Computational Archival Processes & Assessable Sustainability: Challenges and Opportunities

Aurèle Nicolet
Information Science Dpt.
Geneva School of Business Administration
University of Applied Sciences and Arts of Western Switzerland (HES-SO)
Geneva, Switzerland
https://orcid.org/0000-0003-3501-427X

Basma Makhlouf Shabou
Information Science Dpt.
Geneva School of Business Administration
University of Applied Sciences and Arts of Western Switzerland (HES-SO)
Geneva, Switzerland
https://orcid.org/0000-0003-0980-0517

***Abstract*— This article highlights the environmental impacts associated with information and communication technologies (ICT) used for data storage and processing, emphasizing the significant emissions generated during the lifecycle of electronic devices. It addresses the challenges of assessing these environmental impacts using methodologies like the GHG Protocol and Life Cycle Assessment (LCA). It also explores opportunities for mitigating these impacts through better data governance, techniques for reducing digital waste, and sustainability initiatives such as the Arch'Eco project, who aims to assess environmental impacts for the entire lifecycle of data and identify best practices for data management in an environmentally friendly manner.**

Keywords— Computational Archival Science, Sustainability, Environmental Impact

I. Introduction

Climate change and the management of its impact on human societies are among the challenges facing us at the start of the 21st century [1]. Sustainable development and environmental protection are becoming economic, legal, and regulatory issues. This has led to the introduction of an environmental management system, governed by the ISO 14001 standard [2], which covers various areas such as the choice of building materials and waste management. However, the issue of sustainability is not confined to just a few sectors, but applies to all aspects of society, including the processing and management of information, particularly digital information.

Indeed, the digital footprint is not a virtual one, and has very real effects on the environment, such as the depletion of resources in the manufacture of IT hardware and the consumption of water and energy to power it. According to a report on the global digital footprint [3], if the digital world - made up of all electronic equipment using binary data (computers, games consoles, printers, etc.) - were a country, its environmental footprint would be two to three times that of France.

Three of the 17 sustainable development goals to be achieved by 2030 by UN member states [4] require the development of sober and responsible digital technologies: 9) Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation; 12) Ensure sustainable consumption and production patterns; and 17) Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development.

The development of computing in the second half of the 20th century led to an increase in the volume of data generated each year. Since then, this increase has accelerated almost exponentially, from 2 zettabytes in 2010 to 64 zettabytes in 2020 [5], because of several factors such as [6]:

- the increase in the size of mobile applications and web pages [7] or new, more data-hungry uses, such as video streaming [8];
- the rise of digital transformation, which is leading to the dematerialization of media and processes, accelerated by the health crisis linked to COVID-19 [9];
- the exploitation of data through Big Data by companies and institutions to help them make decisions [10][11][12].

The aim of this article is to reflect on the challenges and opportunities facing computational archival processes in the context of climate change and sustainability. Firstly, the authors will outline the challenges by presenting the environmental impacts caused by the production, analysis, and processing of data, and then by presenting the different methodologies currently in use and their main limitations. Secondly, the authors will present the opportunities by outlining several initiatives in the field, particularly the Arch'Eco project.

II. Key Concepts

To better understand our topic, it is important to specify the key concepts, such as computational archival science; conceptual archival processes; sustainability.

According to the working definition [13], Computational Archival Science (CAS) is "A transdisciplinary field that integrates computational and archival theories, methods and resources, both to support the creation and preservation of reliable and authentic records/archives and to address large-scale records/archives processing, analysis, storage, and access, with aim of improving efficiency, productivity and precision, in support of recordkeeping, appraisal, arrangement and description, preservation and access decisions, and engaging and undertaking research with archival material".

Computational Archival Processes (CAP) are the operational level of CAS. They focus on methods, tools, metrics needed to conduct different processes related to archival functions during the entire archives lifecycle.

The sustainable development is the "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" [14]. Although the terms 'sustainability' and 'sustainable development' have some nuances, it is increasingly common practice to regard them as interchangeable terms [15]. Often reduced by the general public to its ecological aspect, the concept of sustainability is based on three pillars: economic, environmental and social [15][16], also known as the triple bottom line. The triple bottom line is a concept first developed in the 1990s by John Elkington [17][18][19]. While it was very quickly adopted in managerial discourse [16][18][19],

Published in 2024 IEEE International Conference on Big Data (BigData), 2024, pp. 2470-2478, which should be cited to refer to this work. doi 10.1109/BigData62323.2024.10826059

the academic world has also adopted these three pillars, 3Ps (people, profit, planet) [Lozano 2022, cited by Aytac et al. 2023], or even proposed new models: 5Ps (people, planet, prosperity, peace, and partnership), 6Ps (people, planet, prosperity, peace, partnership and participation) or 7Ps (people, planet, prosperity, peace, partnership, profit and participation) [16].

The term 'sustainability assessment' encompasses the various processes aimed at evaluating the impact of an activity on the three pillars of sustainability: economic, environmental, and social. This assessment can be prospective (*ex ante*) or retrospective (*ex post*) [20].

An environmental impact is all "change to the environment, whether adverse or beneficial, including possible consequences, wholly or partially resulting from an organization's environmental aspects" [14].

III. Problem

Sustainable development and environmental protection are becoming economic, legal, and regulatory issues. The issue of sustainability does not limit itself to a few sectors but extends to all components of society, including the field of information treatment and management, particularly digital.

Indeed, the digital world is not virtual and has tangible effects on the environment, such as the depletion of resources to produce IT equipment and the consumption of water and energy for its operation.

While digital sobriety struggled in 2019 to be integrated into the strategy of organizations and institutions [21], the gradual adoption of measures by governments encourages them to take a more active approach. For example, the City of Geneva [22] has outlined its digital transition policy in four axes, the first of which is to Promote and implement responsible, ecological, and ethical digital practices in the City of Geneva. However, to establish a strategy, it is necessary to have a clear understanding of the initial situation and to have indicators and metrics that allow for the evaluation of the effectiveness of undertaken measures.

In France, the Regulatory Authority for Electronic Communications, Postal Services, and Press Distribution (Arcep) notes, in its report "For a Sustainable Digital Environment," differences in estimates among the numerous studies on the environmental footprint of digital technology published in recent years, differences that may arise from geographical scope as well as from the methodological choices made. This is why it deems it necessary to develop a finer vision and to define precise measurement methodologies that allow for a more granular evaluation to determine the levers of action and assess their implementation [23].

