Invented Archives: A theoretical Matter for Historical Research or Archival Science

Salvatore Spina*

Department of Humanities (DISUM) of the University of Catania, Italy.

Corresponding Author: Salvatore Spina, Department of Humanities (DISUM) of the University of Catania, Italy.

Abstract

This paper investigates the concept of "invented archives" and its significance for the Archival Science and Historical field in the context of the digital age. These innovative archives, which are frequently thematic and unconventional, arise as a reaction to the digitization movement and the demand for accessible information, thereby challenging traditional archival principles that emphasize the connections among record creators, accumulation processes, and the preservation of documents. Additionally, the paper examines the theoretical foundations of "invented archives," highlighting how these collections of documents are influenced by the methodologies of histoire sérielle that emerged in the 1950s, where the focus on seriality and quantitative historical analysis played a crucial role in shaping a new approach to Historiography. As digital technologies have advanced, these methodologies have transformed, resulting in documentary assemblages prioritizing adaptability and data reconfiguration rather than adhering to conventional archival structures.

1. Introduction

Digitality represents the contemporary dimension of life to which persons belong. It is the metaconcept upon which our modern society and world are constructed, and it emerges from humans' aspirations to boost the interconnectedness of language [1]. Computer Science, often perceived as a branch of Logic and Mathematics, serves as a communication system that humans have created to instruct computers in performing computational tasks or to initiate automated and mechanical operations. The foundation of the digital world, particularly within the Niche Ecological Digital (DEN) can be described as a «cultural state, a mindset, a met concept that encompasses all others, a framework, and a lens through which we perceive existence, simplifying complexity and nuance into basic units of information» [1]. Consequently, this leads to a singular and universal notion of "language." Ironically, it took several decades and advancements in Artificial Intelligence technologies to redirect focus towards 'words' and to conceptualize them as "numbers." This shift allows us to explore the vast opportunities that arise from applying our computational innovations in fields of Knowledge where mathematical computation has not traditionally been regarded as fundamental.

The recognition of our role as architects of a novel communicative heimat, as noted by Ciofalo and Leonzi, has only recently begun to mature [2]. Initially, the digitization within Archival Science did not coincide with a comprehensive grasp of the fundamental processes and languages associated with Computer technology. The inception of the Web-focused primarily on facilitating "rapid access" and widespread distribution of documents and findings rather than emphasizing "data processability." Consequently, the available resources lacked the structure and medium to allow researchers to perform analyses utilizing computational methods. The World Wide Web constitutes a network of interconnected documents facilitated by hypertext "links" and a vast array of URLs. This concept is well-established in both our everyday experiences and the significant body of literature in the field such as the works of whom suggests that no further exploration or clarification is necessary [3-6]. Nonetheless, this technology and its implications warrant attention. Nowadays, computers serve as the Sciences' "linguistic common denominator". Any discussions regarding the frameworks and initiatives aimed at developing new paradigms and conceptualizations must incorporate the principles of Computer Science, emphasizing the formalization of information and the sequential logic inherent in the instructions provided (prompts). It is essential to recognize that without this integration, any

Volume - 3 Issue - 2

Citation: Spina, S. (2026). Invented Archives: A theoretical Matter for Historical Research or Archival Science. *Arch Hum Soc Sci Res, 3*(2), 1-16.

Page 1 of 16

innovations produced merely represent a hybridization that undermines established statutes and principles, ultimately failing to yield significant outcomes.

1.1. Computers are Designed to Process Information

Moreover, the concept of "interconnectedness" plays a crucial role in shaping the decisions made by computer scientists, programmers, researchers, and humanists when content is published online. The Internet functions as an intricate and multifaceted system. Unlike traditional print media, where readers follow a linear progression from one section to another from the "introduction" to the "conclusion" and accompanying indexes the structure of cybertexts, as described by, is "mechanical [7]." This organization immerses web-users in diverse and stimulating neural experiences, even when these stimuli originate from a single document. The essential concept underlying the architecture of websites, the documents they house, and the various textual structures that interconnect their components internally and externally is articulated here. Everything within the network is interconnected; thus, everything must remain "associated" with one another. The challenge resides in the underlying logic of the Web infrastructure, which necessitates a reconfiguration of documents that aligns with the principles of Computer Science. In essence, Sciences that advocate for their digital transformation must reorganize their definitions, concepts, and paradigms based on a computational "language" that considers their existing frameworks and a new interplay dimension [8].

To elucidate, the concept of "document" needs to be reconfigured in a "logical" manner. The conventional definition of "document" employed by the General Diplomat describes it as «a documented account of a legally relevant fact, created per specific formalities designed to confer credibility and evidential weight». However, archivists have applied this term in a broader context, encompassing instances that may not strictly align with formal documentation [9]. The assertion regarding "analogue" presents significant challenges when altering perspectives within the digital world; these critical issues require a solution. Employing terms that reflect the document's purpose, such as "testimony" or "proof," diverts the analysis from the core subject the "document" itself thus hindering a comprehensive understanding of its attributes in the Aristotelian framework. A More Suitable Definition of "Document" Would Be «a sequence of propositions in written and/or unwritten form, which establishes the prerogatives of an entity (natural or legal person) on a movable and/or immovable property, on a condition and/or status, on the will or nolòntas, on the rights and duties per the provisions of the law in force at the time of its creation». This definition can undergo a computational procedure that generates the logical outcomes foundational to Computer Science. It is a definition that machines can process considering these as algorithmic instruments or, in recent years, as AI systems that function accurately only when the data is discrete and appropriately formalized since every term could be resembled as a "second-order abstraction," akin to a numerical field [10-20]. By understanding the fundamental

framework, precisely the concept of "propositions," any algorithm or artificial intelligence system can identify a document. In a suitable research process, these systems can gather essential data for Archival Informatics Science, including tasks like automatic classification and metadata generation, while facilitating searches that yield intricate information.

The "document" conventional definition, a prompt issued to the machine must encompass all relevant details, such as the definitions of "subject," "object," "structure," and "function," along with specific directives (for instance, "do not utilize the 'function' to characterize the 'object'"). Failing to do so could lead to unpredictable outcomes, often referred to in contemporary discussions surrounding AI and Large Language Models as "hallucinations" [21]. This scenario is akin to instructing a computer with the information that "a car car" a sentence where the same term is used for the subject and the verb, creating a non-sense phrase. Consequently, the intellectual efforts dedicated to developing the conceptual framework known as "Records in Contexts" (RiC), under consideration for several years, have resulted in creating a descriptive instrument incorporating Archival Science principles and characterizing complexes grounded in specific ontologies. However, this framework lacks an informatics approach that aligns with the paradigms of Computer Science and the contemporary trend now prevalent in various fields of integrating AI technologies within digital archives. This integration aims to facilitate the utilization of the Web and its diverse content in a "semantic" manner, a focus of AI development firms like OpenAI (refer to Fig. 1). Additionally, it seeks to enhance the automation of specific tasks associated with the document archiving workflow, including creation, approval, and classification, followed by the processing of insights that can contribute to a more objective analysis and information extraction process.

An instance of this boundary is the Finnish event "Memory Campus", taking place from September 4 to 6, 2024, which centres around the topic of "The Future of Digital Archiving" and where, among other participants, Juha Henriksson's reflection highlights the emergence of the Europeana and Time Machine Organization projects, which serve as tangible instances of what we define as "invented archives" in our analysis. In both instances, we encounter initiatives (collecting projects) aimed at gathering cultural artefacts to generate digital resources that can be accessed via the Internet, promoting democratization. However, it is crucial to highlight that both initiatives are fundamentally anchored in the "European cultural identity" concept, which is the central theme guiding the meticulous documentation amassed by the two entities. This European identity is a distinctive thematic element vital to constructing curated archives. It represents the inquiry and historical challenge that unifies the digital items within a curated archive, akin to the traditional constraints of archival practice, from which the accumulation of knowledge arises.

Consequently, the issue of invented archives arises from

discussions within Archival Science, yet it lacks a clear conceptual affiliation with this discipline. While it is accurate to assert that the topic is fundamentally significant, it is even more pertinent to note that such a relationship is not typical of conventional archives, which are characterized by diverse documentation that adheres to a distinctly different rationale.

