



HAL
open science

Beyond capstone: toward a new strategy for appraising and selecting emails to transfer to archives within French public agencies

Edgar Lejeune, Bénédicte Grailles, Touria Aït El Mekki, Patrice Marcilloux

► **To cite this version:**

Edgar Lejeune, Bénédicte Grailles, Touria Aït El Mekki, Patrice Marcilloux. Beyond capstone: toward a new strategy for appraising and selecting emails to transfer to archives within French public agencies. *Archival Science*, 2025, 25 (4), pp.36. <10.1007/s10502-025-09511-y>. <hal-05214379>

HAL Id: hal-05214379

<https://hal.science/hal-05214379v1>

Submitted on 1 Mar 2026

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

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



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



Beyond capstone: toward a new strategy for appraising and selecting emails to transfer to archives within French public agencies

Edgar Lejeune^{1,2} · Bénédicte Grailles¹ · Touria Aït el Mekki³ · Patrice Marcilloux¹

Accepted: 25 July 2025 / Published online: 18 August 2025
© The Author(s) 2025

Abstract

This article introduces two new approaches to email archiving within French public agencies undertaken between 2020 and 2023 at the University of Angers (France). It shows how a qualitative survey of producers' practices and an NLP-based tool might provide the basis for new strategies for the appraisal and selection of email boxes and messages to be transferred to archives. The initial impetus was to overcome difficulties faced by French administrative records managers and archivists in locally implementing the appraisal framework known as the “Capstone” approach, which recommends gathering emails as electronic records and preserving for the long term only specific mailboxes selected in light of the function or position of the email account owner. While Capstone was adopted at the national scale in the French administration, major archival issues remain. Does an approach that focuses on the mailboxes of individuals holding strategic job positions correspond to the reality of bureaucratic work in France? And how should one define accountable criteria for identifying and appraising emails or mailboxes to be preserved on the basis of their enduring value? To address this, we combined two sociological methods, interviews and observation, and developed an interface for classifying messages using NLP. The combination of these two yielded some crucial elements to help archivists shift from the level of a single mailbox to that of a network: a better understanding of the practices of email producers, new criteria for selecting relevant mailboxes, and a new tool for archivists to appraise email archives.

Keywords Email · User behavior · Archival appraisal · Selection · Machine learning · Classification

Introduction

This article introduces two new approaches to email archiving within French public agencies undertaken between 2020 and 2023 at the University of Angers (France). It shows how a qualitative survey that focused on producers' practices along with the development of an NLP-based tool might help develop new strategies for the appraisal and selection of emails boxes and messages to be transferred to archives. The initial impetus was to overcome difficulties faced by administrative records managers and archivists in locally implementing the appraisal framework known as the "Capstone" approach, which recommends gathering emails as electronic records, and preserving for the long term only specific mailboxes selected in light of the function or position of the email account owner (NARA 2013).

Capstone was adopted at the national scale in the French administration following the work of the interdepartmental program Vitam, with the aim of managing the volume of emails to be archived in such contexts, which can approach several hundred or even thousands of mailboxes (Ville 2023; Vitam 2013). However, major archival issues remain. First, does an approach that focuses on the mailboxes of individuals holding strategic job positions correspond to the reality of bureaucratic work in France? And second, how should one define accountable criteria for identifying and appraising emails or mailboxes to be preserved on the basis of their enduring value? Last, how can these choices be both verified and justified?

In our opinion, the difficulties faced by archivists and public record managers lie in a lack of empirical knowledge regarding the ways in which public agents produce and use emails to communicate and make decisions. How do agents use their electronic messaging services? What is the role of email in their professional communications? What value do producers attribute to this kind of document within processes of public decision-making and information publication? What logics of classification are implemented?

To address this, we conducted a sociological survey that combined interviews and observation and developed an interface for examining and classifying messages using NLP methods adapted to the French language (Aït el Mekki et al. 2022; Grailles et al. 2022). The overall goal was to shift from the level of individual mailboxes to that of a network, in order to make sense of a volume of data too great to be comprehended at the human level of an archivist. We also sought to demonstrate the significance of recent research in the fields of personal information management, digital communication, and the sociology of work to discussions within archival science on this subject. The combination of these two approaches has yielded some crucial elements to assist public record managers and archivists in identifying and appraising emails or mailboxes to be preserved on the basis of their enduring value: a better understanding of the practices of email producers; new criteria for selecting relevant mailboxes (based on the identification of personal management profiles); and a new interface that enables archivists to access and appraise e-mail records on the basis of the communication networks to which they belong.

State of the art and literature review

The last 10 years have seen significant transformations in email archiving in France, in line with a more global trend. We first discuss the major developments in appraisal frameworks and tools that originated outside France beginning in the early 1990s. Next, we focus on email archiving practices currently in use in France and the difficulties faced by administrative records managers and archivists in implementing the appraisal framework known as “Capstone.” Finally, we address how research on behavior and management of personal information might help in better understanding how public agents record and manage emails, and thus, how such research could be used to improve the selection and appraisal of mailboxes and messages.

Conceptual approaches, appraisal challenges, and technical solutions

In a recent article, Johnston et al. (2019) highlight the shift from a “print and file” strategy to an electronic-first email policy in the US federal administration between 1990 and 2010. Following early litigation on the archival status of email records—which as early as 1993 were no longer considered deletable “convenience copies” by local courts¹—they show how various strategies were implemented to respond to the legal requirement to archive emails in their electronic forms. In the early 1990s, NARA issued rules that made “agencies responsible for providing training to staff to distinguish records from non-records” (Johnston et al. 2019, p 6). However, it soon became obvious that “the technology and practice capabilities for managing email in electronic form were just not feasible,” and that only a “print and file” approach could be implemented (Johnston et al. 2019, p 8). In 2003, the latter approach was nonetheless considered a failure following public exposure such as the “lost” Bush emails, when up to 22 million emails were lost due to insufficient records management software (Burleigh 2016).

At the same time, the responsibility for email records selection and appraisal fell into the hands of professional archivists, and discussions of email preservation therefore became part of broader research on digital record preservation. In 2003, a decisive report was published by the Dutch National Archives, proposing considering email as pivotal example for defining what digital records are (Dutch National Archives 2003, p 6). In 2006, the Digital Curation Center of the UK proposed establishing a common framework for managing emails as archives (Pennock 2006). The Center emphasized the need to consider the entire life cycle of an email, and thus, all the actors involved in the life of electronic messages, with the aim of ensuring the authenticity and integrity of documents which, given their digital nature, easily “mutate” (Pennock 2006, p 9).

