata, Born-digital, Literary archives

I. INTRODUCTION

Personal digital documents are often fluid, easily modifiable, and scattered across different devices, formats, and online platforms. Unlike paper-based materials, they are affected by the fast pace of technological change, such as outdated software or storage media. These features call for new archival methods that go beyond traditional practices and are specifically designed to handle the dynamic and distributed nature of digital content.

In response to these challenges in the context of literary archive preservation, Italy's first dedicated research initiative was launched in 2009: the *Pavia Archivi Digitali* (PAD) project, based at the Centro Manoscritti of the University of Pavia [3][5].

In the last two years, the project has focused on stabilizing the curation workflow, leading to the development of the Digital Preservation Workflow (DPW) software. This tool integrates all key stages of digital archival processing - from acquisition and preservation to access and consultation - within a single, cohesive environment [5].

A key feature of the DPW is the automated extraction of technical metadata from files within born-digital collections. These include file format, MIME type, character encoding, and language, as



well as document-specific properties such as title, author, creation and modification dates, and content length (e.g., word or page count). Additionally, it is possible to identify and extract embedded metadata from within files, such as information stored in document headers or associated with embedded objects.

II. METHODOLOGY

Technical metadata provides valuable insights into the archival history of the materials, supporting more accurate description, provenance tracking, and long-term preservation strategies.

At the same time, the meaning of the metadata is not always immediately apparent, and their understanding often requires careful interpretation. This interpretation must consider several factors, such as the type and format of the file in question, the type and version of the software used to create it, as well as the tools used for extraction, and the production and use context of those documents.

The extraction and interpretation of metadata of this kind have traditionally been carried out within the field of digital forensics, often for investigative and legal purposes [2] [6] [9]. This research aims to adapt and apply these methods from an archival perspective, exploring their potential to support the preservation, description, and critical analysis of born-digital literary materials.

The DPW software is designed for the analysis and extraction of metadata from various file formats. However, due to the literary nature of the digital archives preserved at the Centro Manoscritti, this study has primarily focused on the interpretation of metadata from textual files. In this context, the research relied on the metadata extracted by Apache Tika (v. 3.0.0) [1] and ExifTool (v. 13.00) [4], both integrated into the DPW.

Considering both tools is essential, as they each offer distinct methods for the automatic interpretation of a file, which can result in complementary or divergent metadata outputs. This dual approach allows for a more comprehensive and critical understanding of the structural and semantic information of the digital objects under analysis.

III. CASE STUDIES: SILVIA AVALLONE AND PAOLO DI PAOLO'S ARCHIVES

As part of an archival description project focused on born-digital materials, recent efforts have been directed toward the personal archives of Silvia Avallone and Paolo di Paolo, preserved at the Centro Manoscritti. Although these two archives were produced over a similar time period and originate from authors of the same generation, they exhibit distinct characteristics that make them particularly relevant for archival metadata analysis as well.

Silvia Avallone, born in Biella in 1984, is one of the most influential voices in the contemporary Italian literary scene. In the summer of 2011, she donated a selection of her born-digital materials to PAD. The core of the collection consists of 69 files organized into eight top-level folders, with a weight of 29.74 MB. The archive does not fully represent the complete virtual laboratory of the writer but constitutes a small, curated *corpus* that Avallone deliberately selected for consultation and preservation. The documents span from October 2007 to July 2011 and include PDFs, JPEGs, DOCs, and DOCX files. Her case exemplifies a form of author-led curation, where materials are consciously prepared and selected ahead of the transfer.

Paolo Di Paolo, born in Rome in 1983, donated his digital archive to PAD in 2015. Unlike Avallone's selective approach, Di Paolo chose to transfer almost the entirety of the materials stored on his digital devices at the time (a hard drive and a Macintosh computer). His archive includes 3,119 files, amounting to 5.43 GB, with a time frame ranging from 2000 to 2014. Except for personal documents and private correspondence, the archive offers a remarkably faithful portrait of the creative workshop of the writer. The heterogeneous extensions of the files (including DOC, DOCX, PDF, MP3, MP4, and JPEG) reflect the breadth of his activities: research, writing, reworking, and audio production. Together, these documents provide an in-depth insight into his professional and creative practices up to 2014.

The materials from these two archives have provided valuable input for studying and discussing



various archival dimensions conveyed through technical metadata – particularly the temporal, authorial, and textual dimensions.