The connection between archives and artificial intelligence technologies necessitates a thorough examination beyond the brief discussion. The Expert Group for Research and Outreach Services (EGRSO) of the International Council on Archives (ICA) contributed significantly to this discourse during a September 19, 2024, debate on the potential for a sophisticated integration of AI tools within archival practices and educational frameworks. This gathering highlighted critical challenges concerning the adoption of technologies in digital archives. Additionally, it underscored the urgent need to address significant issues associated with Large Language Models and generative pre-trained transformers (GPTs), such as "hallucination" and misinformation, which could undermine the integrity of systems and platforms, especially those linked to institutional archives that rely on the provision of verified information.

The upcoming conference titled "Theory Meets Practice: Harnessing AI for Practical Implementations in Digital Archiving," scheduled for September 7, 2024, in Budapest, addresses the pressing necessity for ongoing communication between scholars in the humanities and developers of artificial intelligence. This dialogue aims to bridge the gap between these two domains and foster an interdisciplinary approach that examines archives through the lens of fundamental computer science principles and information manipulability. In September 2021, the Plenary meeting held in Rome initiated the "InterPARES Trust AI (2021-2026)" project, aimed at facilitating a dialogue among various stakeholders to establish the criteria that govern the interaction between Artificial Intelligence systems and Archival Science. This initiative explores how the foundational principles of the latter can aid in developing a "responsible" AI technology a notion that, regrettably, often reflects misunderstandings of these technologies, which lack agency. The so-called "hallucinations" experienced by Large Language Models stem not from conscious choices but from computations designed to fulfil a primary directive: to provide responses that meet the standard of "acceptability" as defined by Curry [22]. Therefore, the core issue resides in the user's ability to assess the accuracy of the information provided rather than the machine's ability to convey "truth." However, this topic requires further exploration and should be approached in alternative forums for discussion.

It is essential to recognize that the initial assumptions we establish significantly influence, logically, the potential outcomes and ultimate success of the projects that are progressively coming to fruition. In addition to national and international evaluations, aside from the various definitions and descriptions associated with the projects mentioned above, it becomes evident that the discourse surrounding "computing" remains largely overlooked, if not absent. This oversight allows for the persistence of traditional roles related to "accumulation" and unrestricted "access," which shape the design of interfaces and provoke considerations of "what the user desires." This focus represents a singular viewpoint within the broader context of digitizing archival and cultural heritage. The realm of archives encompassing physical, digital, conceptual, or participatory forms is marked by a "disconnect" from the authentic viewpoints of the Digital Ecological Niche and Digitality. This disconnect highlights the presence of multiple cognitive entities within this interactive framework. The interconnected agents involved in the person2persons2machine relationships extend beyond just human participants. Furthermore, the "semantics" underpinning this reality undergoes a transformation driven by emerging AI technologies. ChatGPT and its parent organization, OpenAI, represent the most significant advancements in this field (refer to Fig. 1).

September 12, 2024

Learning to Reason with LLMs

We are introducing OpenAI o1, a new large language model trained with reinforcement learning to perform complex reasoning. o1 thinks before it answers —it can produce a long internal chain of thought before responding to the user.

Figure 1: ChatGPT and the "Reasoning"

Numerous intricate document complexes are designed to enhance their content's dynamism an aim that should be prioritized within the RIC framework. However, these complexes often remain confined to the "digital (traditional) indexing paradigm despite intentions to achieve dynamism." This approach primarily facilitates users in locating documents within a site structured as a "monadic" dimension, ultimately creating unmanageable complexities. A pertinent example is the thematic portals of the Italian National Archival System (SAN), which offer a consultation experience reminiscent of a bygone era (analogue, close reading), where the notion of "computing" remains unattainable. We are currently observing a digitization process idea that, despite occasionally involving the transcription of photographed

texts, addresses the essential requirement not technological of "structural obligation." This condition entails the creation of basic metadata that enables web-users to locate the documents they seek. The digitization method does not hinge on whether the content is transcribed digitally, enriched with metadata, tagged, or organized for integration with linked open data. Instead, it diverges from the digitization approach that perceives the "document" in the digital realm as "cybertext," a concept articulated by Aarseth in his work, Cybertext: Perspectives on Ergodic Literature. This notion echoes Ted Nelson's early 1965 advocacy for hypertext.

Nonetheless, these viewpoints and theoretical advancements played a crucial role in shaping the research endeavours of all academic fields focused on the text. Archival Science and historical research, on the other hand, have taken a distinct trajectory, which has only recently begun to produce theoretical advancements. Additionally, the advent of web-based computer technology has surpassed three decades, a crucial factor for contemplation, as this "innovation" has not resulted in a "transformation" within these two fields. The theoretical frameworks surrounding Digital Humanities (DH) and digital archiving distinct from "Archival Informatics Science," which addresses document management as a function of a legal entity require comprehensive definitions from both statutory and methodological perspectives. Furthermore, there is a pressing need for genuine encouragement to perceive Digitality as a transformative force that fosters collaboration in research, encompassing both historical and archival studies, through the utilization of computational tools and artificial intelligence. This necessity is underscored by the challenges in articulating the notion of the "invented archive" and its relevance within the paradigms of these two fields.

1.3. Archiving, Archives and the Web

The Web as Nelson stated «accept large and growing bodies of text and commentary, listed in such complex forms as the user might stipulate. No hierarchical file relations were to be built in; the system would hold any shape imposed on it. It would file texts in any form and arrangement desired […]. Besides the file entries, it would hold commentaries and explanations connected with them» [23]. This assertion positions Computer Science as a valuable resource for disciplines focused on studying "text," including Philology and Computational Linguistics. As previously noted, archival and historical research offers a limited and manageable role for Information and Communication Technology (ICT). However, texts and documents possess distinct significance within these two disciplines, albeit with methodological approaches that diverge from Physical and Natural Sciences. This distinction underscores our argument that, since human beings are the focal point of History and the Humanities, events cannot be systematically categorized and examined as natural occurrences, nor can they be treated as quantifiable mathematical data.

Archival Science has gained recognition in recent years regarding information and digital technologies. It is important to note that, despite being perceived as a potential benefit, the advent of internet infrastructure initiated through various roundtable discussions among archival science scholars, notably the conference held from June 17 to 19, 1985, in Turin titled "Informatica e Archivi" has not been embraced as a positive development for the academic standing of this discipline. While database systems and data networks may hold some merit, the field of Archival Science fundamentally relies on the "preservation and organization" of documentary materials produced by entities, whether physical or legal. For instance, Archival Informatics Science has not engaged with concepts from Computer Science, thereby missing numerous opportunities that could have emerged from a computable viewpoint. However, as the discipline matures, it increasingly addresses the challenges of "preservation." The conventional theoretical model of "archives," now reflected in a digital environment, embodies the significant contributions of researchers who have identified the principles of "organicity." These researchers base their conclusions on the core idea of archiving, which involves the complex relationships between the creator, the creation process, the gathering of information, its utilization, and the strategies for preservation.

The "producer" concept is the fundamental basis for the statutory archival framework, which is crucial in establishing an archive and ensuring its proper identification and utilization. Archives represent the outcomes of management and administrative processes that originate within institutional and legal frameworks, culminating when a historical event prompts the emergence of a past issue. In this final stage, the limitations yield to the central theme, prompting the establishment of a constructed archive. Here, the creator utilizes various archives and sources to gather the sources required to address a historical inquiry. This process connects the documents without adhering to strict archival limitations, aiming instead to form a comprehensive collection of historical sources. Scholars will then base their reasoning and historiographical interpretations on this assemblage. Traditional archives and their digital counterparts form the foundation of "invented archives" and contribute to their development; however, they remain distinct entities. Traditional archives precede the invented ones, potentially perceiving them as mere 'outcomes' of their existence. The digital turn has significantly impacted various dimensions of scientific frameworks, as is widely acknowledged. The fields of Physical and Natural Sciences have notably gained strength, with their methodologies and the "scientific method" achieving analytical depths that have allowed human inquiry to explore previously unimaginable territories. Meanwhile, Humanities Science strive to adapt to the "digital ethos," focusing on challenging numerous methodological and paradigmatic elements. However, this effort often lacks clarity, appearing to aim for a state of maintaining the status quo while simultaneously "advocating for change".