However, no clear strategy for appraising email records was proposed, other than ideally selecting them “at the source and at the point of receipt or creation” (Pennock 2006, p 18). In 2013, NARA promulgated a new policy, known as the

¹ Johnston et al. (2019, p 6) described the District Court’s reasoning: “paper printouts were not exact copies of the originals as they did not reproduce transmission and receipt information that existed in electronic versions.”.

“Capstone approach,” which would turn out to be absolutely pivotal (NARA 2013). It aimed to provide “an alternative after the failure of the ‘print and file’ approach, while also addressing the labor required of users to distinguish record from non-record material” (Johnston et al. 2019, p 13). The Capstone approach proposed shifting away from the time-consuming appraisal method, which required archivists to weigh the value of each message using records schedules or guidelines. Instead, it suggested identifying functional roles within a given organization and capturing complete email accounts—as electronic records rather than as printouts. In doing so, the metadata and context of messages were more likely to be preserved, and the burden of evaluating each message would no longer fall on archivists.

Several projects dedicated to email preservation were launched in the following years, in accordance with both Pennock’s conclusions and NARA recommendations. In 2018, the Mellon Foundation commissioned a report focusing on email archiving that addressed both their significance for historical study and methods of collection, appraisal, conservation, and communication (Prom et al. 2018). Numerous projects were identified: Archivemata, DArcMail, EAS, ePADD, Preservica Standard Edition, TOMES Tool, Archival Email Symposium, Carcanet Press Email Preservation Project, CERP, Kaine Email Project@LVA, MeMail et PeDALS. This report showcased the diversity of types of email collections to be preserved, ranging from small collections—that of a writer for example—to massive sets composed of thousands of mailboxes in many public institutions. It also brought to light the various challenges faced by archivists in the appraisal and selection of emails to be transferred to archives (see also Cocciolo 2016).

The report states that the Capstone guidelines were relevant for handling high-volume email archives within academic and government institutions (Prom et al. 2018, p 16). It also points to the various ways in which the guidelines might be implemented, and the technical and organizational challenges faced by archivists and administrative records managers in doing so. On the one hand, the report attests that in many cases, the sheer volume of mailboxes to be appraised precludes simply archiving accounts from the top of the hierarchy, due to storage or workforce limitations. On the other hand, appraisal at the message level remains mandatory in many cases, because of the need to identify—and delete—sensitive or private information within messages, as well as spams and duplicates.

Several institutions worldwide tried to address these issues by developing pre-acquisition appraisal tools. For example, the open-source software ePADD from Stanford University aims to facilitate the processing, examination, appraisal, and communication of archival corpora (Schneider 2016; Baker et al. 2012). The Review, Appraisal and Triage of Mail Project (RATOM) launched by the University of North Carolina draws on the volume of messages produced to automate the tasks of selecting and appraising emails, using methods taken from artificial intelligence. Both used the following suite of processes: implementation of named entity recognition; navigation and visualization of metadata; modules for appraising and classifying that include searches based on recurring expressions. However, ePADD and RATOM have primarily been used with special collections, often linked to a single author—Salman Rushdie, Ian McEwan, Wendy Cope, etc. (Schneider 2019). The appraisal issues faced in such cases are different from those confronting records

managers and archivists working at the level of an institution, where legal requirements compel them to define accountable criteria for appraising and selecting emails in light of both their enduring and probative values. Moreover, RATOM and ePADD were designed to process English language content, reducing their potential applications outside the English-speaking world.

The French context

Today, email is the primary channel of communication used by French public officials and is considered by many agents to be central to the organization of their professional tasks. Ten years ago, sociologists studying the context of territorial civil services showed that “channeling the flow” of messages had already become a structuring activity in decision-making (Bretesché et al. 2012; 2014). They also showed that the use of email leads to extreme fragility in the conservation of data and information, making it necessary for public servants to deploy diverse strategies to ensure traceability (Bretesché et al. 2018).

These conclusions reflect the situation faced by archivists in France, where, just as in the USA, the strategic value of emails and the growing volume of records to archive imply a radical transformation in appraisal practices and norms.² French archivists began examining these questions in the late 2000s within the business world and under the pressure of business needs, with a focus on the question of evidence (Ballet and Rietsch 2008). Just as Johnston et al. show in their 2019 article regarding the US context, this movement was initiated following litigations. In France, since 2000, electronic documents have held the same weight in court as paper documents, and in 2004 electronic mail was given a legal definition and the concept of a digital original was absorbed in 2005.³

In 2011, the project team Vitam (Valeurs immatérielles transmises aux archives pour mémoire; Immaterial Values Transmitted to Archives for Memory) was entrusted with the task of reflecting on the digital preservation of material produced by state administrative agencies, and it validated the Capstone (2013) approach (Vitam 2013; Ville 2023). The study we undertook in 2020 confirmed the deployment of a strategy focusing on the emails of key individuals in political and administrative decision-making (cabinet ministers, directors, deputy directors) at the national level (Lejeune 2022). The Ministry for Europe and Foreign Affairs has, for example, targeted deputy directors and assistant directors, identifying 500 email accounts among the 15,000 it manages (Texier 2020). The Mission Archives of the Prime Minister’s Office focused instead on direct collaborators and “the most senior civil servants,” while also considering “strategic emails to be archived” (Ville 2023, p 3; Mission des archives auprès du Premier Ministre 2018).⁴ In contrast, regional archives departments have gathered few emails or none at all, due to a lack of clear

² At the Ministry for Europe and Foreign Affairs, email records now constitute “more than 40% of the stock preserved in their digital repository” (Ville 2023, p 6).

³ Law of March 13, 2000; law no. 2004–575 of June 21, 2004 regarding trust in the digital economy; and the ordonnance of June 16, 2005.

⁴ A similar strategy was implemented at the Council of State (see Meynaud 2021).

appraisal policies and tools adapted to the French language that can provide access to concepts and composition styles particular to French.