IV. TEMPORAL DIMENSIONS

A. *The enigma of document dating*

With the use of the DPW application, notable discrepancies emerged between the creation dates as indicated by file system metadata, and the ones automatically extracted by Apache Tika and ExifTool.

On the one hand, in the Avallone collection, file system metadata associated with folders and with most of the textual files point to July 2011 as the date of their creation and last modification. This is the period during which Avallone was preparing the materials for transfer to PAD. However, when analyzing the metadata automatically extracted, different scenarios open up.

In the first place, extracted metadata often reveals more plausible creation dates. This is the case with Avallone's novel *Acciaio* (Rizzoli, 2010), where the Apache Tika library shows the metadata `dcterms:created = 2008-05-29T12:11:00Z`. Similarly, for the poetry collection *Il libro dei vent'anni* (Edizioni della Meridiana, 2007) `dcterms:created = 2006-12-19T21:44:00Z` is displayed.

In other cases, the metadata may suggest that the author transferred texts by copying or retyping them into new documents, thereby overwriting or erasing valuable philological and chronological traces. A notable example is the typescript *Alcune ragazze lo sanno.docx*. Indeed, the text we find in this file was published in the newspaper *Corriere della Sera* in August 2010, while the file carries the metadata `dcterms:created = 2010-11-06T09:05:00Z`, which postdates its publication. This retrospective analysis is further supported by metadata

concerning the total editing time of the file (see Section IV, Subsection B), which both Apache Tika and ExifTool report as five minutes. Although the document is relatively short (457 words), such timing is more consistent with a transcription activity rather than with the act of original composition.

As for the last modification dates, at first glance, the file system metadata and the information extracted through both Tika and ExifTool appear to be consistent. Nearly all Avallone's files were last modified between July 19 and July 20, 2011, suggesting that the author made final changes to the texts shortly before their transfer (see Table 1).

However, a closer inspection reveals a noteworthy detail: the automatically extracted modification dates differ from the file system timestamps by exactly two hours. The most plausible explanation for this systematic discrepancy lies in the handling of time zones.

Metadata extracted by Apache Tika and ExifTool are often normalized to Coordinated Universal Time (UTC), whereas file system metadata (such as those recorded by FAT) may reflect the local system time at the moment of file creation or modification. If the computer was operating under Central European Summer Time (CEST), i.e., UTC+2 – standard in Italy during the summer months – this would fully account for the observed two-hour difference.

On the other hand, in the Di Paolo collection, an analysis of the documents related to the novel *Dove eravate tutti* (Feltrinelli, 2011), allows for different considerations. The extracted metadata offer insights into a partial reconstruction of the novel's writing process: Paolo Di Paolo likely began drafting the text in 2009 and created at least ten different documents containing different drafting steps (see Table 2).



Table 1. Creation and last modification dates of Avallone's files, reported by the file system, Tika, and ExifTool.

File name	File system modification data	Tika (dcterms:created)	Exiftool(Create Date)	Tika (dcterms:modified)	Exiftool (Modify Date)
Acciaio – Dattiloscritto del 18 settembre2009.doc	7/19/2011 15:10	2008-05- 29T12:11:00Z	2008:05:29 12:11:00	2011-07- 19T13:10:00Z	2011:07:19 13:10:00
Il libro dei vent'anni.doc	7/19/2011 15:10	2006-12- 19T21:44:00Z	2006:12:19 21:44:00	2011-07- 19T13:11:00Z	2011:07:19 13:11:00
Alcune ragazze lo sanno.docx	7/19/2011 15:10	2010-11- 06T09:05:00Z	2010:11:06 09:05:00Z	2011-07- 19T13:23:00Z	2011:07:19 13:23:00Z

Table 2. Comparative overview of the temporal metadata associated with the documents related to Di Paolo's novel *Dove eravate tutti*.