Currently, reflections on the environmental impact of digital technology are primarily based on hardware and aim to improve tools and equipment (eco-design) to reduce consumption and prolong their use. Through this project, we aim to broaden the approach to include human behavior, particularly information practices, which are often outside the scope of investigation. However, these practices can have significant implications, such as the unintentional redundancy of data leading to an increased need for storage capacity or a multiplication of queries to search for information, which will demand more from the hardware. Currently, these implications do not have factual measurements; a gap that our project aims to fill.

IV. Objectives and Methodology

Our study, as part of the Arch'Eco project, is structured in three main dimensions.

The first consists of an assessment of practices and a mapping of the processes and tasks applied to the acquisition and processing of data and archives. It includes a systematic literature review and case studies to identify best practices (methods, tools, metrics) in terms of sustainable data and archive management.

The second aims to develop a conceptual model that provides an overview of archival functions, related tasks, and tools, with the help of a mapping of metrics and consultation sessions with partners to gather academic and professional validations.

The third presents the functional and technical specifications of a management and control tool, in addition to the development and testing of decision support devices. It will also propose the design of short training sessions and awareness materials.

V. Main Challenges

A. Information and communication technology (ICT) environmental impacts

To be stored and processed, data requires a whole ecosystem of IT equipment, which can be grouped into three main categories: client, server, database. The manufacture, use and disposal of the hardware has a significant impact on the environment.

The manufacturing phase has the greatest impact on the environment. For example, the Agence de l'Environnement et de la Maîtrise de l'Énergie (ADEME) [24] estimates the carbon footprint of a laptop at 169 kg of CO2 for its entire life cycle, including 124 kg for its manufacture alone, i.e. almost three-quarters (73%) of its footprint.

Computer hardware is made up of dozens of different components [25]. These include plastics and synthetic materials; glass and ceramics; ferrous and non-ferrous metals (copper, aluminium, zinc, tin, chromium, nickel, etc.), precious metals (gold, silver, platinum, palladium, etc.) and rare earths (europium, yttrium, terbium, gallium, etc.) [24].

Metals and rare earths require a significant amount of material to be extracted, around 800 kg for the components of a laptop [24]. This requires energy and chemical treatment, and in the case of rare earths also generates radioactive pollution due to the nature of these elements [26]. In addition, working conditions in mines and factories are often poor [27]. Although this factor is not considered in 'traditional' life cycle analysis methodologies, it is now considered in life cycle analyses that include the social dimension, the Social Life Cycle Assessment (S-LCA).

During the use phase, the environmental impact of the digital world is mainly reflected in two elements: the means needed to produce the electricity that powers the devices and the cooling system of the data centers [27][10]. Even though a major effort is being made to encourage and develop renewable energies [29][30], electricity production still relies heavily on fossil fuels [31]. Data centers consume an enormous amount of water [32]. For example, the consumption of data centers in the US in 2014 was estimated at around 626 billion liters [33].

Finally, the disposal or recycling of waste electrical and electronic equipment (WEEE) or electronic wastes (e-wastes) is extremely complex because of the toxicity and diversity of the materials. This is why, in 2019, only 9.3 million tonnes out of 53.6 million tonnes (17.4%) were properly collected and recycled [34][35]. Although the recycling of WEEE is particularly complex because of the toxicity and diversity of the materials (already mentioned in the manufacturing phase), it offers several advantages. Firstly, it significantly reduces the environmental impact of extracting components. Secondly, it can also provide a good opportunity for employment [27][36]. Finally, recycling can also combat the shortages that are likely to occur in certain materials [37]. Mountains of waste thus become 'urban mines' [25].

B. How to assess the environmental impacts

There are various methodologies used to assess the environmental impact of human activities. The two main ones are the GHG Protocol and the life cycle assessment (LCA).

1) GHG Protocol

The GHG Protocol Initiative is "a multi-stakeholder partnership of businesses, non-governmental organizations (NGOs), governments, and others convened by the World Resources Institute (WRI), a U.S.-based environmental NGO, and the World Business Council for Sustainable Development (WBCSD), a Geneva-based coalition of 170 international companies. Launched in 1998, the Initiative's mission is to develop internationally accepted greenhouse gas (GHG) accounting and reporting standards for business and to promote their broad adoption" [38]. Over the years, the GHG Protocol Initiative has issued several standards, briefly summarized in the following paragraphs.

The GHG Protocol Corporate Accounting and Reporting Standard was first published in September 2001, then revised in March 2004. It focuses solely on the accounting and reporting of emissions [38]. For each section, the document presents the requirements of the standard, as well as advice on how to apply them.

The inventory is built around three scopes. Scope 1 focuses on direct emissions, i.e. emissions from sources owned or controlled by the organization, primarily [38]: Production of electricity, heat or steam, from the combustion of fuels; Physical or chemical treatment, resulting from the transformation of products (manufacturing); Transport of materials, products, waste and employees by the organization's vehicles/devices; Fugitive emissions, which correspond to releases linked to leaks, for example.

Scope 2 corresponds to indirect emissions linked to electricity consumption. Transmission and distribution losses are only counted by the organization distributing the electricity [38].

Scope 3 corresponds to indirect emissions that are not covered by scope 2. This includes the extraction, production, and transport of purchased materials. The choice of categories is left to the organizations, which means that no precise comparison can be made [38].

The GHG Protocol for Project Accounting, published in 2005, provides specific principles, concepts and methods for quantifying and reporting GHG reductions - i.e. decreases in GHG emissions, or increases in removals and/or storage - from climate change mitigation projects (GHG projects).

The two standards presented above form the basis of the GHG Protocol and were adopted in 2006 as the first version of the ISO 14064 standard, after several years of negotiation between the International Organization for Standardization (ISO), the World Resources Institute (WRI) and the World Business Council for Sustainable Development (WBCSD) [39]. Since then, the GHG Protocol Initiative has issued several standards that complement this base.

The Corporate Value Chain (Scope 3) Standard, published in 2011, complements the GHG Protocol Corporate Accounting and Reporting Standard by providing a standardized approach for preparing and reporting a GHG emissions inventory that includes indirect emissions resulting from value chain activities (i.e. Scope 3 emissions) [40].