From this perspective, it is interesting to note that his assertions hold validity. Since the launch of the inaugural

website (http://info.cern.ch/hypertext/WWW/TheProject. html), which offered web users complimentary access to the concepts and features surrounding the "invention" of the "WWW," the content that has been "uploaded" bears minimal relation to the notion of an "archive." Furthermore, these newly established archives do not originate from the Web; instead, they regard this infrastructure merely as a subsequent phase that serves as a crucial medium for assembling the documentation curated by their creators. Consequently, digitization does not drive the rapid emergence of these invented archives. A particularly illustrative example, as noted by Gianluca Braschi, is the "Piancastelli Collection," which aims to encapsulate the identity of Romagna, a Region that lacks a traditional producing entity [24]. In this instance, the collection is designed to articulate the identity of Romagna, yet it deviates from conventional archival principles due to the absence of a "producer"; instead, a creator is involved, specifically identified as the conservator, which in this case is the State Archive of Forlì-Cesena. Since 1991, websites have continuously emerged featuring archival materials that present historical initiatives, research concepts, and individual interests. The introduction of "Mosaic Netscape" and "Netscape Navigator" in 1994 empowered users to curate their collections, transforming the Internet into a comprehensive global information network. Historical narratives proliferated, and notably, individuals eagerly began sharing their personal stories online, facilitated by establishing platforms focused on their preferred topics and themes.

Although intriguing and "intrusive," we are not dealing with mere archives; instead, we encounter what characterizes as "the outcome of the efforts of numerous individuals, driven by diverse and varied motivations [25]". "Entities" that do not align with the definitive traits of the "originator" but rather correspond to the role of the creator, who, in pursuit of upholding democratic ideals of "widespread distribution," compiles and gathers diverse documentation encountered within archives. This process involves connecting various documents and establishing them as foundational elements for research trajectories, whether they encompass a broad perspective or "support" for a specific event confined by spatial and temporal parameters. It is crucial to highlight that the aim behind creating these "collections" arises not solely from the enthusiasm of committed scholars and researchers; instead, a more appropriate designation would be "documentary complex." This project seeks to construct a thorough framework to guide and shape historical investigations within that field. Between 1995 and 2000, numerous collaborative projects emerged and continue to expand, each with distinct scientific objectives. Although these partnerships lack an archival connection, the criteria guiding document access selection are articulated, distinguishing them from others.

For example, in the case of the History of France, the network offers the possibility of consulting sites created with the aim of collecting documentation on this subject:
• Bibliothèque Nationale de France.

• Centre d'accueil et de recherche des Archives nationales.
• Archives du service historique de l'Armée.
• La Nouvelle France.
• GIS and Remote Sensing for Archeology: Burgundy, France.
• Château de Versailles.
• The National Huguenot Society; 8) French Revolutionary Pamphlets.

For Italy, users can access:
• al sito "Cronologia".
• all'Archivio Storico della Camera dei Deputati.
• alla Fondazione Istituto Gramsci.
• History of Italy Primary Documents, Harold.
o Lee Library, Brigham Young University.
• International Institute of Social History.
• Journal of Modern Italian Studies.

However, we do not observe any conventional archives within this collection. This observation prompts a consideration that leans more towards the historical aspects of Archival Science rather than its outlook on digital developments. When examined through the lens of the underlying computer infrastructure, these intricate systems pertain to "semantic relations." This particular statement, however, proves to be challenging to dissect and analyze. Dealing with "semantics" involves exploring the intricate pathways of human interaction, specifically the concepts of "meaning and interpretations". The focus of Humanistic inquiry extends beyond mere data; it encompasses both the tangible and intangible outputs of human cognition. These outputs do not conform to rigid rules or laws but are represented through symbols and signs belonging to the realm of Knowledge. As Cassirer describes, these are "denotators" that, according to Sebeok, are signs linked by convention rather than by similarity or proximity between their signifiers and what they denote this suggests a conventional framework where the denoted entity exists independently [26,27]. This premise invites a thorough exploration of the notion of "text," positing that each author who creates "documents" assigns meaning to a sign based on self-selected rules. Consequently, Computer Science needs to reconstruct this relationship to translate the intended information into computational language accurately and, if feasible, capture the author's intent or awareness that the message is designed to elicit a specific response.

In the field of Computing, drawing from the insights of Ada Lovelace and Charles Babbage, who emphasized the "discrete" nature of the processed information, the concept of "meanings" and, consequently, Semantics garnered minimal interest until the 1950s. This shift occurred in response to Alan Turing's pivotal inquiry: "Can Machines Think [28]". This inquiry has led to numerous research endeavours to decipher the mechanisms that need to be established and subjected to what was then referred to as "artificial intelligence." This exploration significantly diverges from traditional archival paradigms, which have historically grounded their conservation and utilization methodologies in approaches that have overlooked mainly the concept of

"meaning." A document lacks inherent 'meaning' as it exists within a specific "context" (such as a series or background); it occupies a particular "situation" defined as a spatiotemporal instance within the archive, where its legal reason resides and where the producers intentionally positioned it. Semantics do not exist in isolation; they may potentially arise through rigorous intellectual examination and interpretation.

Archives are structured in a hierarchical manner rather than a relational one. This structure resembles a "graph" or a "tree," which is the foundation for various archival description standards, including ISAD, ISAAR, ISDIAH, and ISDF. These standards are further developed through the "RiC" model, which is suggested as a relational alternative [29].

The Internet and newly created archives demand a greater scope; "The Internet not only provides digital representations of pre-existing archives but also introduces entirely novel archives specifically tailored for online use. [...] Newly established archives focus on gathering and providing access to documents that are dispersed across various physical archives" [30,31].
• What, then, constitutes this "other"? Furthermore, what additional inquiries persist regarding the definition of "archives"?
• If it is accurate to assert that the created archives, in some respects, assume the designation of "archive," which does not accurately reflect their true nature, what kind of connection exists between the Web and these "heritages"?

Beginning in 2006, which marked a decade after the launch of the first website, archivists were prompted to engage in profound contemplation centred around the "external perspective" concerning archival practices [32]. This shift in focus moved away from the internal significance of archives for archivists, redirecting attention to the value and relevance of archives for a global audience that would increasingly seek access through websites and digital repositories. The digitization process involves generating copies that are easily accessible and completely searchable. It is essential to distinguish this from the notion of "dematerialization," which operates under a different framework and carries distinct implications, particularly in Ontology and Philosophy. This distinction expands the scope of the discussion by introducing concepts that may lead to complex theoretical challenges. In the context of Archival Science, "digitization" refers to the production of copies and "reproductions" that possess a specific "original" materiality, existing within the realms of servers, bits, bytes, networks, and the Web. Ultimately, these copies primarily function as 'service copies. The digital format may enhance the physical archive by offering a form of 'increased protection' and undoubtedly benefits online users. However, it fails to support the establishment of the archive as a fundamental concept—this is even more true for archivists, who act as the sole intermediaries between external users and the materials stored within the archive.

The Internet allows users to browse freely without an expert's guidance, resulting in a less attuned portal to user needs. This situation arises because online consultations often lead to digital inventories that mirror the structure of traditional paper archives. Consequently, these online inventories frequently hinder precise identification and fail to streamline the data and associated datasets contained within documents. As a result, users lack an analytical framework for their inquiries. The term "invented" transcends the cognitive framework of both the archivist and the archives, which do not focus on establishing a chronological order but rather on situating the document within a clearly defined context. This process generates a meaningful relationship among the records, facilitating the reconstruction of the archival connection. As attention shifts to the digital world, the resulting online platforms function as official "catalogues" for archives, embodying both the curators and the creators. These websites provide internet users with access to digital collections in their existing formats and/or databases that strive to replicate the hierarchical organization of the collections, often to enhance analytical investigation through comprehensive metadata. Consequently, these platforms do not represent original works; instead, they reflect the promotional intentions of the curators and producers—suggesting that if something is available online, it must be real.

The notion of "invented archives" does not imply the significance of digitization initiatives, even when archival materials or portions of them are made available online this means that the resulting searchable formats lack meaningful connections among the data, except for the relationships between individual records, their context, and the association with the originating entity. This entity remains the sole authority capable of reliably validating the processes that led to the creation of each specific document. Consequently, there is a disconnection between Archival Science and the concept of "invented archives", which continues to be a topic characterized by significant dynamics and critical issues concerning the preservation of the Web, including its various sites and documents.