Other than these local endeavors, no in-depth evaluation of the appropriateness of the Capstone appraisal framework to French public agencies was conducted, either within the Vitam project, which focused instead on acquisition, preservation and access, or elsewhere. The two programs conducted at Angers between 2020 and 2023 aimed at undertaking such an evaluation at a theoretical and practical level. The hypothesis that the enduring value of emails derives from their position in decisional processes orients preservation toward certain email accounts at the top of the organizational chart (even if Capstone does allow for ad hoc designation of particular mailboxes outside of the organizational hierarchy). This is based on several presuppositions: that all strategic information flows through these email accounts, due to the hierarchical positions of the individuals they belong to; that assets that have value other than informational need not be documented; and that individual ways of managing information are relatively homogeneous. However, in contrast to these assumptions, administrative records managers and archivists assert their difficulties in selecting a manageable number of relevant mailboxes and accounting for the rationale behind their choices.

Records managers and archivists also face difficulties linked to the specificity of the archival object in question. Email accounts are highly heterogeneous due to the heterogeneity of users' behaviors and practices. This heterogeneity may stem from the informational content they contain: emails that may or may not have been sent as copies; forwarded emails; email loops; personal and/or functional addresses, attachments, mailing lists, calendars, etc. From an archival point of view, heterogeneity is further increased by the range of content the messages themselves contain, which can include various forms of discourse existing side-by-side (professional, personal, informal, structured, unstructured, etc.). In dealing with such heterogeneity, a final hurdle is the lack of tools adapted to the French language, since the supply of libraries or pre-trained NLP models varies from one language to another. This means that internal selection is significantly hindered, analysis of a mailbox impossible, and archival description limited. In this regard, examination of the research instruments offered by the French National Archives is telling: 130 mailboxes were indexed in 2022, predominantly using the "content" format, which provides only the name and job position of the producer (Naud 2022).⁵

Personal management of information

The issue of heterogeneity calls for thorough attention to studies in personal management of information. In recent years, numerous studies of professional email practices have been conducted outside of the field of archival science, guided by the idea that email is paradigmatic of technology that is appropriated and repurposed by its users (Tran 2010). Whereas it was designed for asynchronous and

⁵ The National Archives is currently experimenting with various solutions for accessing and giving access to email accounts and messages, including EPADD+ +, RATOM tools, and the Pêle-mél interface that came out of our research in Angers.

predominantly downward communication, new uses have emerged: delegation of professional tasks, follow-up on these tasks, personal archives storage, synchronous communication in the form of conversations, transfer of documents between computers, reminder emails sent to oneself. Email's multiformity, combined with its omnipresence across professional activities, has led it to be identified as a virtual space of work and socialization (Ducheneaut and Bellotti 2001) and a site where the boundaries between the public and the private blur, as evidenced by the transfer of emails from personal accounts to professional accounts (Capra et al. 2013).

Tran (2010) emphasizes the three functions of email: synchronous communication, coordination, and "organizational memory." Indeed, each individual user organizes a more or less enduring system of knowledge that is adapted to their own style and reflects how they both use and retrieve information (Messaoui 2020). Studies of personal information management have identified a number of constants. Malone (1983) highlighted two major organizational modes, which he connects to two different cognitive functions he refers to as "refinding" and "reminding." "Refinding" is accompanied by the creation of files labeled and organized according to a plan that makes it possible to identify where information is kept, while "reminding" makes it possible to prioritize matters that need to be handled while providing a reminder of what remains to be done and ensuring that frequently used documents are readily available. These two modalities are not in opposition to one another, and the same person might turn to either or blend the two attitudes depending on the information they are responsible for, the tools they have, the type of information being processed, as well as their experience (Carlson and Zmud 1999). It has also been noted that users navigate three strata, three "ages" or temporalities of information (Barreau and Nardi 1995): ephemeral information, working information, and archived information. This observation from the digital realm corresponds to categories proposed by Cole (1982) for information on paper: "action information," "personal work files," and "archived information."

Many authors have sought to categorize user behaviors more specifically within electronic messaging. Alberts (2013) distinguishes between users who delete emails once a task has been completed, those who preserve them in their mailboxes to track an activity, and those who save them on a hard drive or print them in order to better benefit from the information they contain. Three other types have also been identified:

- 1) "filers," who keep items organized in folders—whether organized by sender, organization, project, or personal significance (Ducheneaut and Bellotti 2001);
- 2) "pilers," who allow them to accumulate in their inbox; and
- 3) "spring-cleaners," who periodically delete items, and have been divided into "folderless cleaners" and "folderless spring-cleaners" (Bälter 1997).

Other behaviors have also been described: Mackay (1988) distinguishes between "prioritizers," who focus on incoming emails, and "archivers," who anticipate future uses by saving and organizing emails. The action of filing/no filing has been

analyzed in terms of “no filers” and “frequent filers” (Whittaker and Sidner 1996). Approaches based on the intensity of storing have been proposed, as well as cultural models—Europeans tend to organize more than Americans (Tang et al. 2014)—and approaches based on job responsibilities. The consensus seems to be that an individual’s way of managing their work email is similar to how they organize their personal email (Capra et al. 2013), and that individual differences remain relevant (Dabbish et al. 2005).

Research has also been conducted regarding the consequences of “workplace telepressure” and management on user behavior (Kefi et al. 2021). The trend toward synchronous use has been identified as imposing “email reactivity” (Paczkowski and Kuruzovich 2016). Alberts (2013) points out that one source of difficulty linked to email use stems from a lack of protocols for digital conservation, which generates a feeling of insecurity, a fear of creating false documents, and a desire to protect oneself. Such professional anxiety leads to what has been labeled “keeping behavior”: greater conservation of emails in professional email accounts than in personal ones, and greater efforts at organization and saving (Capra et al. 2013).

The diversity of human interactions with electronic artifacts has also been analyzed at the level of a community or network, providing great methodological insights for archival sciences. The analysis of email genres underlines the importance of considering messages as contents “anchored in a context” and shared among a community bound by use and discourse (Gagnon-Arguin et al. 2015). It implies socially situated practices: each community generates a system of genres, and each message might then be appraised in light of these practices. Network analysis has been used to explore such interactions on the basis of the contribution statistics of each contact within a network (Karagiannis and Vojnovic 2009), by focusing on the objective of emails sent—to inform, investigate, or plan (Lockerd and Selker 2003)—or by automatic extraction of social hierarchies (Creamer 2009).