The Product Life Cycle Accounting and Reporting Standard, published in 2011, provides requirements and guidance for companies and other organizations to quantify and publicly report an inventory of GHG emissions and removals associated with a specific product [41].

The GHG Protocol Mitigation Goal Standard, published in 2014, provides guidance for designing mitigation goals at national and sub-national levels, as well as a standardized approach for assessing and reporting progress towards goals [42].

The GHG Protocol Policy and Action Standard, published in 2014, provides a standardized approach for estimating and reporting the change in GHG emissions and removals resulting from policies and actions [43].

The Global Protocol for Community-Scale Greenhouse Gas Inventories: An Accounting and Reporting Standard for Cities was first released in 2014 and will be updated in 2021. It provides a framework for calculating and reporting GHG emissions at a city scale, classifying greenhouse gas emissions into six main sectors: 1) stationary energy (building heating or electricity generation), 2) transport, 3) waste, 4) industrial processes and product use, 5) agriculture, forestry, and other land uses, and 6) any other emissions occurring outside geographical boundaries and resulting from the city's activities.

2) Life cycle assessment (LCA)

The life cycle assessment (LCA) is the "compilation and assessment of the inputs, outputs and the potential environmental impacts of a product system throughout its life cycle" [14], so LCA can be applied to the construction of a house as well as to its rental. The case of an organization is more complex, as it is made up of many goods and services. This is why its analysis is the subject of a specific standard, ISO 14072:2014.

LCA is an iterative technique, the principles and framework of which are defined by ISO 14040:2006 and the requirements and guidelines by ISO 14044:2006. LCA consists of four phases: 1) definition of the goals and scope of the study, which aims to clearly define the field of application of the study (purpose, scope, public concerned, use of results, etc.); 2) inventory analysis, which includes both the collection and calculation of data to quantify the relevant inputs and outputs; 3) impact assessment, which seeks to estimate the extent of potential environmental impacts based on the results of the inventory; 4) interpretation, which combines the results of the inventory and the assessment and checks their consistency with the objectives and scope of the study established in the first phase.

The LCA method offers a great deal of flexibility and requires choices to be made that will have an impact on the final results, which makes it difficult to make comparisons between different similar organizations. For this reason, several impact assessment methods have emerged, each of which proposes characterization factors that enable "to convert an assigned life cycle inventory analysis result to the common unit of the category indicator" [14]. For example, the climate change impact category of the ReCiPe 2016 method has, as a characterization factor, the global warming potential expressed in kilo of CO2 equivalent [44].

These main methods include [45][46][47]: CML (for Centrum voor Milieukunde Leiden); ILCD (for International Life Cycle Data); ReCiPe (for RAISONnement COMparative Effect Pathways for the inclusion of midpoint and Endpoint effects) and USEtox (for USEful toxicity model).

In addition to impact assessment methods, harmonization efforts are continuing, with guides and standards being proposed to standardize assessments and facilitate comparisons within the same sector. The European Community first developed the European International Life Cycle Data System (ILCD), which provides a series of practical guides for setting up an LCA, and then undertook the development of the Product Environmental Footprint Category Rules (PEFCRs) and the Organization Environmental Footprint Sector Rules (OEFSRs), which provide precise methodologies for specific groups of products.

In the digital sector, standards are beginning to be developed. In France, the NegaOctet consortium with ADEME has developed four environmental assessment methodologies: 1) for Internet Service Providers (ISPs) [48]; 2) for data center hosting and cloud services [49]; 3) for LAN networks and corporate telephony services [50]; and 4) for digital services [51].

VI. Main opportunities

A. *Contribution to reduce the environmental impacts*

The environmental impacts presented in the previous paragraphs are linked to the equipment needed to store and process the data. However, a significant proportion of data is unused, or even considered to be waste, even if the exact proportion is difficult to estimate - between 80% and 99% of all data [33]. These Dark data are therefore expensive in terms of storage and processing... costs that are not only economic, but also ecological [33][52][53]. By helping to combat dark data, CAS can indirectly contribute to reducing the environmental impact of data.

1) Recent initiatives

Several studies have looked at the environmental impact of data management, either by calculating the carbon footprint or by proposing a model to support more sustainable data management.

For example, Kinnaman and Munshower examined the information practices of Virginia Tech libraries and were particularly interested in estimating the energy consumption and carbon emissions involved in calculating the checksum used to verify the integrity of digital documents. To do this, they used a study by Damasevicius, Ziberkas, Stuikys, and Toldinas [54] which estimated the energy consumption of hash functions at 40mj for a 1mb file, and cross-referenced it with data from the U.S. Energy Information Administration to obtain CO2 emissions. Based on these initial results, they proposed a series of recommendations, like determine acceptable loss or reduce the size of objects to be preserved [55].

Wijsman, Groen, van Zwol and Gillesse looked at the carbon emissions associated with the storage and use of digital objects and data, based mainly on the energy consumed by servers [56].

Tiainen, Lehtonen, Helin and Kylander sought to calculate the carbon footprint of two data repository services funded by the Ministry of Education and Culture of Finland: 1) Digital Preservation Service for Cultural Heritage and 2) The Digital Preservation Service for Research Data [57]. To do this, they broke down data repository services into different elements: 1) hardware needed for the preservation; 2) data centers; 3) network; 4) administration work; 5) development work; and 6) supporting ICT-services. The study excludes the optical network, communication data or common support elements for production and development. The method for calculating the carbon footprint itself is not really explained and is based mainly on the use of two tools:

- Product Attribute to Impact Algorithm (PAIA), developed by the Massachusetts Institute of Technology (MIT) to calculate the environmental footprints of Information & Communication Technology (ICT) products.
- Y-HIILARI Hiilijalanjälki -työkalu, a carbon footprint calculation tool developed by the Finnish Environment Institute.

The three studies, although very interesting, focus mainly on the carbon footprint and do not take other impact categories into account. They also focus on the use of devices rather than their construction or disposal, and finally they only look at preservation without considering the entire life cycle of the data.

For their part, Bussel, Smit, and De Pas have sought to adapt the Green IT model and thus propose Green Archiving [58][59]. This model is made up of three elements: Green Computing, Archival Retention Levels, and the Information Value Chain.