1.4. Internet Archive: Archive? Library? Invented?
On August 6, 1991, Tim Berners-Lee, a British physicist, introduced the first website, marking the inception of the World Wide Web, which he conceptualized and developed. Berners-Lee envisioned and successfully developed an information network that enables users to browse interconnected content via hyperlinks. This content, which includes website pages, images, videos, and various official documents from institutional websites (such as those belonging to governments and ministries), possesses a distinct address known as the URL (Uniform Resource Locator). Users can utilize this address to access and retrieve information. We are encountering an unprecedented situation. All users connected to the network can browse any content available online, provided they utilize search engines. However, this accessibility raises significant concerns regarding the ultimate "aim" of such information, the potential obsolescence of formats, and the transformative

Citation: Spina, S. (2026). Invented Archives: A theoretical Matter for Historical Research or Archival Science. *Arch Hum Soc Sci Res, 3*(2), 1-16.

Page 6 of 16

impact of technology across various knowledge domains. This situation prompts critical inquiries for humanists. Are archivists and historians implicated in this issue: should the continuous stream of documents be integrated into an archive? What form should such an archive take? In what ways will historians and archivists engage with this content.

While numerous inquiries have received answers over time, it remains evident that three decades is insufficient for a millennia-old discipline like Archival to establish "definitive" conceptual frameworks regarding the challenges of digitization and the management/storage of online documents. For instance, what implications arise for communications from political organizations utilizing instant messaging platforms—as Durrant and Allegrezza suggest [57,58]? «The Web is having a dramatic impact on how we research and understand the recent past. Historians, who have long laboured under conditions of source scarcity we wish we had more information about the past, but it was not recorded or preserved are now confronted with primary sources on a scale that defies both conventional methodologies and standard computational methods». «For the most part, historians cannot write histories of the 1990s unless they use web archives: with them, military historians will have access to the voices of rank-and-file soldiers on discussion boards; political historians, to blogs, the cut and thrust of websites, electoral commentary and beyond; and course, social and cultural historians, to the voices of the people on a scale never before possible» [33].

Since its inception, the Web has recognized the importance of establishing a repository for preservation. All forms of information, originating from the dawn of humanity, inevitably transition into History as time progresses. Humans possess an intrinsic drive to archive all they create, encompassing written works, ideas, and spoken expressions, regardless of the medium employed. The Internet serves as a novel medium; however, it can mislead users when confronted with the term "archive," commonly employed by website creators aiming to compile a collection of information and/or documents. Digital archives, which gather data regarding events and individuals, do not equate to traditional archives. Furthermore, institutional archive websites operate under a different framework, as previously noted. In the context of the Web, when we encounter the Internet Archive (https://archive.org), we categorize it as an archive. Its unique characteristics align with the principles that characterize a conservative institution. Brewster Kahle established the Internet Archive in 1996, initiating a project

that involved the development of a digital archival system designed to safeguard websites created and published online. Many of these sites lack ongoing financial support, leading to their potential removal from the Internet.

The collecting entity consists of a non-profit organization formed by a diverse array of governmental and non-governmental institutions, which has established a digital repository that aggregates "millions of books, films, software, music, websites, and more, all available at no cost," located in San Francisco. The framework operates on an "entry point" model that enables internet users to find specific content and trace the historical development of a website, including updates to its interface and its initial graphic and functional design. Additionally, users can browse and download previously available documents on the site, but no longer accessible offline. Through its "Wayback Machine", the Internet Archive facilitates the retrieval of PDFs and other files from its collection, allowing for remote storage and download. Since its inception, this platform has been regarded as a national repository for the United States; however, the advent of the Internet and the World Wide Web enables it to transcend geographical boundaries, facilitating the preservation of all online content at the moment the "crawling" process is initiated. Can we identify a conservative subject in a specific context? It is essential to clarify who generates the archive—or the various archives that will be treated as collections once they are incorporated into the Wayback Machine or, more appropriately, as catalogues.

The crawling mechanism also incorporates the websites of various institutional entities, such as municipalities, regional authorities, and ministries, into the Internet Archive. When this process successfully captures a site's pages, the "time machine" reconstructs its architecture and internal hyperlinks. For institutional websites, pages related to the praetorian court register and public documents will probably be included in the archive. However, an examination of the platform reveals that the documentation associated with the different entries in the Register and the Official Acts is not available for download. This limitation effectively turns the website pages into a comprehensive online catalogue, assisting users in locating documents that may not be included in the most recent update of the official site. For instance, the Municipality of Riposto's website, located in the province of Catania (Italy), underwent its initial archiving on April 10, 2001. The crawling process has been executed 3,583 times (see Fig. 1), enabling users to explore the complete progression of the Agency's website interface (see Fig. 2-5).
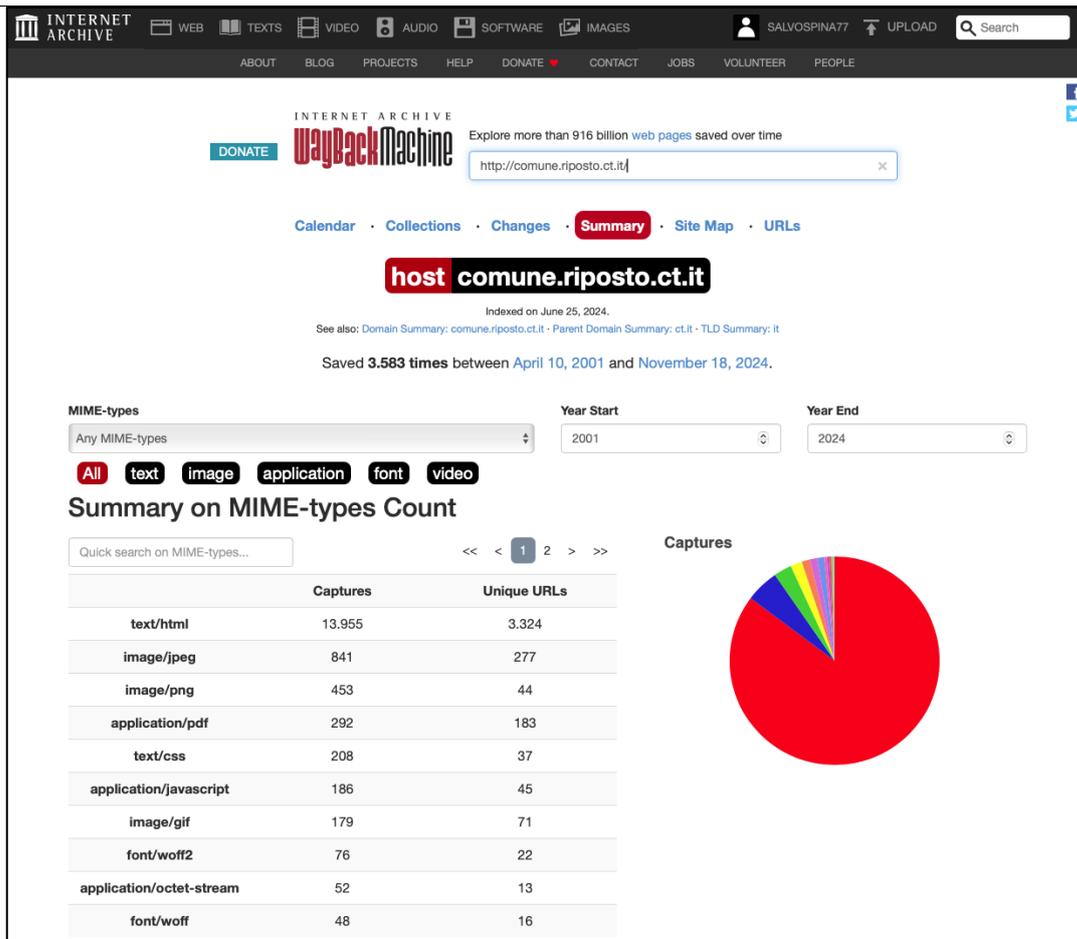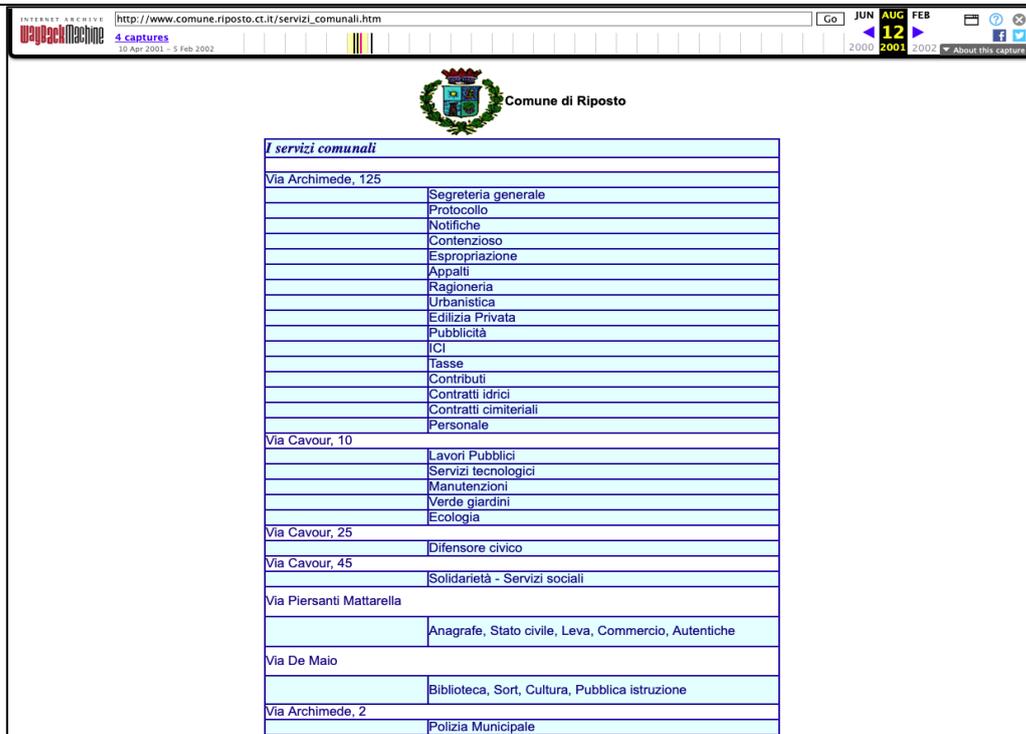
Citation: Spina, S. (2026). Invented Archives: A theoretical Matter for Historical Research or Archival Science. *Arch Hum Soc Sci Res, 3*(2), 1-16.