File name	File system modification data	Tika (dcterms:created)	Exiftool(Create Date)	Tika (dcterms:modified)	Exiftool (Modify Date)
storia degli anni senza nome.doc	12/4/2009 15:48	2009-11- 25T08:42:00Z	2009:11:25 08:42:00	2009-12- 04T14:47:00Z	2009:12:04 14:47:00
crescere con Berlusconi.doc	12/30/2009 0:27	2009-12- 29T17:34:00Z	2009:12:29 17:34:00	2009-12- 29T23:27:00Z	2009:12:29 23:27:00
la cosa meno seria.doc	11/10/2010 10:08	2010-11- 10T09:07:40Z	2010:11:10 09:07:40	2010-11- 10T09:07:40Z	0
dov'eravate.doc	1/26/2011 11:24	2010-02- 15T13:18:00Z	2010:02:15 13:18:00	2011-01- 26T10:24:00Z	2011:01:26 10:24:00
dov'eravate tutti dic010.doc	3/3/2011 19:44	2010-02- 15T13:18:00Z	2010:02:15 13:18:00	2011-03- 03T18:44:00Z	2011:03:03 18:44:00
aggiunte.doc	3/15/2011 11:03	2011-03- 15T09:51:00Z	2011:03:15 09:51:00	2011-03- 15T10:03:00Z	2011:03:15 10:03:00
DOVE ERAVATE TUTTI 2011.doc	3/28/2011 19:46	2011-03- 03T18:53:00Z	2011:03:03 18:53:00	2011-03- 28T17:46:00Z	2011:03:28 17:46:00
incipit Paolo Di Paolo.doc	4/4/2011 13:32	2011-04- 04T11:17:00Z	2011:04:04 11:17:00	2011-04- 04T11:32:00Z	2011:04:04 11:32:00
DOVE ERAVATE TUTTI def maggio 2011.doc	5/18/2011 16:52	2011-05- 16T14:37:00Z	2011:05:16 14:37:00	2011-05- 18T14:52:00Z	2011:05:18 14:52:00
aggiunte Dove eravate tutti.doc	6/24/2011 23:27	2011-06- 10T14:29:00Z	2011:06:10 14:29:00	2011-06- 24T21:27:00Z	2011:06:24 21:27:00

The analysis of documents from both archives also highlighted how identifying the final modification made before publication – a crucial element for both philological and archival inquiry – proves to be a particularly complex task. In the digital context, the phenomena of *epigenesis* – that is, the continuation of a text's genesis even after its publication [10] – may occur with greater frequency, due to the ease and fluidity with which digital texts can be revised, duplicated, or updated. Paradoxically, such post-publication developments can also be more readily detectable, provided that a combined approach is adopted: one that integrates metadata analysis with automatic textual comparison techniques. The latter, however, depends on the availability of the final published text

for such operations, which cannot be taken for granted due to copyright restrictions.

This brief overview has shown how, even from a limited set of examples, the assessment of the temporal dimension of born-digital documents within a private archive requires the consideration of a complex interplay of technical and conceptual factors.



B. New data, new challenges

A particularly interesting and entirely unprecedented piece of information – compared to analog documents – is the metadata related to the total time spent working on the file.

While this information is intriguing, it is simultaneously quite obscure. For instance, Word documentation does not provide specific information on how this time is calculated, and there are instances where the data appears unreliable. For example, in some cases, the reported time seems disproportionately high when compared with other metadata, raising concerns about its accuracy.

A clear difference in the handling of this data emerges between Apache Tika and ExifTool results from these two libraries. The issue is especially evident in the data extracted by Apache Tika from Word documents with .doc extension. The result provided by Apache Tika in these cases often proves difficult to interpret: the reported values are unusually large, and the unit of measurement is not explicitly indicated. This suggests that the figure may represent a submultiple of the second, such as milliseconds or nanoseconds.

To address this inconsistency, integrating the ExifTool library into the analysis workflow proved crucial. Unlike Apache Tika, ExifTool provides the total time value in a more human-readable format, expressed in hours, minutes, or days. This makes it easier to read and interpret the data (and invites further investigation into how Apache Tika calculates and represents this particular field). A preliminary comparison suggests that Apache Tika may be reporting the value in nanoseconds but using an incorrect base, leading to misleading results.

For example, the file “aggiunte.doc” from Di Paolo’s archive shows a total editing time reported by Apache Tika as “600000000”, while ExifTool reports it as “1 minute” (see Table 3). Microsoft Word stores duration fields in units of ticks, where one tick equals 100 nanoseconds [8]. Apache Tika returns the raw value in ticks without converting it to seconds or minutes. If one mistakenly assumes that the

600000000 value represents nanoseconds, the conversion to minutes would yield 0.01 minutes, not 1 minute. This discrepancy does not indicate an error in Tika per se, but rather reflects that its output is in ticks, which can be misinterpreted if the user is unaware of the unit used by Word. ExifTool, by contrast, automatically converts ticks to seconds or minutes, producing the more immediately understandable value of 1 minute.