Green Computing is the study and practice of designing, manufacturing, using, and disposing of computers, servers and associated subsystems such as monitors, printers, storage devices and networking and communications systems, effectively and efficiently, with minimal or no impact on the environment. It can be broken down into six categories: 1) Product Longevity; 2) Optimization; 3) Power Management; 4) Recycling; 5) Telecommuting; and 6) Low Power IT.

The Archival Retention Level defines the detailed functional (organizational) responsibilities for the analysis, processing, and storage of specific data.

The Information Value Chain is the use of the informational and evidential value of data in business processes to improve the management of trusted data and the performance of business processes. This includes all information and data management processes, such as processing, structuring, publishing, appraisal, etc.

In their article [28], Pendergrass, Sampson, Walsh and Alagna present a reflection on evaluation to have sustainable preservation in every sense of the word, thanks to a more drastic evaluation of archives and less redundant preservation, even if it means having losses.

Based on the article by Pendergrass et al [28], Paschalidou, Fafet and Milios have developed a framework for sustainable data preservation based on three strategies [60]:

- Sufficiency strategy, aimed at reducing the amount of data to be preserved by only collecting cultural objects that correspond to the institution's policy or by only digitizing what is explicitly necessary;
- Efficiency strategy, aimed at using less energy and equipment to preserve the same amount of data;
- Digital Preservation-Efficiency Strategy, which mainly aims to determine a sufficient level of digital preservation with an acceptable loss threshold and to adjust the level of preservation quality according to the value of the data to be preserved.

The aim is to provide a framework for the sustainable digitization (in both senses of the term) of cultural heritage.

2) Identified gaps

The literature review allowed to identify several difficulties and challenges concerning environmental impact assessment. Three seem important to note.

Firstly, the lack of unified methods for assessing environmental impacts makes it difficult to compare the ecological effects of different practices or technologies. This often leads to inconsistent data and conclusions that are difficult to generalize or apply across different studies or sectors. As mentioned above, the NegaOctet consortium has begun to develop repositories for the digital domain, but similar work is lacking for the records management and archives domain.

Secondly, many studies attending to measure environmental impact frequently focus on carbon footprint. However, it is essential to recognize that it is not the sole indicator of environmental health. Other factors also play crucial roles and should be integrated into environmental assessments to provide a more comprehensive understanding. For example, the NegaOctet consortium consider as mandatory to take into account five environmental indicators [51]: 1) Abiotic depletion (minerals and metals), measured in amount of material missing for future generations (kg Sb eq); 2) Climate change, measured in radiative forcing as global warming potential (kg CO_2 eq); 3) Acidification, measured in accumulated exceedance (mol H^+ eq); 4) Particulate matter, measured in impact on human health (disease incidence); 5) Ionising radiation, measured in human exposure efficiency relative to U-235 (kBq U-235 eq).

Finally, we can note a lack of holistic perspective on environmental impacts due to the scarcity of studies examining the entire life cycle of data. Many evaluations concentrate on singular aspects such as data preservation, but this narrow focus fails to capture the broader environmental consequences associated with data production, storage, usage, and disposal. In France, the Serda group has developed the first reference framework for responsible digital information management, proposing 21 key indicators for responsible information management [61]. These indicators correspond to actions, such as sending an email or scanning a page, to which a carbon footprint in grams of CO_2 equivalent with a high and a low value is assigned. However, these are only theoretical estimates and are not based on concrete cases.

The Arch'Eco project, presented below, aims to fill these gaps.

3) Arch'Eco

Financed by the RCSO IS-net fund from the University of Applied Sciences and Arts of Western Switzerland, Arch'Eco is a project leaded by the Geneva School of Business Administration's Information Science Department with private French partners. The Arch'Eco project has several objectives: 1) Identify best practices in data and archive management; 2) Characterize the ecological issues involved in acquiring and processing data and archives; 3) Identify the environmental impacts of data processing via the life cycle of data and archives; 4) Quantify the environmental impacts of data and archive acquisition and processing; 5) Define the functional and technical specifications of a tool to manage and control environmental costs; 6) Recommend resources to better support ecological sustainability among decision-makers.

The methodology rest on three main methods: 1) literature review on the information / records management best practices, the environmental impact / sustainability assessment methods, and the existing tools; 2) several case studies (research repository, public administration, international organization, massive data producer…) and 3) consultations with the academic and professional community for testing the conceptual model.

Several deliverables are expected: 1) a synthesis of best practices and recurring issues in data and archive processing in an ecologically sustainable approach; 2) a summary of calculation tools and simulators for measuring carbon footprints and ecological costs; 3) a conceptual model defining the axes, indicators and metrics of data and archive impact management; 4) the production of several monitoring tools, like a simulator to specify the technical specifications of the control tool and their feasibility or awareness-raising and training materials.

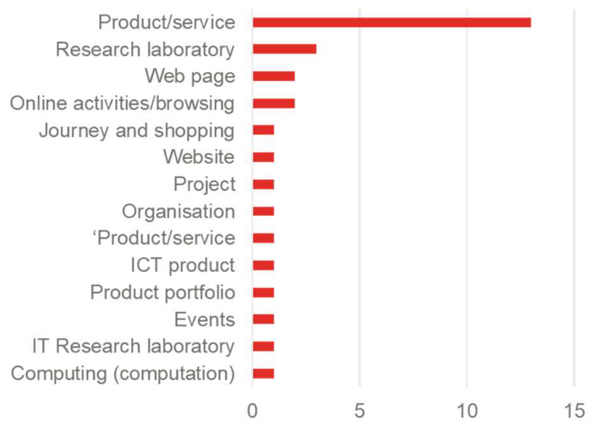

Fig. 1. Scope of the 29 identified tools

As a preliminary result, the literature review has so far identified 29 tools for assessing environmental impacts. The majority (17) propose a multi-criteria assessment based on one or more LCA methods, while the remainder (12) focus on a single impact category, climate change, with CO_2 as the indicator. These tools are mainly aimed at assessing the environmental impact of products or services (fig.1).

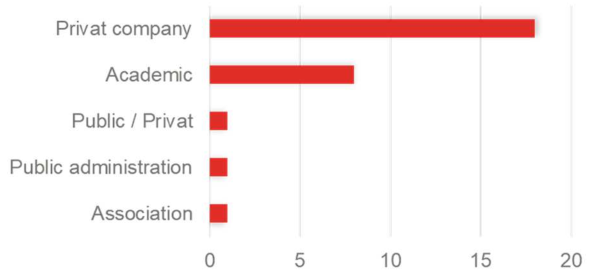

Fig. 2. Type of organisation that developed the tools

Most of these tools have been developed by private companies (fig.2), which use proprietary code. Only 10 tools are open source. Most of these tools are web applications or cloud solutions (fig.3).