Page 7 of 16

Figure 2

Figure 3

Figure 4

Figure 5

Upon attempting to access the Praetorian Register's webpage, users can view a list of various acts; however, they cannot download any attachments. This limitation stems from the site's restrictions against external organizations performing crawling activities (see Figures 6 and 7). In this instance, could we envision a "Riposto Praetorian Register" repository managed by a curator, specifically the Internet Archive? The likely response is affirmative; however, we must remain open to the potential to examine this issue thoroughly. Undoubtedly, this can be interpreted as evidence that neither the Internet Archive nor the Municipality di Riposto's website represents a fabricated archive.
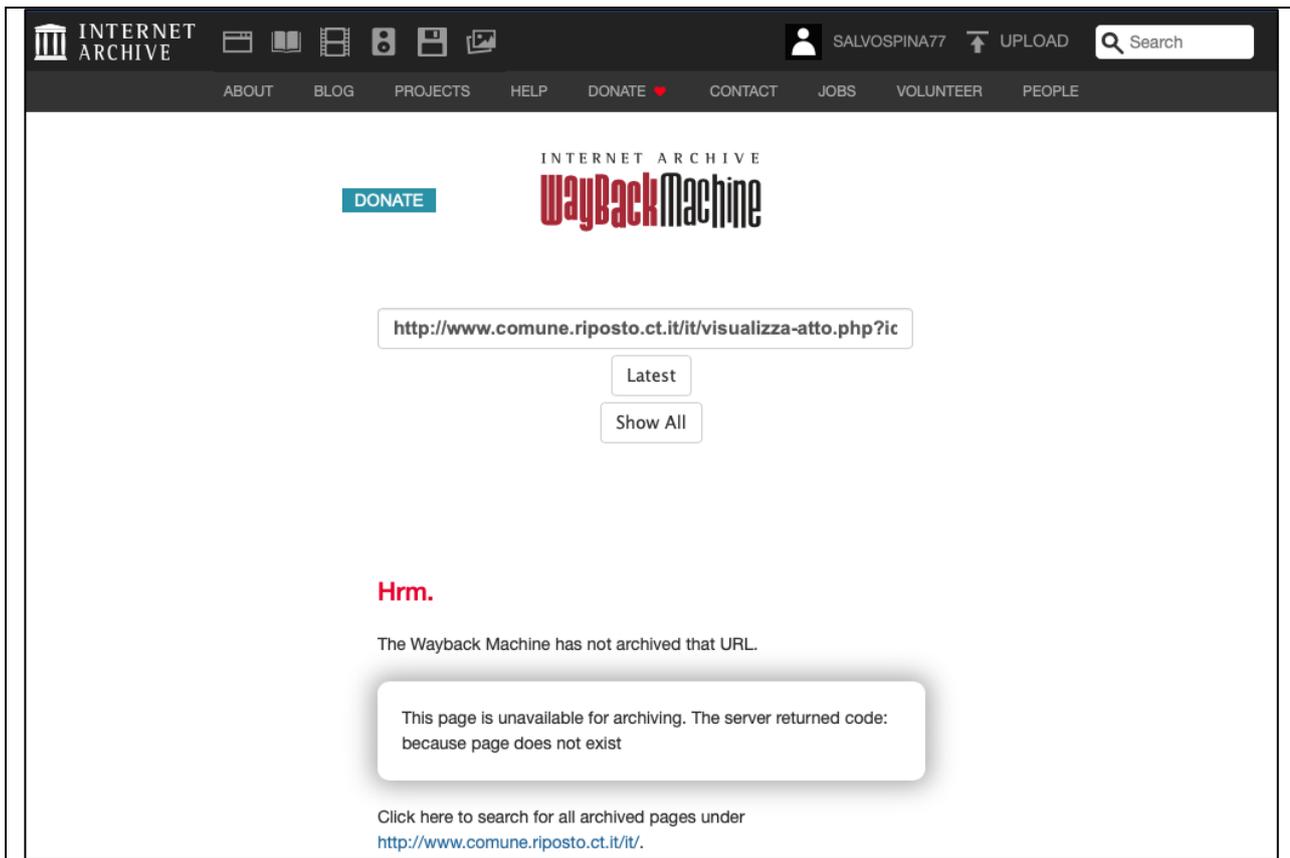

Figure 6

Figure 7

We pose additional inquiries concerning the "archive of archives," mainly focusing on the criteria for classification utilized by the organization. It is essential to recognize that classification is pivotal in establishing and administrating an archive, especially in digital production. Furthermore, we question whether the Internet Archive search engine and the Wayback Machine function as a thesaurus or merely as a comprehensive inventory. The platform lacks search functionality for its content and fails to offer comprehensive or concise bibliographic details regarding the included sites. Additionally, the archive does not reveal information about the sites it hosts, including the number and type of their domain names, nor the countries they are associated with.

The primary concern is that the archive lacks an objective system for categorizing or classifying its sites—this necessitates that users know the URL to find the "archived document" amidst the vast collection. This situation hinders rapid access and undermines the general and standardized principles that inform the notion of "placement" (such as room, shelf, folder, file, etc.). The archival theory has established clear guidelines that facilitate the identification and location of the fundamental archival unit. As noted by Stefano, "documents related to the same matter are organized into a file, which in the physical realm is stored within a shirt [that must display] the identifying information of the file," and multiple files "are typically housed in a container (folders, envelopes, bundles, box files) [34]." The digital dimension also adheres to this framework, albeit with

different terminology, as outlined by Stefano Allegrezza.