As previously mentioned, this issue only arises with .doc files and does not occur with .docx files. Metadata extracted from the latter is generally more consistent, readable, and aligned across different tools. This could be explained as the .doc files – in contrast to .docx – are based on a proprietary binary format, so they are more prone to anomalies, both in metadata structure and in the accuracy of their extraction.

Table 3. Total time spent working on a file: Tika vs ExifTool

File name	Tika (extended-properties:TotalTime)	ExifTool (Total Edit Time)
aggiunte.doc	600000000	1 minute
Alcune ragazze lo sanno.docx	5	5 minutes
la cosa meno seria.doc	0	0

Another important inconsistency was observed when comparing file size and content complexity with total time. In several cases, documents of substantial length and intricacy surprisingly displayed low editing times. For example, the Paolo Di Paolo document titled *la cosa meno seria.doc*, despite being composed of 1010 words, shows a total editing time of 0 (see Table 3). This may suggest that the content was copied and pasted into a new document, thus resetting the metadata.

These discrepancies highlight the need to approach the total time as an indicative rather than a definitive metric. It is a value that should always be interpreted with caution and contextualized through cross-analysis with other metadata fields (e.g., creation/modification dates, user logs) and internal textual elements. Furthermore, this example underscores the importance of cross-checking between metadata extraction libraries. One must always bear in mind that these tools act as



“interpreters” of the same underlying data, meaning that multiple (and potentially different) layers of interpretation exist between the raw bitstream and the final metadata we can read.

V. AUTHORIAL DIMENSIONS

A. Authorship Attribution

Another noteworthy aspect concerns the metadata related to authorship attribution. Files often appear to have been created by individuals other than the actual owners or creators of the donated archive. This discrepancy may come from the use of shared or institutional devices, collaborative writing environments, or the transfer and re-saving of documents over time – all of which complicate authorship attribution and pose significant challenges for the archival interpretation.

The presence of `dc:creator` metadata that differs from the name of the author may reveal underlying relationships with publishers, family members, or colleagues. For instance, in Avallone’s collection, the metadata of four text documents lists “Ritanna Armeni” as their creator – a detail that suggests a collaborative link between the two writers, likely related to Avallone’s contribution to an entry in the *Dizionario Femminile* edited by Armeni. However, the documents in question do not contain the published text. A possible explanation is that Avallone reused a document originally sent to her by another individual as a template for drafting new materials.

In the case of Paolo Di Paolo’s archive, the `dc:creator` metadata is occasionally associated with individuals other than him. In some instances, the metadata indicates a different name, yet the content corresponds to a work by Paolo Di Paolo, or his name appears at the end of the document. Conversely, there are cases in which texts authored by others appear under the `dc:creator` value “Paolo Di Paolo” suggesting that his computer may have been used by someone else or that he might have a role in the creative process. This situation introduces complications, particularly in the context of unpublished works, where distinguishing between different authors becomes difficult and can lead to the misattribution of authorship.

In one instance within Paolo Di Paolo’s archive, the `dc:creator` metadata contains two names, indicating a case of co-authorship between Di Paolo and another writer. This occurrence draws attention to another important aspect of the `dc:creator` field – namely, the possibility of manual input or editing within Microsoft Word. Word indeed allows users to freely modify the author field, even inserting more than one name as a single string.

In some cases, the presence of an editorial or institutional relationship can also be inferred from the `meta:last-author` or the `extended-properties:Company` (which typically records the organization associated with the Microsoft Word license used to create or modify the file). Although not a definitive indicator, these metadata can nonetheless contribute to a broader reconstruction of the document’s provenance and circulation history. Indeed, authorial cannot always be taken as a reliable or automatically generated reflection of authorship; rather, it must be critically assessed in light of both the file’s content and its broader production context.