Methods combining social science and algorithms

This article draws on two projects carried out at the University of Angers with the aim of helping archivists with the appraisal and selection of massive collections of emails. The first project, CARAMèls (2021–2022), explored the interplay between the practices of email producers and the difficulties faced by archivists in choosing appropriate mailboxes for long-term preservation. It provided an overview of the field of social science literature on professional email practices and gathered qualitative data from interviews with email users and producers in various national and regional agencies, as well as from direct observation of mailboxes. The second project, Pêlè-mél, tested various strategies and methods based on natural language processing (NLP) and artificial neural network (ANN) between 2020 and 2023. It culminated in the development of two prototypes that can be used to classify messages and to mine their content (Grailles et al. 2022). Here, we draw on the results of this project as applied to the issue of appraisal rather than access—that is, we consider the uses of the two prototypes as post-acquisition appraisal tools. We show how the analysis of two mailboxes belonging to two ministerial advisors—the

minister's closest collaborators—provides meaningful insights into the communication network of the Minister's Office and its relations with other mailboxes outside this network.⁶

Investigating producer practices

In CARAMèls, the profiles targeted were not just those recommended by the Capstone approach (directors and deputy directors). We decided in addition to go down the hierarchy by including departments heads and members of ministerial staff. These meetings led to eventually carrying out interviews with civil servants holding posts at an even lower level within the hierarchy, such as an executive assistant and an inspector of social affairs. Another originality of the survey lay in gathering interviews and observation from both national and regional departments.

The profiles selected were the following: directors, deputy directors, and department heads of national departments currently at their posts; directors, deputy directors, and department heads of regional departments currently at their posts; and members of ministerial staff currently at their posts or having recently left. There were three phases of the study. First, we conducted seven exploratory interviews with archivists with the aim of identifying the practical difficulties they face in appraising email from public agencies at the national and regional level. Then, we carried out twelve semi-structured interviews within national departments (5), territorial communities and decentralized departments (6), and Ministerial offices (1). Finally, we took advantage of these interviews to observe the mailboxes of certain key informants.

Among these twelve interviews, it should be noted that multiple agents holding similar job positions were interviewed: inspectors (2), members of regional leadership (3), members of a *conseil départemental* (departmental council) (2), agents involved in budget laws (2), as well as: a mayor (1), a secretary at the Cour des Comptes (Court of Auditors) (1) and a ministerial staff member (1). This helped in analyzing the interplay of professional norms and idiosyncratic or local practices at the level of a given department. The sample was also well balanced in terms of gender (5 women and 7 men), age (average age 46.8, range 31–68 years old), and length of career in public service (average 22.9 years, range 6–43 years).

Interviews were semi-structured and focused on: a) the technical environment within which agents work (laptop, personnel, software, etc.); b) the flow and nature of electronic messages (number of emails per day, number of interlocutors per day, type of interlocutors); c) strategies for managing the flow of these messages (time spent cleaning and organizing one's mailbox, motivations behind such practices); and d) practices of storing, archiving, and deleting emails (see Lejeune 2022). Interviews also provided an opportunity to gather documents shedding light on certain individuals' professional profiles (CVs), the regulatory context surrounding their

⁶ These emails dated from 2007–2011 and had been converted to .eml format using the MailExtract library developed as part of the VITAM program. The folder hierarchy was preserved. Each level has an XML manifest that corresponds to the French standard SEDA (Standard of Exchange of Data for Archiving). The export file is in .eml format.

practices (charts, regulations), and their methods of classifying and archiving emails (classification trees).

Analyzing electronic mailboxes

By combining NLP, metadata, and observation of classification practices (trees), the Pêle-mél project produced a prototype for exploring and visualizing acquired email accounts, using extracted messages in.eml format structured as a database. It aimed to facilitate access to acquired emails and to offer an interface for archivists to better understand a given email network. One original aspect of our project is that it sought to understand and analyze groups of mailboxes, with the intention of facilitating comprehension of an entire electronic mailbox network. It was thus a matter of describing collective, and not only individual, practices. Its application to two mailboxes belonging to ministerial advisors brought to light several elements for better understanding email practices.

Four main functionalities were implemented. First, the prototype offers a simple graphic interface for retrieving, exploring, and displaying acquired mailboxes and messages using a search engine (Fig. 1). An advanced search interface gives access to further functionalities, such as annotation tools and modification of the databases (Grailles et al. 2022). Next, the prototype provides the user with basic quantitative information about the acquired accounts, e.g., number of messages (received and sent), number of words per message, number of attachments, number of correspondents, etc. It allows emails to be sorted using metadata, and indicators such as the internal and external degrees of traffic to be computed. The interface prototype also offers visualizations using customizable graphs and, for networks of correspondents, dynamic graphs, generated using Pyvis (Fig. 2). Address and date filtering as well as search functions can be used to refine the graph, which can be exported as an image.

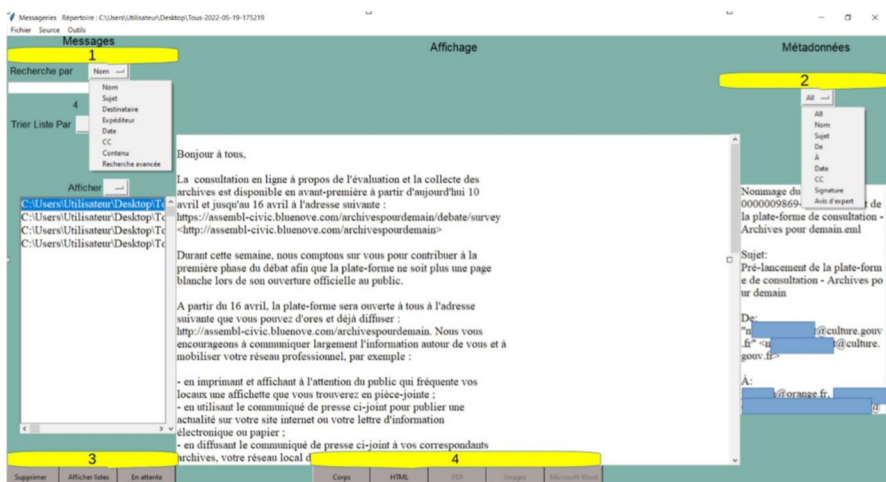


Fig. 1 Pêle-mél interface (main page)