As we transition into Digitality, traditional paper documents are supplanted by digital files, essentially sequences of bits. These files require a storage medium to exist. In the digital environment, files are organized into 'folders' or other document collections within the operating file system. Consequently, the significance of unit conditioning diminishes due to the distinct 'physicality' of digital formats, leading to the obsolescence of physical shelves. The Internet Archive deviates from this reasoning. While it possesses a conservation entity, it lacks a systematic classification and inventory that would assist users in identifying specific archival units. Furthermore, the notion of "placement" in the digital realm is predominantly represented by URLs. The Wayback Machine facilitates the identification of a website referred to as its "bottom" but fails to provide access to its content. Consequently, an inventory exists, created by the conservation entity, yet the only means to explore the contents of the archives involves navigating through the pages of a preserved site. Users cannot utilize the internal search functionality of the original site, as it is no longer operational on the original server, rendering the dataset inaccessible. Thus, the sole method to locate a "drive" is knowing its precise URL. In the case already considered as an example, the website of the Municipality of Riposto is "http://comune.riposto.ct.it/", but within the Wayback Machine and the Internet Archive, the URL (i.e. the "placement") for its identification becomes "https://web.

archive.org/web/20130913175143/http://www.comune.riposto.ct.it/it/elenco-albo-pretorio.php". Soon, scholars and historians eager to document this organization's History and related events will pursue this connection. To address our inquiry regarding exploring the invented archive concept, it becomes evident that we do not encounter an "invented" notion even in web archiving. This observation suggests that, once again, it fails to constitute a compelling topic within Archival Science.

1.5. Internet and Historical Research and Invented Archives as Histoire Sérielle

While invented archives are thematic collections centred around diverse research topics and objects, assembled by individual scholars both professionals and amateurs as well as by research communities and academic institutions (Rosenzweig), it is crucial to acknowledge that this characterization, though accurate, compels us to examine the concept of "serial History" which necessitates considering the methodological framework that emphasizes the identification of specific archival documents and their organization into defined series aimed at elucidating historical phenomena. As articulated by Furet, this approach is a fundamental prerequisite for developing the Quantitative History Field [35-40]. Pierre Chaunu introduced the term "histoire sérielle" to elucidate the framework of his study, Séville et l'Atlantique, 1504-1650 [41]. Subsequent publications further refined his concept, including Histoire quantitative ou histoire sérielle and La durata, lo spazio e l'uomo nell'epoca moderna. La storia come scienza sociale invited readers to focus on the recurring elements situated within a context of uniformity rather than isolated events to construct a comprehensive set of analyzable data, employing the principles of social mathematics [42,43].

Seriality supplants Positivist Historiography's "elusive" event characteristic, establishing a consistent recurrence of chosen data. This framework allows historians to navigate within a complex structure organized into subsystems. As notes, historians can suggest or omit the internal connections of these subsystems, enabling them to address their inquiries regarding History. We stand on the threshold of Digital History, particularly concerning the contemporary notion of the "invented archive". The 1990s marked the emergence of innovative approaches within the field of History. Computer technologies transformed the discipline and prompted a reevaluation of perspectives, revealing crises and scepticism, as noted, alongside aspirations for advancement in historical scholarship [44]. While Braudel highlighted the "general crisis of human sciences", he acknowledged the importance of analyzing phenomena to develop a unique understanding of short-, medium-, and long-term temporal dimensions [45]. While it may not be immediately apparent in all these instances, the consistent presence of "data" is essential for theoretical development and Historiography that recognizes "omnipresent" elements across time. This perspective challenges the narratives of significant divisions, which must rely on examining identified elements or data that elucidate the series, quantity, and extended duration. Furet

emphasizes that "serial history [is] one of the most fruitful ways of developing historical knowledge", highlighting its value not only for its contributions but also for fostering the necessary conditions for the emergence of new concepts. Following extensive discussions, these concepts have revealed the "promising" aspects of Digitality, including datasets, databases, data organization, Big Data, and digital archives, among others.

Every historian systematically arranges their findings and evidence within their scholarly inquiry, integrating them into a historiographical work that reflects their interpretation of historical events and topics. In both an evenemential and long-term approach, the initial phase of historians' work is defined by the necessity to systematically arrange and connect the information obtained through "analogue acquisition," which involves a close reading methodology. The necessity for this development became evident, mainly influenced by the concepts presented by Vannevar Bush in his work As We May Think [46]. This realization emerged distinctly in the 1950s with the advent of computers and software designed to systematically arrange historical data into structured datasets. According to Oscar Itzcovich, historians can utilize the database as a highly adaptable and cutting-edge instrument for constructing and analyzing information complexes, moving beyond traditional close reading methodologies [47,48]. In this context, integrating mathematical methods into written records has become a prevalent approach within Humanities research. Spina states that many archival materials can be chronologically categorized and incorporated into centralized systems, databases, and digital repositories [49]. The 20th century, in contrast, revealed numerous additional opportunities. While the database has established the foundational concepts for the invented archives, it is even more accurate to assert that the Internet infrastructure and the Web have significantly enhanced their accessibility.

On December 20, 1990, CERN introduced the "WWW" initiative to its staff. Subsequently, on August 6, 1991, the inaugural website emerged on the network. By April 30, 1993, CERN opted to release the World Wide Web as open source, placing it in the public domain. This decision significantly transformed how research findings, including those in the Humanities and History, are produced and communicated. In the spring of 1993, the Centre for Research, Development and Higher Education in Sardinia (CRS4) launched Italy's inaugural website, which remains operational at https://www.crs4.it. The Internet has undergone significant transformations and continues to evolve, particularly in areas where we struggle to adopt a computational mindset. This evolution profoundly impacts researchers engaged in this pivotal era, making it impossible to remain detached from this overarching medium that permeates daily life. The increasing opportunities for interaction not only reshape the methodologies and timelines of historical research but also profoundly affect the practices of disseminating and sharing knowledge, as well as the pedagogy of History and its communication [50]. Research on the Past has taken a

completely different methodological meaning, which is not easy to define: "There are no products that are already good and ready, suitable for use by historians, nor will they ever exist" [44].

The notion of an invented archive becomes relevant and broadens in this context: the network can encompass everything, facilitating communication and, crucially, enhancement of information. It represents an open environment that allows for the sharing of an astonishing volume of resources, presenting scholars with a wealth of previously unimaginable materials, which often remains beyond control [51-56]. The core of our analysis lies in the observation that while Archivistica is beginning to focus on these matters with increased scrutiny, it appears to lack a distinct conceptual framework for the term "invented archive." Furthermore, it is essential to note that historical research plays a crucial role in underpinning its establishment and development. Historical websites were everywhere, dealing with all aspects of historical matters. Researchers and academics at different tiers actively upload extensive collections of documents online, drawing the interest of fellow historians and scholars who require access to these valuable resources. In their 2000 publication, *The History Highway 2000,* Dennis Trinkle and Scott Merriman articulate the aspiration to foster a critical mass in historical research. Their work emphasizes that engaging with History can evolve into a more dynamic and participatory endeavour, aligning with Public History's foundational principle. Is the resurgence of the serialized narrative a possibility? Likely. Like the offline databases that scholars initially utilized, invented archives necessitate a systematic approach to dating and adding metadata to the documents they contain (records). Consequently, it is insufficient to merely select a historical topic and seek out archival materials (both public and private) to create an "invented archive." While the theme serves as the guiding principle, the primary purpose of the database or digital archive is to establish a network of interconnected data. This interconnectedness allows researchers to gain deeper analytical insights and to identify patterns that can elucidate a specific event or a series of events that collectively illustrate the foundational theme of the documentary collection.

The emergence of digital archives and their rapid expansion, along with the ongoing innovation of platforms and tools for their development, such as Omeka-S (https://omeka.org/s/) and Arches (https://www.archesproject.org), captivates researchers associated with qualitative methodology. As noted by Furet, these scholars perceive greater efficacy in serial data and quantification, mainly when the availability of archives and documentation, coupled with the lack of frequent disruptions, as seen in Modern History, enhances their analytical framework. Consequently, the invented archives inadvertently reflect the methodology of "seriality," effectively capturing the challenges faced by historians who seek to quantify transformations, trace the origins of phenomena, and analyze their geographical distribution. A pertinent example is the database on slavery, particularly

Slave Voyages (https://www.slavevoyages.org), which illustrates the intricate and widespread nature of this issue. Historians require a systematic approach to organizing their documents to pinpoint those that can serve as sources for analyzing and elucidating their historical inquiries. In a digital context, this necessity intensifies, leading to a more intricate process of data dating that ultimately results in the development of a dynamic document, referred to as a cybertext. In the context of the invented archives, the documentation should cater to the requirements of historians rather than prioritize "accessibility" within the archival framework. The significance lies not only in the physical placement of the documents but also in their interconnections with other sources in the database, even though the IRC considers this viewpoint.