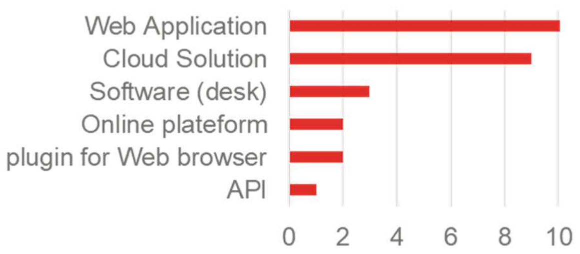

Fig. 3. Type of tool configuration

For the next stages, the Arch'Eco project aims to define functional and technical specifications for measuring tool of piloting the environmental impacts applicable on data processing, and then to test a simulator on OLOS, a service of research data repository [62], as part of a doctoral research project.

B. *Contribution to study on climate change*

Indirectly, the analysis of large-scale archives using the possibilities offered by Big Data can help us to better understand climate change or how previous populations were able to overcome similar problems [63].

Oceanography is a good example of this, because over the course of the 17th, 18th and especially 19th centuries, the meteorological observations recorded in ships' logbooks gradually became codified. In 1853, the Brussels Maritime Conference sought to set up an international system for collecting and exchanging meteorological data on the oceans, thus laying the foundations for the establishment of the World Meteorological Organization [64].

Recently, Walker and Ummenhofer, a maritime historian and an oceanographer, seeked to extract Global Maritime Weather Data from approximately 4,200 North American whaling and 2,200 Portuguese Navy logbooks dating to the middle eighteenth century [65]. To achieve this, a team of trained human readers has been tasked with manually extracting data from the logbooks. While the technology is already well advanced for medieval manuscripts [66][67][68][69], this is not yet the case for logbooks, which present several challenges, such as a wide variety of forms and a medium that is very often deteriorated [65].

VII. Discussions

The previous sections may give the impression of advocating for an extremely sparing use of ICTs at the risk of deteriorating the management and preservation of archives. This not really the intension. Indeed, while the accurate estimation of the potential environmental impacts of ICTs should bring insights and will educate strategies and discussions about archives management. Therefore, it is essential to measure the environmental impact throughout the data and archives lifecycle to identify key areas to focus on and make rational use of the available technologies.

This reflection aims to consider the implications of digital sobriety for archives and archivists by presenting some issues related to the information and data lifecycle.

It is not possible to keep everything. In the face of the exponential growth of generated data, evaluation and its automation become indispensable. Several studies have addressed this issue [70] [71] [72].

The preservation of data requires redundant backups (archival copy, distribution copy, etc.) and regular migration of media to address the evolution and risk of obsolescence. In addition, data integrity checks using checksums are necessary. To reduce energy consumption or hardware fabrication costs, several considerations and positions should be taken on the frequency of migrations and integrity checks or extending the lifespan of media. In the medium and long term, new storage media, such as DNA or quartz, may offer alternative reduction paths.

Regarding the sharing and access to archives, while digitization and online access have negative environmental impacts due to server usage, they can also have positive environmental impacts by, for example, reducing user and researcher travel to archives.

Implementing responsible information management presents numerous challenges for archivists. However, it is also an opportunity to advocate for archivists by highlighting the importance of controlled information governance to reduce not only financial costs but also the environmental impact of an organization.

VIII. Conclusion

In an increasingly data-driven world, the intersection of Computational Archival Processes (CAP) and sustainability underscores the pressing need for innovative approaches to managing environmental impacts associated with information and communication technologies. As this paper has demonstrated, the digital footprint of archival practices is substantial, with the environmental consequences of data storage and processing warranting careful consideration. Notably, the lifecycle emissions generated by electronic devices present significant challenges in accurately assessing sustainability.

Through our exploration of methodologies such as the GHG Protocol and Life Cycle Assessment (LCA), we have identified gaps in current literature regarding comprehensive evaluations of the entire lifecycle of data, from acquisition through to disposal. Existing studies often focus narrowly on specific phases or aspects, neglecting the broader context of data management in the sustainability dialogue.

The Arch'Eco project represents an essential step toward addressing these gaps by establishing best practices in sustainable data governance and resource management. By promoting methodologies that encompass the entire lifecycle of data, this initiative aims to facilitate actionable insights for organizations to evaluate and mitigate their environmental impact.

In conclusion, fostering a culture of sustainable archival practices is crucial as we navigate the challenges posed by climate change and environmental degradation. By integrating principles of sustainability into the core of archival science, we can create a more robust framework that not only preserves information but also safeguards our environment for future generations. Emphasizing a holistic understanding of data management and encouraging collaboration across disciplines will be essential for achieving sustainable development in this rapidly evolving digital landscape.

Acknowledgment

We would like to thank our academic and professional partners in the project Arch'Eco. We are grateful to Prof Giovanna Di Marzo Serugendo from the University of Geneva for her support.

References

[1] F. Durand, *Le réchauffement climatique : enjeu crucial du XXIe siècle*. éditions Ellipses, 2020. Accessed: Jan. 17, 2021. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02549992

[2] International Organization for Standardization, "Environmental management systems — Requirements with guidance for use," International Organization for Standardization, Geneva, ISO 14001, Sep. 2015.

[3] F. Bordage, "Empreinte environnementale du numérique mondial." GreenIT, Sep. 2019. [Online]. Available: https://www.greenit.fr/wp-content/uploads/2019/10/2019-10-GREENIT-etude_EENM-rapport-accessible.VF_.pdf

[4] "The 17 Goals", United Nations [Online]. Available: https://sdgs.un.org/goals

[5] T. Gaudiaut, "Le Big Bang du Big Data," Statista Infographies. Accessed: Sep. 02, 2022. [Online]. Available: https://fr.statista.com/infographie/17800/big-data-evolution-volume-donnees-numeriques-genere-dans-le-monde/

[6] A. Nicolet, E. Crausaz, E. Jaomazava, and A. Sieber, "Identification et prise en compte de l'impact environnemental des données numériques : un nouveau défi pour la gouvernance de l'information," in *Données documents connaissances perspectives de recherche et d'enseignement: Actes du 22e colloque international sur le document numérique. 9-10 décembre 2021 ¦ CNAM - Paris-France*, G. Chartron, M. Ihadjadene, and E. Broudoux, Eds., Paris, France: Europia, Jun. 2022, pp. 270–289.