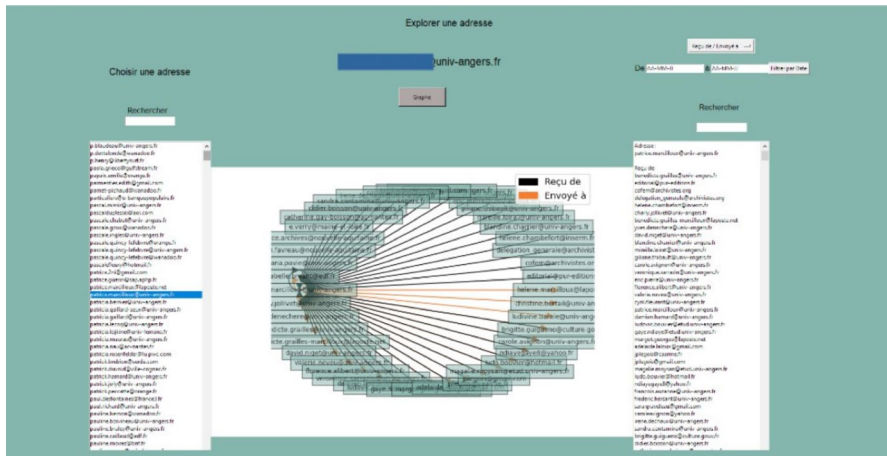


Fig. 2 Pêle-mél interface (visualizations module)

Finally, the project deployed an original classification function, which relies on named entities and term extraction and seeks to determine the semantic relations between them.⁷ Classification took place in four stages: a) extraction and creation of a list of terms and named entities; b) structuring these lists by using a linguistic approach to link terms and named entities according to semantic relationships; c) replacement of each term with a term cloud, using machine learning techniques; and d) classification of messages using the term clouds and a learning method. Archivists and records managers were asked to establish a list of themes corresponding to various domains of the ministry’s activities during the time in question. This list was based on their knowledge and methods of analyzing document flow and was structured around the organization’s missions, remits, actions, and programs. The semantic relationships between the themes identified and lexical entities—that is, terms⁸ and named entities extracted from the corpus beforehand—were detected. Grammatical categories, morphosyntactic information, and lemmatization were also labeled beforehand.

The goal was to create categories of terms surrounding each theme. The approach used here differs from previous approaches, notably Decker’s (2022), in its intention of defining a theme by a term cloud—that is, a set of terms in direct or indirect relation to a parent term. Our objective was to generate term clouds with strong similarity in order to then use them to classify messages. To establish relationships,

⁷ Results: before human supervision, 92% of extracted terms were correct; for named entities, 94% of personal names and 72% of organizations were correct. Almost all similar terms extracted were correct. To extract entities, we mainly used the SpaCy library [<https://spacy.io>] and for terms, our programmer Chafik Akmuouche adapted a PLDAC extractor, developed by Antoine Cadiou [<https://antocad.github.io/PLDAC/>].

⁸ A term is a lexical unit made up of one or more words, which expresses a specific concept or notion. For example: “voluntary termination of a pregnancy”; “sexual violence”.

two methods were combined: lexical-syntactic patterns⁹ (111 general patterns were created) and word embeddings. For the latter, words were represented in their context, capturing semantic and syntactic similarities. The word embedding model used (Word2Vec) represents words as vectors.¹⁰ Words which appear in similar contexts are mapped to nearby vectors, measured by cosine similarity. The corpus was processed using a model pre-trained in French.¹¹

The second phase of classification focused on messages. The base unit for this classification was not the message alone, but rather a message along with its attachments and file name; the entirety was represented by a vector. A similarity calculation was then performed in comparison with the vectors derived from the term clouds. To do this, we chose to use clustering with Doc2Vec, a word2vec extension for learning document embeddings.¹²

Initial results: from analysis to categorization to appraisal

By combining interviews, observations of mailboxes, and data produced by computational terminology and classification, we now have a nuanced understanding of the practices of email producers within French administrative agencies. We are therefore in a position to propose new strategies for categorizing, appraising and selecting mailboxes.

The heterogeneity of producers' practices

Mailboxes at the heart of continuity in administrative production

Our research confirmed the fundamental importance of electronic mailboxes to the operation of public administrative organizations. The two ministerial staff advisors' mailboxes analyzed between 2007 and 2011 as part of the Pêle-mél project, already demonstrated high activity in terms of emails sent, in keeping with the aforementioned sociological research (Bretesché et al. 2018). The results of the survey conducted within the CARAMèls project attest that emails are today the first and foremost channel of communication within public agencies, and a pivotal resource in organizing tasks and decision-making. Both projects concluded that electronic mailboxes act as tools for memory, proof, validation, and thus, continuity within

⁹ A lexical-syntactic pattern is an automaton to extract relationships. For example: the pattern Term1 + be + det + Term2 identifies the hypernymic relationship between “AIDS” and “incurable disease” in the sentence “Aids is an incurable disease.”

¹⁰ We also tested CamemBERT, a model derived from BERT (Bidirectional Encoder Representations from Transformers) adapted to French (see <https://camembert-model.fr/>). The results from Word2Vec were cleaner (Documentation officielle de Gensim—Documentation officielle de Word2Vec. <https://radimrehurek.com/gensim/models/word2vec.html>).

¹¹ Official documentation available at <https://radimrehurek.com/gensim/models/doc2vec.html>. We used a pre-trained French model developed by Jean-Philippe Fauconnier [<http://fauconnier.github.io>].

¹² The same library was used. See the documentation at <https://radimrehurek.com/gensim/models/doc2vec.html>. And the model for French developed by Fauconnier [<http://fauconnier.github.io>].

administrative work. These essential functions apply at both the collective and individual levels.