The proposed archive would accommodate extensive and increasing amounts of text and commentary, organized in intricate formats as specified by the user. It should not establish hierarchical relationships among files; the system must remain free from any imposed structure. It would allow storing texts in any preferred format and configuration, effectively merging the functionalities of a card catalogue, a flip book, and similar tools. The archive would categorize entries without limit and incorporate the trails outlined by Bush. Beyond the primary entries, the system would include associated comments and explanations. Furthermore, it would offer various mechanisms for modification, enabling users to alter both the content and organization of their files. Structures would exist for revising and rephrasing text, and users would be able to adjust the arrangement of the components within the dossier, including modifications to sequence, labelling, indexing, and commentary [23]. The invented archive must align with the evolving nature of historical inquiry, which adjusts its sources according to the specific objectives of the research. The seriality concept compels historians to explore sources' diverse interpretative and analytical potential continually. In this context, the framework of the invented archive—similar to what Nelson refers to as the Evolutionary List File (ELF)" can be utilized and combined in numerous ways" [23].

While organized to serve as a unifying element of the theme that inspired the creation of the archive, they exhibit flexibility that transforms the "fact" into an interpretation grounded in the relationships among them. "The document, the data, does not exist for [themselves] but concerning the series that precedes and follows them; what becomes objective is their relative value and not their relationship with an elusive "real" substance. Thus, the old problem of "criticism" of the historical document" is also being shifted. The invented archive, therefore, guarantees what Furet defines as an "external" critique, which "is no longer based on credibility based on the comparison of the document with contemporary texts of different nature, but on consistency with a text of the same nature that has a different position in the same time series, that is to say, preceding or following it. [On the contrary, the "internal" criticism is all the more simplified as many operations of "scraping" data can be

Citation: Spina, S. (2026). Invented Archives: A theoretical Matter for Historical Research or Archival Science. *Arch Hum Soc Sci Res, 3*(2), 1-16.

Page 13 of 16

inserted into the memory of the electronic computer". It is in the statements of Furet that we can identify a general, decisive and first definition of "invented archive". «Questo incontro di una rivoluzione metodologica e di una rivoluzione tecnica, che non sono d'altra parte estranee l'una all'altra, permette di prendere in considerazione la costituzione di archivi nuovi, conservati su schede perforate, che rimandano non soltanto a un nuovo sistema di classificazione, ma soprattutto a una critica documentaria diversa da quella del secolo XIX. [...] La coerenza [tra i dati] è assicurata [...] da un minimo di formalizzazione del documento, in modo che si possano ritrovare, per un lungo periodo di tempo e per ogni unità di tempo, gli stessi dati, nella stessa successione logica. Da questo punto di vista, l'utilizzazione del calcolatore elettronico per lo storico non costituisce solo un immenso progresso pratico, per il risparmio di tempo, [ma] è anche una coazione teorica molto utile, nella misura in cui la formalizzazione di una serie documentaria destinata a essere programmata obbliga fin dall'inizio lo storico a rinunciare alla sua ingenuità epistemologica, a costruire il suo oggetto di ricerca, a riflettere sulle sue ipotesi, e a passare dall'implicito all'esplicito» (The revolution in methodology and technology enables the creation of new types of archives, such as those utilizing punched cards. These archives introduce a novel classification framework and, more importantly, a form of documentary criticism that diverges from 19th-century practices. Within these archives, the integrity of the data is maintained through the formalization of documents, allowing for the retrieval of identical data in a consistent logical order. Consequently, employing computers for historical analysis represents not only a significant advancement in efficiency but also imposes a valuable theoretical framework. This formalization process compels historians to relinquish their epistemological creativity, systematically construct their research focus, critically evaluate their hypotheses, and transition from implicit assumptions to explicit formulations) ».

2. Conclusion

The notion of "invented archives" presents both a challenge and an opportunity for research in archiving and Contemporary History, though the impact may not be equally distributed. On one hand, the digitization process and the rise of invented archives within the digital landscape prompt critical inquiries regarding conventional archives' essence, role, and significance. These invented archives often fail to adhere to established archival principles grounded in the connections among the document's creator, accumulation, utilization, and preservation. The motivations behind their creation vary widely, encompassing aspirations for increased accessibility and the democratization of knowledge and efforts to address gaps in documentation by consolidating scattered sources. While these diverse objectives enhance access to historical data, they may undermine the archival relationships and internal consistency that define traditional archives. Consequently, an "invented archive" may prioritize the creator's intentions over the actual documentary context of the original producer, thereby introducing an epistemological challenge for the field of archiving. Conversely, constructed

archives present significant opportunities and avenues for historical inquiry. Their thematic focus and non-linear organization enable reevaluating how historical information is structured and analyzed, moving beyond conventional archival practices. The digital framework of the Internet supports the collection and examination of documents from various sources, effectively addressing the physical and logistical constraints associated with traditional archives. Furthermore, the capacity to connect documents via hyperlinks and to manipulate data digitally fosters a dynamic research environment where sources can be continually reorganized and reinterpreted. This fluid methodology resembles the approaches utilized in the histoire sérielle of the 1950s, which emphasized quantitative analysis and the serial nature of data to elucidate historical events. Indeed, constructed archives can be regarded as a contemporary advancement of that methodology, merging seriality with the adaptability of digital frameworks, thereby enabling a broader spectrum of interpretations and historical analyses.

Invented archives expand the avenues available for historical inquiry and introduce considerable challenges. The absence of standardized protocols for classifying, describing, and preserving documents in digital formats can jeopardize the credibility of sources and the overall quality of research. Furthermore, the reliance on search engines for accessing digital materials and the dependence on specific URLs for data retrieval underscores a fundamental vulnerability in their preservation. The lack of a systematic inventory and a cohesive classification framework diminishes the capacity of invented archives to function as trustworthy repositories of historical information. This situation prompts additional concerns regarding their viability over time and the reliability of the data they contain. In Web archiving and its inherent challenges, an essential consideration arises: the necessity for a hybrid methodology that merges the advantages of digitization with established archival practices. The development of digital archives must extend beyond simply converting paper documents into digital formats; it should involve reorganizing data that honours both the principles of traditional archival systems and the evolving demands for access and interaction within digital contexts. While platforms like the Internet Archive strive to safeguard digital heritage, they exhibit significant shortcomings regarding classification, accessibility, and sustainable preservation. The lack of a definitive producer and the absence of a coherent structure for archival placement underscores the challenges in recognizing such initiatives as authentic archives in the conventional sense.

In conclusion, although the emergence of invented archives addresses the increasing demand for immediate and unrestricted access to historical materials, it is equally important to thoroughly and critically examine archives' function in the digital era. Archivists face a dual challenge: first, they must acknowledge and adapt to the technological and methodological shifts introduced by digitization and the establishment of invented archives; second, they must uphold and reinforce their core principles that ensure the

integrity, contextualization, and continuity of documentary evidence. Collaboration among archivists, historians, and computer scientists is crucial to devise methodologies that reconcile these seemingly opposing domains, safeguarding historical records' reliability and fostering meticulous and innovative research. In this context, the "archive" concept, which has historically posed difficulties for scholars in archival Science, requires a revision to accommodate the contemporary landscape of digital documentation. As articulated by Gianni Penzo Doria [55], this revision must consider the infrastructure of "locations" such as servers and media for connectivity and storage, the influence of companies that offer 'spaces' and various cloud services—acknowledging the potential risk of "vendor lock-in," which creates a dependency on the service provider. Additionally, it must address the contributions of web users who inadvertently leave "data" on servers and websites during their interactions. This updated definition should also reflect the complexities of these digital documents from a Computer Science perspective while remaining grounded in the theoretical and methodological principles of the humanities.