[7] D. Bezaleli, J. Gutman, and R. Nossenson, "Using the ZDelta compression algorithm for data reduction in cellular networks," in *2013 Future Network Mobile Summit*, Jul. 2013, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/6633522

[8] K. Widdicks, M. Hazas, O. Bates, and A. Friday, "Streaming, Multi-Screens and YouTube: The New (Unsustainable) Ways of Watching in the Home," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow Scotland Uk: ACM, May 2019, pp. 1–13. doi: 10.1145/3290605.3300696.

[9] R. Subramaniam, S. P. Singh, P. Padmanabhan, B. Gulyás, P. Palakkeel, and R. Sreedharan, "Positive and Negative Impacts of COVID-19 in Digital Transformation," *Sustainability*, vol. 13, no. 16, p. 9470, Aug. 2021, doi: 10.3390/su13169470.

[10] F. Lucivero, "Big Data, Big Waste? A Reflection on the Environmental Sustainability of Big Data Initiatives," *Sci Eng Ethics*, vol. 26, no. 2, pp. 1009–1030, Apr. 2020, doi: 10.1007/s11948-019-00171-7.

[11] C. Sauvajol-Rialland, "Infobésité, gros risques et vrais remèdes," *L'Expansion Management Review*, vol. N° 152, no. 1, pp. 110–118, 2014, doi: 10.3917/emr.152.0110.

[12] S. Taffel, "Data and oil: Metaphor, materiality and metabolic rifts," *New Media & Society*, Jun. 2021, doi: 10.1177/14614448211017887.

[13] 9th COMPUTATIONAL ARCHIVAL SCIENCE (CAS) WORKSHOP, [Online]. Available: https://ai-collaboratory.net/cas/cas-workshops/2024-9th-cas-workshop/

[14] International Organization for Standardization, "Environmental management — Vocabulary," International Organization for Standardization, Geneva, ISO 14050, Jul. 2020.

[15] P. Halla and C. R. Binder, "Sustainability Assessment: Introduction and Framework," in *Sustainability Assessment of Urban Systems*, 1st ed., C. R. Binder, R. Wyss, and E. Massaro, Eds., Cambridge University Press, 2020, pp. 7–29. doi: 10.1017/9781108574334.

[16] S. Aytac, N. Bautista-Puig, E. Orduña-Malea, and C. Y. Tran, "Contribution of carbon footprint research towards the triple bottom line of sustainability," *Environ Sci Pollut Res*, vol. 30, no. 38, pp. 88331–88349, Jul. 2023, doi: 10.1007/s11356-023-28799-w.

[17] J. Elkington, "ACCOUNTING FOR THE TRIPLE BOTTOM LINE," *Measuring Business Excellence*, vol. 2, no. 3, pp. 18–22, Mar. 1998, doi: 10.1108/eb025539.

[18] Q. Farooq, P. Fu, X. Liu, and Y. Hao, "Basics of macro to microlevel corporate social responsibility and advancement in triple bottom line theory," *Corp Soc Responsibility Env*, vol. 28, no. 3, pp. 969–979, May 2021, doi: 10.1002/csr.2069.

[19] F. Quairel and M. Capron, "Le couplage « responsabilité sociale des entreprises » et « développement durable » : mise en perspective, enjeux et limites," *Revue Française de Socio-Économie*, vol. 11, no. 1, pp. 125–144, 2013, doi: 10.3917/rfse.011.0125.

[20] J. Pope, A. Bond, J. Hugé, and A. Morrison-Saunders, "Reconceptualising sustainability assessment," *Environmental Impact Assessment Review*, vol. 62, pp. 205–215, Jan. 2017, doi: 10.1016/j.eiar.2016.11.002.

[21] J. Le Bolzer, "« La sobriété numérique peine à s'inscrire à l'agenda des directions générales »," Les Echos. Accessed: Mar. 11, 2020. [Online]. Available: https://www.lesechos.fr/thema/economie-nouvelle-generation/la-sobriete-numerique-peine-a-sinscrire-a-lagenda-des-directions-generales-1140894

[22] Ville de Genève, "Politique de transition numérique: l'essentiel en bref." 2022. [Online]. Available: https://www.geneve.ch/sites/default/files/2022-03/en-bref-politique_transition_numerique-ville-geneve.pdf

[23] Autorité de régulation des communications électroniques, des postes et de la distribution de la presse (Arcep), "Pour un numérique soutenable." Autorité de régulation des communications électroniques, des postes et de la distribution de la presse (Arcep), Dec. 15, 2020. Accessed: Oct. 12, 2022. [Online]. Available: https://www.arcep.fr/uploads/tx_gspublication/rapport-pour-un-numerique-soutenable_dec2020.pdf

[24] Agence de l'Environnement et de la Maîtrise de l'Énergie, *La Face cachée du numérique: réduire les impacts du numérique sur l'environnement*. Angers: Agence de l'Environnement et de la Maîtrise de l'Énergie, 2019. [Online]. Available: https://librairie.ademe.fr/cadic/2351/guide-pratique-face-cachee-numerique.pdf

[25] A. Geldron, "Métaux stratégiques : la mine urbaine française:," *Annales des Mines - Responsabilité et environnement*, vol. N° 82, no. 2, pp. 67–73, Apr. 2016, doi: 10.3917/re1.082.0067.

[26] Agence de l'Environnement et de la Maîtrise de l'Énergie, "Terres rares, énergies renouvelables et stockage d'énergies." Oct. 2020. [Online]. Available: https://librairie.ademe.fr/energies-renouvelables-reseaux-et-stockage/492-terres-rares-energies-renouvelables-et-stockage-d-energies.html

[27] É. Lèbre *et al.*, "The social and environmental complexities of extracting energy transition metals," *Nat Commun*, vol. 11, no. 1, p. 4823, Dec. 2020, doi: 10.1038/s41467-020-18661-9.