One indicator of the centrality of email to administrative functioning is the volume of emails received daily, which fluctuates between 50 and 200 according to interviewees. The variance between these extremes is difficult to explain, but the disproportion between the number of emails received and the number sent should be noted: the latter fluctuates between 20 and 80. A striking observation was that email took precedence over telephone conversations for six of the twelve interviewees, four of whom stated that they use email even more than face-to-face conversations. This predominance of email over oral exchanges is primarily found in departments whose missions take place outside of their ministries—for example, inspection, oversight, and auditing missions, and can be explained by the fact that members of these departments do not share a workplace. However, other respondents also testify to this prevalence of email and observe that it has greatly increased since the 2020 COVID pandemic, due to an increase in remote work.

A second indicator points to the essentiality of email to civil servants' work: time spent on emails. Within a typical workday, the interviewees spent at least 40% of their time responding to emails, and almost half of them stated that they do so continuously over the course of the day. For those whose jobs primarily involve information, coordination, and management, handling email represents the bulk of their work activities. In addition to being the predominant mode of communication among civil servants, email messages and mailboxes are also the main tools used by both individuals and departments to organize work. With their task management tools, address books, calendars, etc., mailboxes act as platforms for managing information and hierarchically organizing tasks.

At the collective level, emails now perform the functions previously carried out by paper mail and letter files. Today, they serve to “formalize an order” and to “call to action”—in short, to ensure the division of administrative labor. For this reason, practices of copying and forwarding emails are often considered by civil servants to be fundamental to the smooth functioning of a department, and the issue of copies for information versus copies for action is considered key. This aspect of email use becomes even more striking when we observe that emails addressed to multiple recipients are far more common (around 60% of emails according to interviewees; 46% of the messages received in the second mailbox analyzed in *Pêle-mél*). The use of email to put agents into action must be done in keeping with the traditional administrative chain of command. Thus, the vast majority of civil servants interviewed state that they directly contact the individuals in question for a given order, while emphasizing that they always send information copies to their superiors.

Such practices may take various technical and communicational forms—for example, mailing lists and shared functional addresses, which act as tools for organizing the dissemination of information and the delegation of tasks. For the mailboxes studied as part of the *Pêle-mél* project, there were two functional addresses and 30 mailing lists linked to a single office.¹³ For the most part, these addresses

¹³ These can be detected through regular expressions (regex), a sequence of characters that specifies a match pattern in a text.

correspond to the operational organization of the office (basically, its divisions), but also indicate needs that cut across divisions (mailing lists monitoring the press and the minister's speeches; management of confidential emails; information for advisors; special events). Email's official and easily reproducible nature makes it an absolutely essential type of document within administrative work, over and above traditional documents that might serve as proof of decisions. The interviews conducted during the CARAMèls projects nonetheless warn us of a rapid transformation in this regard, due to the growing use of cell phones. This is particularly the case among ministry staff, since urgent situations require decisions that are often made rapidly and ratified through text or instant messages (Signal, Telegram, WhatsApp). This opens up new and crucial challenges for archival scientists, archivists, and records managers.

Our study also showed that mailboxes act as essential tools for organizing individual work, since agents all use them as interfaces for managing orders they receive. Read and sent emails are selected, sorted, and classified as tasks are completed, and these practices emerge as structuring elements in the organization of administrative work at the individual level. One interviewee even stated that she would be able to keep working as long as she had access to her email, even if all her other work tools disappeared—files, paper, the office, her phone, etc. This is because mailboxes act as tools for memory, proof, and validation. We see this in civil servants' practices of following up on the matters, assignments, and orders for which they are responsible almost exclusively via their mailboxes and received messages. In some rare cases, following up on a matter entails making a request of certain colleagues—assistants or secretaries—who look in their own inboxes to retrieve the information. Some interviewees also indicated that they use self-archiving practices directly within their mailboxes for attachments or notes they intend to re-use, which was confirmed by network graphs generated with the Pêle-mél interface. Follow-up and self-archiving practices demonstrate ways in which mailboxes serve as agents' memory.

Emails also act as proof when a civil servant must give evidence of a decision made by themselves or a colleague. Email, by virtue of its written nature, makes decisions public and attests to their definitive nature. Thus, email makes matters official in a way that spoken words or text messages never can. One interviewee stated that use of email can even prevent breaches of chain of command—for example, requests made by higher-ups to act beyond of the scope of an individual's position. By copying an email to multiple individuals, a form of bureaucratic insecurity can be prevented. In terms of day-to-day matters, email is also a powerful tool ensuring continuity in administrative decisions. In settings such as budget departments, where staff frequently change positions, new hires are often given access to the mailboxes and emails of their predecessors so that they may benefit from their work. This practice reflects the fact that mailboxes do more than just ensure a single civil servant's memory: they are tools for departmental memory, independent of individuals, and thus constitute a veritable foundation for information and continuity.

The three points emphasized here—email as the predominant form of communication, as an organizational tool, and as a tool for continuity in administrative work—make a case for the profound transformation of the administrative production

in recent years. It also unfolds the difficulty in identifying relevant Capstone users without having empirical insights into producers' practices.

Identifying mailbox profiles

Our study also sought to analyze mailboxes and messages in order to help administrative archivists and records managers overcome certain difficulties they face. To this end, and contrary to the Capstone guidelines, which recommend considering the position of the account owner within a given hierarchy (while also allowing the selection of specific individuals beyond the top of the organization), we chose instead to consider mailbox profiles and behavioral profiles together. These profiles were identified and distinguished on the basis of metrics such as the volume of message received and sent, the type of "traffic" in a given mailbox, the proportion of mail with several recipients or copied message, as well as the job duties of the civil servants in charge of these mailboxes, the modes of classification they use, their professional trajectories, and are not mutually exclusive. This approach provides new criteria for considering which accounts should be selected and long-term preserved, and thus aims to help in implementing a more nuanced strategy than the one proposed by Capstone, which does not address characteristics of mailbox usage.

One category is that of mailboxes primarily used for information purposes. We include here mailboxes used for internal or external communication, which, because of the duties of the individuals in charge of them, contain a high volume of general information regarding a given department. So-called "functional" mailboxes, which, according to our interviewees are mainly used to receive messages from individuals outside the department, or act as a kind of internal mailing list, also belong to this category. In both cases, their role is to collect and redistribute information. The mailbox belonging to the press relations agent gathered as part of the *Pêle-mél* project also falls under this category, as evidenced by the large number of messages sent externally and by the extremely wide network of senders and recipients. Here, the term "functional" is not used in the same way as in the NARA guidelines, as it does not refer to the function of the mailbox holder, but rather to the function of the mailbox itself.