Compliance with Ethical Standards

References

1. Spina, S. (2024). Digitality as a "longue durée" historical phenomenon. arXiv preprint arXiv:2403.03869.
2. Ciofalo, G., & Leonzi, S. (2013). Homo communicans: una specie di/in evoluzione.
3. Xu, Z., & Zhang, J. (2021). Computational Thinking: A Perspective on Computer Science. Springer.
4. Gonzalez, T., Diaz-Herrera, J., & Tucker, A. (Eds.). (2014). Computing handbook: computer science and software engineering (Vol. 1). CRC Press.
5. Reilly, E. D. (1993). Encyclopedia of computer science. Van Nostrand Reinhold Company.
6. Clerici, A., & De Pra, M. (2012). Informatica e web. EGEA spa.
7. Aarseth, E. J. (1997). Cybertext: Perspectives on ergodic literature. JHU Press.
8. Spina, S. (2023). Homo-Loggatus. The anthropological condition of historians in the digital world. arXiv preprint arXiv:2307.02590.
9. Guercio, M. R. (2019). Archivistica informatica. I documenti in ambiente digitale. Roma: Carocci.
10. Frege, G. (1980). The foundations of arithmetic: A logico-mathematical enquiry into the concept of number. John Wiley & Sons.
11. Frege, G., Penco, C., & Picardi, E. (2005). Senso, funzione e concetto: scritti filosofici: 1891-1897. Laterza.
12. Frege, G. (2021). L'alfabeto del pensiero. LIT EDIZIONI.
13. Bianchi, C. (2003). La filosofia di Gottlob Frege (Vol. 75). FrancoAngeli.
14. Roselló, J. (2019). Hilbert, Göttingen and the development of modern mathematics. Cambridge Scholars Publishing.
15. Hilbert, D., & Cohn-Vossen, S. (2021). Geometry and the Imagination (Vol. 87). American Mathematical Soc..
16. Corry, L. (2004). David Hilbert and the axiomatization of physics (1898-1918): From Grundlagen der Geometrie to Grundlagen der Physik (Vol. 10). Dordrecht: Kluwer.
17. Hilbert, D., & Ackermann, W. (1972). Grundzuge der theoretischen logik: 6. Auflage. Springer.
18. Whitehead, A. N., & Russell, B. (1994). Principia Mathematica: Vorwort und Einleitungen. Suhrkamp.
19. Gödel, K. (1992). On formally undecidable propositions of Principia Mathematica and related systems. Courier Corporation.
20. Rodriguez-Consuegra, F. (Ed.). (1995). Kurt Gödel: unpublished philosophical essays. Springer Science & Business Media.
21. Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large language models are zero-shot reasoners. Advances in neural information processing systems, 35, 22199-22213.
22. Curry, H. B. (1954). Remarks on the definition and nature of mathematics. Dialectica, 228-233.
23. Nelson, T. H. (1965, August). Complex information processing: a file structure for the complex, the changing and the indeterminate. In Proceedings of the 1965 20th national conference (pp. 84-100).
24. Braschi, Gianluca. 2024. "Archivi Inventati." Medium
25. Vitali, S. (2004). Passato digitale: le fonti dello storico nell'era del computer. Pearson Italia Spa.
26. Cassirer, E. (2021). The philosophy of symbolic forms: Three volume set. Routledge.
27. Sebeok, T. A., & Danesi, M. (Eds.). (1986). Encyclopedic dictionary of semiotics. de Gruyter.
28. Turing, A. M. (2007). Computing machinery and intelligence. In Parsing the Turing test: Philosophical and methodological issues in the quest for the thinking computer (pp. 23-65). Dordrecht: Springer Netherlands.
29. International Council on Archives. 2023. "Records in Contexts. Conceptual ModelECORDS IN CONTEXTS CONCEPTUAL MODEL." International Council on Archives Expert Group on Archival Description.
30. O'Malley, M., & Rosenzweig, R. (1997). Brave New World or Blind Alley? American History on the World Wide Web. The Journal of American History, 84(1), 132-155.
31. Rosenzweig, R. (2001). The road to Xanadu: Public and private pathways on the history web. The Journal of American History, 88(2), 548-579.
32. Ministero per i Beni Culturali e Ambientali Ufficio Centrale per i Beni Archivistici Archivio di Stato di Torino, Regione Piemonte - Assessorato alla Cultura, CSI Piemonte, and Istituto Bancario S. Paolo di Torino. 1986. Informatica e Archivi. Atti Del Convegno. Torino 17-19 Giugno 1985. Roma: Istituto Poligrafico e Zecca dello Stato - Libreria dello Stato.
33. Graham, S., Milligan, I., Weingart, S. B., & Martin, K. (2016). Exploring big historical data: the historian's macroscope.
34. Allegrezza, S. (2017). Towards a new archival economy: the development of the discipline in the transition

from analogue to digital. JLIS: Italian Journal of Library, Archives and Information Science= Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione: 8, 1, 2017, 114-126.

35. Furet, F. (1981). Il quantitativo in storia. Fare storia. Temi e metodi della nuova storiografia, 3-24.

36. Furet, F. (1971, February). L'histoire quantitative et construction du fait historique. In Annales. Histoire, Sciences Sociales (Vol. 26, No. 1, pp. 63-75). Cambridge University Press.-

37. Darcy, R., & Rohrs, R. C. (1995). A guide to quantitative history. (No Title).

38. Dollar, C. M., & Jensen, R. J. (1971). Historian's guide to statistics: quantitative analysis and historical research. (No Title).

39. Furet, F. (1968). Sur quelques problèmes posés par le développement de l'histoire quantitative. Social Science Information, 7(1), 71-82.

40. Itzcovitch, O. (1996). Dal mainframe al personal: il computer nella storiografia quantitativa. Storia&Computer: Alla ricerca del passato con l'informatica, 30-47.

41. Chaunu, P. (2019). Séville et l'Atlantique, 1504-1650: Structures et conjoncture de l'Atlantique espagnol et hispano-américain (1504-1650). Tome II, volume 1: La conjoncture (1504-1592). Éditions de l'IHEAL.

42. Chaunu, P. (1964). Histoire quantitative ou histoire sérielle. Cahiers Vilfredo Pareto, 2(3), 165-176.

43. Trojano, P. R., & Coppolino, S. (1983). La storia come scienza sociale.

44. Rowland, R. (1991). L'informatica e il mestiere dello storico. Quaderni storici, 693-720.

45. Braudel, F., Tenenti, A., & Salsano, A. (1980). Scritti sulla storia. Mondadori.

46. Bush, V. (1945). As we may think. The atlantic monthly, 176(1), 101-108.

47. Itzcovich, O. (1989). Lo storico e il database. Quaderni

storici, 24(70 (1), 321-325.

48. Itzcovich, O. (1993). L'uso del calcolatore in storiografia (Vol. 5). FrancoAngeli.

49. Spina, S. (2021). Per una Digital History delle pandemie. Prospettive d'interconnessione e multidisciplinarità. Umanistica Digitale, (10), 271-288.

50. Salvatori, E. (2017). Digital (Public) History: la nuova strada di una antica disciplina. RIME–RIVISTA DELL'ISTITUTO DI STORIA DELL'EUROPA MEDITERRANEA, 1(I), 57-94.

51. Ahmed, S. E. (Ed.). (2014). Perspectives on Big Data analysis: methodologies and applications (Vol. 622). American Mathematical Society.

52. Eijnatten, J. V., Pieters, T., & Verheul, J. (2013). Big data for global history: The transformative promise of digital humanities. BMGN-Low Countries Historical Review, 128(4), 55-77.

53. Franzosi, R. (2017). A third road to the past? Historical scholarship in the age of big data. Historical Methods: A Journal of Quantitative and Interdisciplinary History, 50(4), 227-244.

54. Lane, R. (2016). The big humanities: Digital humanities/ digital laboratories. Routledge.

55. Doria, G. P. (2022). A new archives definition. JLIS. it, 13(2), 156-173.

56. Spina, A. (2021). Preservice Mathematics Teachers' Conceptualization of the Standards for Mathematical Practice: A Study Across Four Universities (Doctoral dissertation, University of California, Santa Barbara).

57. Durrant, T., Lilly, A., & Tingay, P. (2022). WhatsApp in government: How ministers and officials should use messaging apps–and how they shouldn't.

58. Allegrezza, S. (2023). "Effimeri ma non troppo". La conservazione dei messaggi istantanei: è ora di cominciare ad occuparcene?. *AIDAinformazioni*, (1-2), 9-31.