VI. TEXTUAL DIMENSIONS

A. Titles and denominations

In Microsoft Word documents, the metadata field labeled as `dc:title` by Apache Tika or Title by ExifTool refers to the title property assigned within the document. This value may reflect the official title of the text or a provisional one, and – similarly to the `dc:creator` field – it can be manually edited or assigned by the user. In some cases, particularly in older .doc files, the title metadata may coincide with a working title or even with the document’s first written line or paragraph. Notably, this metadata is often not aligned with the actual file name. When the discrepancy between the `dc:title` and the file name is significant, it may indicate that the file was reused as a template for drafting multiple texts, a pattern observed, for instance, in the Avallone archive.

However, while in Avallone’s case this information is not particularly useful for reconstructing a pre-textual phase, in the Paolo Di Paolo collection, the `dc:title` values are, in most cases, meaningfully related to the document’s actual content – offering a more reliable insight into the development of the text.



B. Revisions

Among the metadata automatically extracted by tools such as Apache Tika and ExifTool is the number of revisions or saves a document has undergone. However, this value gains interpretive significance only when examined with other metadata fields—particularly the creation and last modification dates, as well as the total editing time. It is through this combined analysis that a more comprehensive understanding of a document's editing history can be formed.

Federico Milone, in his work on the digital archive of Laura Pugno – also preserved at the Centro Manoscritti – has explored this dimension in depth. His analysis underlines a broader trend among contemporary authors: rather than repeatedly editing a single file, many prefer to create separate versions at various stages of the writing process. Laura Pugno exemplifies this approach, consistently opting for new files over incremental revisions within a single document [7]. That said, in the case of Avallone and Di Paolo archives, the revision counts did not yield particularly meaningful insights into their compositional practices.

C. Duplications

A relevant aspect of the textual dimension is the occurrence of duplicate documents within the digital archive. In this respect, Paolo Di Paolo's archive includes numerous instances of identical files saved or copied across different directories, whereas Silvia Avallone's archive exhibits no such redundancies. The presence or absence of duplicate files can offer valuable insight into the extent of curatorial intervention before donation. A lack of duplication often points to a deliberate and selective preparation of the archive, while the recurrence of duplicates may indicate a more unfiltered archive. Moreover, such patterns provide further clues about the author's documentary practices and *modus operandi* in the creation and management of their archive.

To assist in the identification of redundant files, the DPW software relies on file hashes analysis rather than on specific metadata fields. In digital environments, hash functions serve as unique fingerprints for each file, making them a fundamental tool not only for ensuring the

authenticity of documents but also for detecting exact content duplicates.

However, file hashes alone are not sufficient for identifying related or near-duplicate texts, which may be dispersed across different directories or differ slightly in content. In these cases, a meticulous comparison of metadata can provide valuable support. Additionally, DPW incorporates various similarity index algorithms, which analyze textual resemblance and generate a ranked list of comparable documents. These tools aid archivists and researchers in reconstructing the documentary history of a given corpus, offering a deeper understanding of its structure and evolution [5].

VII. CONCLUSIONS

The analysis of technical metadata automatically extracted from born-digital literary archives offers valuable insights into creative practices, authorship, and the management of digital documents. Through the case studies of Silvia Avallone and Paolo Di Paolo, this study has demonstrated how metadata can substantially enrich both archival description and philological inquiry. In particular, technical metadata contributes to reconstructing the compositional history of texts and tracing the author's writing and archiving behaviors over time.

At the same time, the investigation has brought to light the complexities inherent in the interpretation of metadata. Discrepancies between system-generated and extracted metadata, as well as ambiguities in several fields, highlight the need for a critical and contextual approach. Metadata cannot be taken at face value; it must be scrutinized, especially when dealing with legacy formats. Tools such as Apache Tika and ExifTool are indispensable in revealing the hidden technical layers of digital documents, yet their results require thorough comparison and verification as well.

Ultimately, the examples discussed in this study point to the necessity of moving beyond traditional methodologies. A truly effective approach to the understanding and description of born-digital archives must be interdisciplinary, integrating archival science, philology, and, crucially, computational expertise. These combined perspectives are essential for navigating the complex terrain of digital documentality and for responsibly



preserving and interpreting contemporary literary heritage. More broadly, this study highlights the continuous processes, challenges, and learning experiences that practitioners encounter in their efforts to manage and safeguard born-digital materials.

AUTHOR BIOGRAPHIES

Adele Gorini is an archivist and PhD student at the University of Bologna, specializing in personal digital archiving. She collaborates with the Centro Manoscritti at the University of Pavia on born-digital writers' archives. Her research expands traditional archival expertise from paper-based collections to the preservation and management of digital materials.