The second category covers more operational mailboxes, which are to be defined by two features. They are owned by civil servants with roles pertaining to the day-to-day work of administrative bodies: oversight, analysis, inspection, expertise, auditing, etc. (6 out of 12 interviewees). The content of their mailboxes reveals the behind-the-scenes of administrative work: working drafts of documents, technical discussions, and conversational, synchronous back-and-forth exchanges. The second mailbox analyzed as part of the *Pêle-mél* project also highlights another feature of this kind of mailbox: most of the conversations within it are internal to the department or division, and there is more internal traffic than external.

The third type of mailbox that can be identified is linked to coordination or management functions. Nine of the 12 interviewees had such roles within their departments or offices. Thus, their mailboxes send and receive a great deal of strategic information connected to decision-making, arbitration, and potential conflicts. If

Table 1 Typology of classification styles and folders

| Types of folders | Types of classification |
|------------------|---|
| Thematic | By function (management, task, institution, etc.) By task (budget, security, external policy, etc.) By type of document (newsletter, spam, etc.) By content (personal/professional) By response status (read, unread) |
| Chronological | By processing status (emails that have been addressed/those in process of being addressed) By date (year, month) |

three of these nine also perform tasks that we qualify as day-to-day, their roles as managers or coordinators contrasts with the bulk of the administrative work embodied by operational mailboxes. “Managerial” mailboxes thus offer richer and more heterogeneous content in comparison with the two previous types.

The second distinction we can observe has to do not with mailbox functions, roles or responsibilities but with behavioral profiles based on individuals’ practices of classifying and deleting. Two families of organization, which are sometimes combined, can be observed: thematic classification and chronological classification (see Table 1 below).

Two major elements should be highlighted here. The first, as Dabbish and Kraut (2005) note, has to do with the salience of individual differences and the difficulty of explaining the behaviors observed. The civil servants we interviewed each offered different reasons for their classification practices: professional training, experience within the organization, a personal desire for organizing one’s workspace. From a quantitative point of view, the vast majority of those interviewed opt for chronological classification, most often by month, due to lack of time to implement a truly functional thematic classification system. We also observe that the more “telepressure” there is, the more agents opt for chronological classification. Our research also revealed, however, that those who spend time on classification and internal sorting are more likely to be women trained at prestigious administrative training institutes (the *École nationale d’administration*, regional administrative institutes, etc.) who are sensitive to historical issues and who hold positions as office heads rather than as department directors or assistant directors.

The second noteworthy element overlaps with the typology established by Alberts in 2013, which can be specified to fit the case of civil servants. First, none of the subjects interviewed behaves entirely as a “piler”: all perform some kind of classification of their inboxes by moving emails into folders at one point or another. A very small minority can be considered “spring-cleaners,” for the practice of deleting messages is in reality quite rare, and when it does occur, it is often the result of necessity due to storage limits. Here again, hierarchical rank is significant, since directors and assistant directors have the privilege of being able to request increased storage, and thus do not delete anything.

The two mailboxes analyzed as part of the *Pêlè-mél* program make it possible to specify the extent of such classification, as the folder hierarchies for these mailboxes

were preserved. They show that the larger mailbox was partially organized into folders, which names, however, were mostly unprecise and prevent from contextualizing their content—i.e. names that contain acronyms or labels such as “various folders,” “supplementary elements.” One of the two mailboxes did contain a folder named “Notes to keep,” demonstrating a desire to create a personal space for preserving the memory of activities performed; the other mailbox contained a folder named “cab” (short for *cabinet*; “office”), demonstrating an attempt to separate operational information from that relating to the advisor’s activities.

Cross-referencing information gathered during interviews with quantitative analysis demonstrates the existence of three mailbox profiles: “functional,” “operational” and “managerial.” Research on the personal management of information shows the importance of individual behaviors on the internal organization of mailboxes and the structure of a given communication network. These two types of profiles are to be considered useful for archivists and records managers wishing to better identify hub-mailboxes, to filter messages, or to improve their description for communicative purposes.

Appraising differently in order to select differently

Selecting hub mailboxes

Appraisal of electronic mailboxes generally involves two levels of selection: certain mailboxes are preserved because the job responsibilities of their owners mean they are considered to be of strategic value; and selection also occurs within mailboxes. Our hypothesis is that, in contrast to current practices among French records managers and archivists, there exist hub-mailboxes within organizations that do not necessarily belong to the individuals with the greatest responsibilities. Indeed, the interviews we conducted showed that an individual’s strategic position does not necessarily reflect the centrality of their mailbox at the level of the network. Thus, an office head’s mailbox may turn out to be richer and more meaningful to preserve than that of their hierarchical superior.

Two elements appear crucial to us here: the number of emails sent/received and practices of copying. If we look at the first variable, our empirical results attest to no clear correlation with hierarchical position. Moreover, the agent who claimed to receive the highest number of emails in our interviews was an office head, that is, a position that was not targeted in current archivists’ practices. Thus, there appears to be no evidence that the more responsibility an individual holds the more emails they send or receive. However, we should also specify that interviews do not provide precise or accountable information, since it is reasonable to assume that the numbers reported by individuals in interviews are estimates rather than exact numerical realities.

In our opinion, copying practices yield more precise information regarding the centrality of a mailbox. Interviewees told us that the purpose of sending a copy of an email is to pass on an order—making it official while at the same time informing

one's colleagues and superiors. In departments where copying practices comply with the rules in force (which was the case for all our interviewees' departments), direct superiors are thus systematically copied when it is a matter of strategic information (decisions, important information, taking action, arbitration, etc.). Another element shows that it might make it advantageous to prioritize archiving the mailboxes of office heads as well as directors and assistant directors: they are all less likely to delete emails since they enjoy technical privileges in terms of storage space. This has implications for the rate at which records managers gather data, which must be higher when it is a matter of departments or individuals contending with more storage limitations. For all these reasons, we advise records managers and archivists to prioritize gathering the mailboxes of directors, assistant directors, as well as office heads, after ensuring that best practices in terms of sending copies for information and copies for action have been respected.