[28] K. L. Pendergrass, W. Sampson, T. Walsh, and L. Alagna, "Toward Environmentally Sustainable Digital Preservation," *The American Archivist*, vol. 82, no. 1, pp. 165–206, Mar. 2019, doi: 10.17723/0360-9081-82.1.165.

[29] J. P. Ríos-Ocampo, S. Arango-Aramburo, and E. R. Larsen, "Renewable energy penetration and energy security in electricity markets," *Intl J of Energy Research*, vol. 45, no. 12, pp. 17767–17783, Oct. 2021, doi: 10.1002/er.6897.

[30] J.-P. Sasse and E. Trutnevyte, "Cost-effective options and regional interdependencies of reaching a low-carbon European electricity system in 2035," *Energy*, vol. 282, p. 128774, Nov. 2023, doi: 10.1016/j.energy.2023.128774.

[31] S. Zhang, Z. Liu, X. Zhang, Z. Wu, and Z. Hu, “Sustainable thermal energy harvest for generating electricity,” *The Innovation*, vol. 5, no. 2, p. 100591, Mar. 2024, doi: 10.1016/j.xinn.2024.100591.

[32] N. Sattiraju, “The Secret Cost of Google’s Data Centers: Billions of Gallons of Water to Cool Servers,” *Time*, Apr. 02, 2020.

[33] C. J. Corbett, “How Sustainable Is Big Data?,” *Production and Operations Management*, vol. 27, no. 9, pp. 1685–1695, Sep. 2018, doi: 10.1111/poms.12837.

[34] V. Forti, C. P. Baldé, R. Kuehr, and G. Bel, “The Global E-waste Monitor 2020.” United Nations University (UNU)/United Nations Institute for Training and Research (UNITAR) – co-hosted SCYCLE Programme, International Telecommunication Union (ITU) & International Solid Waste Association (ISWA), 2020. [Online]. Available: https://www.itu.int/myitu/-/media/Publications/2020-Publications/Global-E-waste-Monitor-2020.pdf

[35] S. C. Chakraborty *et al.*, “Metals in e-waste: Occurrence, fate, impacts and remediation technologies,” *Process Safety and Environmental Protection*, vol. 162, pp. 230–252, Jun. 2022, doi: 10.1016/j.psep.2022.04.011.

[36] Z. Yang *et al.*, “Increasing the energy efficiency of a data center based on machine learning,” *Journal of Industrial Ecology*, p. jiec.13155, Jun. 2021, doi: 10.1111/jiec.13155.

[37] E. Olivié-Paul, “Pour un numérique responsable, repenser les approches vis-à-vis des matières premières minérales,” AdVaes. Accessed: Sep. 30, 2022. [Online]. Available: https://advaes.fr/2021/09/16/pour-un-numerique-responsable-repenser-les-approches-vis-a-vis-des-matieres-premieres-minerales/

[38] World Business Council for Sustainable Development and World Resources Institute, Eds., *The greenhouse gas protocol: a corporate accounting and reporting standard*, Revised ed. o.O: World Resources Institute and World Business Council for Sustainable Development, 2004.

[39] J. F. Green, “Private Standards in the Climate Regime: The Greenhouse Gas Protocol,” *Bus. polit.*, vol. 12, no. 3, pp. 1–37, Oct. 2010, doi: 10.2202/1469-3569.1318.

[40] World Business Council for Sustainable Development and World Resources Institute, Eds., *Greenhouse gas protocol: corporate value chain (Scope 3) accounting and reporting standard : supplement to the GHG protocol corporate accounting and reporting standard*. Washington, DC], [Geneva, Switzerland: World Resources Institute ; World Business Council for Sustainable Development, 2011.

[41] World Resources Institute, Ed., *Greenhouse gas protocol: product life cycle accounting and reporting standard*. Washington, DC, Geneva, Switzerland: World Resources Institute ; World Business Council for Sustainable Development, 2011.

[42] World Resources Institute, Ed., *Greenhouse gas protocol: mitigation goal standard*. Washington: World Resources Institute, 2014.

[43] World Resources Institute, Ed., *Greenhouse gas protocol: policy and action standard*. Washington: World Resources Institute, 2014.

[44] M. A. J. Huijbregts *et al.*, “ReCiPe 2016 v1.1: A harmonized life cycle impact assessment method at midpoint and endpoint level – Report I: Characterization,” National Institute for Public Health and the Environment, Bilthoven (Netherlands), RIVM Report 2016-0104a, 2017. Accessed: Aug. 24, 2024. [Online]. Available: https://www.rivm.nl/documenten/recipe2016v11

[45] A. P. Acero, C. Rodríguez, and A. Ciroth, “LCIA methods: Impact assessment methods in Life Cycle Assessment and their impact categories.” GreenDelta, May 11, 2016. [Online]. Available: https://www.openlca.org/wp-content/uploads/2016/08/LCIA-METHODS-v.1.5.5.pdf

[46] Asha T H, “Life Cycle Impact Assessment (LCIA) Methodologies,” Linkedin. Accessed: Aug. 14, 2024. [Online]. Available: https://www.linkedin.com/pulse/life-cycle-impact-assessment-lcia-methodologies-asha-t-h-rh6pc/

[47] PRé Sustainability, “SimaPro database manual: Methods library.” PRé Sustainability, Mar. 2024. [Online]. Available: https://simapro.com/wp-content/uploads/2024/04/DatabaseManualMethods.pdf

[48] NégaOctet and ADEME, “Référentiel méthodologique d’évaluation environnementale de la Fourniture d’Accès Internet (FAI).” ADEME, Jan. 2023. [Online]. Available: https://negaoctet.org/wp-content/uploads/2023/10/referentiel_ae_fourniture_acces_internet.pdf

[49] NégaOctet and ADEME, “Référentiel méthodologique d’évaluation environnementale des Services d’hébergement informatique en centre de données et de Services Cloud.” ADEME, Jan. 2023. [Online]. Available: https://negaoctet.org/wp-content/uploads/2023/10/referentiel_rcp_datacenter_services_cloud.pdf

[50] NégaOctet and ADEME, “Référentiel méthodologique d’évaluation environnementale d’un réseau LAN et des services de téléphonie d’entreprise.” ADEME, Jan. 2023. [Online]. Available: https://negaoctet.org/wp-content/uploads/2023/10/referentiel_rcp_reseau_lan-services_telephonie_entreprise.pdf

[51] NégaOctet and ADEME, “Référentiel méthodologique d’évaluation environnementale des services numériques.” ADEME, Jul. 2021. [Online]. Available: https://negaoctet.org/wp-content/uploads/2023/10/Referentiel-services-numeriques_v1.0_FR.pdf

[52] A. F. Md Ajis and S. Hajar Baharin, “Dark Data Management as frontier of Information Governance,” in *2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Malaysia: IEEE, Apr. 2019, pp. 34–37. doi: 10.1109/ISCAIE.2019.8743915.