Lucia Giagnolini holds an MA in Digital Humanities and a diploma in Archival Studies. She is an archivist at the University of Bologna's Historical Archive and a PhD student researching methods for the representation of born-digital personal archives. Since March 2023, she has collaborated on the Pavia Archivi Digitali (PAD) project at Centro Manoscritti at the University of Pavia.

ACKNOWLEDGMENTS

This study was conducted as part of the doctoral research project of Lucia Giagnolini and Adele Gorini. Gorini's project is funded by the European Union – Next Generation EU, Investment I.4.1 PNRR Scholarships for Cultural Heritage, Ministerial Decree No. 118/2023. Giagnolini's project is funded by the European Union – Next Generation EU, Investment I.4.1 PNRR Scholarships for Cultural Heritage, Ministerial Decree No. 351 of April 9, 2022.

The authors acknowledge the Centro Manoscritti for their significant investment in the Pavia Archivi Digitali project, specifically recognizing the contributions of Giuseppe Antonelli, Chiara Andreatta, and Sara Pizzi. Particular recognition is given to Primo Baldini for his efforts in developing the DPW application and for his fundamental availability and support.

AUTHOR CONTRIBUTIONS

All authors contributed to the conception and design of the study. Material preparation and data collection were carried out by A. Gorini, while the analysis was performed jointly by A. Gorini and L. Giagnolini. Chapters I, III, and VII were co-written by both authors; Chapter II and IV were written by L. Giagnolini, whereas Chapters V and VI were prepared by A. Gorini. All authors read and approved the final manuscript.

REFERENCES

- [1] Apache Tika, "Apache Tika – a content analysis toolkit," [Online]. Available: <https://tika.apache.org/>. [Accessed: Apr. 14, 2025].
- [2] V. Fernando, "Cyber Forensics Tools: A Review on Mechanism and Emerging Challenges," *2021 11th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, Paris, France, 2021, pp. 1-7, [Online]. doi: 10.1109/NTMS49979.2021.9432641.
- [3] E. Carbé, *Digitale d'autore: macchina, archivi, letteratura*. Firenze, Siena: Firenze University Press; USiena Press, 2023. doi: 10.36253/979-12-215-0023-3.
- [4] "ExifTool," exiftool.org, [Online]. Available: <https://exiftool.org/>. [Accessed: Apr. 14, 2025].
- [5] L. Giagnolini and P. Baldini, "Literary Bytes: New Approaches to Born-Digital Archives at Pavia Archivi Digitali," in *The Intangible Papers. Authorial philology and born-digital texts*, G. Antonelli, L. Giagnolini, and F. Milone, Eds. Bologna: il Mulino, 2025, pp. 57–78. doi: 10.1401/9788815414328/c3
- [6] B. J. Ismaylovna, "Methods of Extracting and Analyzing Metadata for Evidentiary Purposes," *Uzbek Journal of Law and Digital Policy*, 2024. [Online]. Available: <https://api.semanticscholar.org/CorpusID:271716331>
- [7] F. Milone, "Philological Investigation in the Digital Era: The Case of Laura Pugno's Sirene," in *The Intangible Papers. Authorial Philology and Born-Digital Texts*, G. Antonelli, L. Giagnolini, and F. Milone, Eds. Bologna: il Mulino, 2025, pp. 169–190. doi: 10.1401/9788815414328/c8
- [8] Microsoft Corporation, "[MS-NRBF]: .NET Remoting: Binary Format Data Structure - DateTime Structure," Microsoft Open Specifications Documentation, Microsoft Learn, [Online]. Available: https://learn.microsoft.com/en-us/openspecs/windows_protocols/ms-nrbf/f05212bd-57f4-4c4b-9d98-b84c7c658054. [Accessed: Aug. 25, 2025].
- [9] J. Oh, S. Lee, and H. Hwang, "Forensic Recovery of File System Metadata for Digital Forensic Investigation," *IEEE Access*, vol. 10, pp. 111591–111606, 2022.
- [10] D. Van Hulle, *Modern Manuscripts: The Extended Mind and Creative Undoing from Darwin to Beckett and Beyond*, London: Bloomsbury Academic, 2014, p. 14.