[53] A. F. Md Ajis, J. Md Jali, I. Ishak, and Q. N. Harun, “Enlightening the Repercussion of Dark Data Management towards Malaysian SMEs Sustainability,” *E-BPJ*, vol. 8, no. SI15, pp. 223–229, Sep. 2023, doi: 10.21834/e-bpj.v8iSI15.5070.

[54] R. Damasevicius, G. Ziberkas, V. Stuikys, and J. Toldinas, “Energy Consumption of Hash Functions,” *ElAEE*, vol. 18, no. 10, pp. 81–84, Dec. 2012, doi: 10.5755/j01.eee.18.10.3069.

[55] A. Kinnaman and A. Munshower, “Green GoesS with Anything: Decreasing Environmental Impact of Digital Libraries at Virginia Tech,” presented at the iPres 2022: The 18th International Conference on Digital Preservation, Glasgow Scotland Uk, Sep. 2022. [Online]. Available: http://hdl.handle.net/10919/112392

[56] L. Wijsman, A. Groen, T. van Zwol, and R. Gillesse, “The CO2 Emissions Of Storage And Use Of Digital Objects And Data,” Nov. 2022, doi: 10.17605/OSF.IO/6D87G.

[57] M. Tiainen, J. Lehtonen, H. Helin, and J. Kylander, “Calculating the Carbon Footprint of Digital Preservation: A Case Study,” presented at the iPRES 2023: The 19th International Conference on Digital Preservation, Champaign-Urbana, Sep. 2023. [Online]. Available: https://hdl.handle.net/2142/121104

[58] G.-J. van Bussel and N. Smit, “Building a Green Archiving Model: Archival Retention Levels, Information Value Chain and Green Computing,” in *Proceedings of the 8th European Conference on IS Management and Evaluation*, University of Ghent, Sep. 2014, pp. 271–277. [Online]. Available: https://www.researchgate.net/publication/273777831_Building_a_Green_Archiving_Model_Archival_Retention_Levels_Information_Value_Chain_and_Green_Computing

[59] G.-J. van Bussel, N. Smit, and J. van de Pas, “Digital Archiving, Green IT and Environment. Deleting Data to Manage Critical Effects of the Data Deluge,” *The Electronic Journal Information Systems Evaluation*, vol. 18, no. 2, pp. 188–199, 2015.

[60] E. Paschalidou, C. Fafet, and L. Milios, “A Strong Sustainability Framework for Digital Preservation of Cultural Heritage: Introducing the Eco-Sufficiency Perspective,” *Heritage*, vol. 5, no. 2, pp. 1066–1088, May 2022, doi: 10.3390/heritage5020058.

[61] Serda Conseil and Coopérative Carbone, “Le référentiel 2024 des indicateurs clés de la dématérialisation écoresponsable.” Serda Conseil, 2024.

[62] “Integrated data management solution”, OLOS [Online]. Available: https://olos.swiss/

[63] A. F. More *et al.*, “The Role of Historical Context in Understanding Past Climate, Pollution and Health Data in Trans-disciplinary Studies: Reply to Comments on More et al., 2017,” *GeoHealth*, vol. 2, no. 5, pp. 162–170, May 2018, doi: 10.1029/2017GH000121.

[64] C. Roy and M.-H. Durand, “Des archives climatiques rendues accessibles,” in *Du bon usage des ressources renouvelables*, Y. Gillon, C. Chaboud, J. Boutrais, C. Mullon, and J. Weber, Eds., IRD Éditions, 2000, pp. 359–374. doi: 10.4000/books.irdeditions.25454.

[65] T. D. Walker and C. C. Ummenhofer, “Extracting Global Maritime Weather Data from New England Whaling and Portuguese Navy Logbooks (1740-1960),” *MAINSHEET*, vol. 1, Mar. 2024, doi: 10.61355/001c.94790.

[66] A. Fischer *et al.*, “Automatic Transcription of Handwritten Medieval Documents,” in *2009 15th International Conference on Virtual*

Systems and Multimedia, Vienna, Austria: IEEE, Sep. 2009, pp. 137–142. doi: 10.1109/VSMM.2009.26.

[67] J. Schoen and G. E. Saretto, “Optical Character Recognition (OCR) and Medieval Manuscripts: Reconsidering Transcriptions in the Digital Age,” *Digital Philology*, vol. 11, no. 1, pp. 174–206, Mar. 2022, doi: 10.1353/dph.2022.0010.

[68] S. Torres Aguilar and V. Jolivet, “Handwritten Text Recognition for Documentary Medieval Manuscripts,” *Journal of Data Mining & Digital Humanities*, vol. Historical Documents and..., p. 10484, Dec. 2023, doi: 10.46298/jdmdh.10484.

[69] E. Guéville and D. J. Wrisley, “Transcribing Medieval Manuscripts for Machine Learning,” *Journal of Data Mining & Digital Humanities*, vol. On the Way to the Future of..., p. 9805, Jul. 2024, doi: 10.46298/jdmdh.9805.

[70] T. Antsoupova, “To keep or not to keep: Automating appraisal and selection at the National Archives of Australia,” *JDMM*, vol. 4, no. 1, p. 82, Nov. 2015, doi: 10.69554/LJWM9147.

[71] G. Rolan, G. Humphries, L. Jeffrey, E. Samaras, T. Antsoupova, and K. Stuart, “More human than human? Artificial intelligence in the archive,” *Archives and Manuscripts*, vol. 47, no. 2, pp. 179–203, May 2019, doi: 10.1080/01576895.2018.1502088.

[72] B. Makhlouf Shabou, J. Tièche, J. Knafou, and A. Gaudinat, “Algorithmic methods to explore the automation of the appraisal of structured and unstructured digital data,” *RMJ*, vol. 30, no. 2, pp. 175–200, Jul. 2020, doi: 10.1108/RMJ-09-2019-0049.