However, individual practices remain determinant, and each organization might present a different network configuration. Graphs that dynamically depict networks are thus indispensable for confirming an archivist's choice of mailboxes. The Pêlemél interface makes it possible to utilize metadata and to produce graphs using various indicators: number of recipients, number of senders, frequency of exchanges, concentration of associated mailing lists or functional addresses, variety of correspondents. To this list we may add, thanks to the thematic classification of messages performed by Word2Vec, the diversity of subjects addressed. Analysis and visualization of a network of mailboxes—either all mailboxes in the case of small organizations, or a portion of them in the case of larger ones—should make it possible to identify significant vertices or hubs, which may be of various types. A mailbox may be central, at the level either of the graph or of a sub-graph delimiting a service or a division. Or, it may be indispensable to the connectivity of the network. It also may be a transmission node, i.e. it makes it possible to accelerate circulation by reducing the length of a directed path.

The combination of empirical survey and visualization makes it possible to utilize the typology of profiles presented in the previous section. We have shown that mailboxes with management and/or coordination roles are in principle the richest in strategic information. We have also emphasized that certain behavioral profiles are to be considered differently. For example, the majority of interviewees transfer all their sent and received messages into a folder corresponding to the month. In this respect they may be considered "filers," but the different modalities of classification used make it possible to distinguish profiles in a more nuanced way. Interviews showed that three quarters of the interviewees are actually "prioritizers," since what counts above all for them is to address incoming emails. However, they all have elements of "archivers," since they all classify their emails into folders in the goal of more easily locating them for future use. Yet the rate at which this classification occurs varies widely from one person to another—some file incoming emails as soon as they are received, while others do so only at the end of the month. It is difficult to explain the reasons behind these behaviors, although the technical limits of storage space represent a significant variable.

These observations make it possible to qualify the relevance of the categories proposed by Alberts, McKay, and Bälter, for individuals' sorting and classification

practices all obey different temporalities. The most relevant element here is thus the rate of classification and the time agents dedicate to implementing and executing it. Such behaviors have an impact on data gathering, but once again we observe here that these behaviors are quite individualized. However, it is clear that the details of the missions of the departments under consideration must be taken into account when creating a gathering schedule—for example, emphasizing periods of low activity (i.e. December for budget laws), as well as longer timeframes, such as yearly or multiple years.

The major advantage of the approach we propose is that it positions archival appraisal at the level of a group of mailboxes rather than a single mailbox and thus corrects for the effects of individual behaviors. It also makes it possible to intervene at the level of the email server rather than the client interface. However, identifying hub-mailboxes does not solve all the problems posed by the archiving of emails, in particular the heterogeneity of message content and the issue of appraising at the message-level.

Filtering the content of a mailbox

In this last section, we would like to suggest to archivists and public record managers a few ideas for identifying and filtering important messages within a given mailbox, with the aim of enhancing appraisal and selection. The most common issue faced is eliminating messages without much content and those that are private. Based on both the CARAMèls and Pêle-mél research programs, we have shown that these might be indicated by “ping-pong” style exchanges (a rapid succession), or by an abnormally long series of messages (as defined by the average length of conversational exchanges multiplied by 2). In contrast, the mailing lists used to obtain implicit or explicit ratification, as well as messages with multiple recipients copied—all of which are indicators of the workflow of bureaucratic work—must be carefully preserved. The CARAMèls project confirmed that all our interviewees considered messages with recipients or senders copied to be particularly strategic for administrative functioning, and thus, demonstrate records importance. This complicates certain notions that short or seemingly informal emails should not necessarily be archived. The common case of a decision ratified by a simple “ok” is one example demonstrating the difficulty of filtering mailboxes by overly simplistic criteria such as number of words or lexicon used.

Visualizations produced with Pêle-mél also make it possible to identify practices of self-archiving (sending an email to oneself or to a different professional or personal address identifiable by a regular expression), rendered as a loop in the node. Such messages and their attachments should in principle be selected, for this practice suggests the mailbox owner’s particular interest and investment in the contents of these messages, although they might also be the result of other intentions—bypassing file copy limitations or storage space limits, or simplifying a transfer between two software programs that communicate poorly.

Another bias for filtering emails within a mailbox relies on the fact that users attempt to organize their mailboxes. For example, most of our interviewees created specific folders for personal messages. It might therefore seem tempting to make

selections based on the files they have created. However, interviews have shown the difficulties actors face in maintaining and sticking to their classification methods over time, both in terms of receiving and sending. For other type of messages, observation of mailboxes also highlighted the difficulty of interpreting the file names used, between acronyms, jargon, and shortened forms unique to an individual. Examination of the folder hierarchies of archived mailboxes showed that even when their owners had purged them of personal messages, some remained; it also became clear that particular attention should be paid to photographic attachments. Searches using regular expressions, in conjunction with the word embedding method, can help identify private messages as well as so-called “engaging” ones.¹⁴

Pêle-mél experiments also offer several leads toward better selection at the level of the message. The clustering performed using Doc2Vec makes it possible to identify and follow conversation topics, and above all to trace their evolution over time. This might help in considering the importance of a channel of messages based on the importance of the topic. Another approach developed by the Pêle-mél project was to assess the quality and complexity of writing, with the aim of filtering messages with more elaborate content. This was attempted by performing close analysis of text that included average number of words per sentence as well as part-of-speech tagging. However, the results are still scarce, and this way of accessing the content of messages requires a number of archived resources that must either be reconstituted or exist as open data: directories, organization charts, rules governing the naming of lists, individuals belonging to certain lists, definitions of acronyms, etc.

Conclusions

The two research programs presented here sought to shed light on new strategies for appraising and selecting email records to transfer to archives. The goal was to identify mailbox profiles and/or hub mailboxes whose archiving would be of long-term interest. We draw on the hypothesis that a better knowledge of email practices could enhance the implementation of the Capstone approach at a local level, by defining categories and indicators. Our study provides two types of results for archival scientists, records managers, and archivists. We emphasize the need to shift email appraisal from the level of the individual mailbox to the level of a network of mailboxes. In order to do so, we propose conceptual tools for better selecting and appraising mailboxes—such as profiles and categories— as well as technical resources: a model of a questionnaire and an interface to explore acquired email records.

Several paths forward were proposed to consider both the specificity of each organization and the heterogeneity of behaviors with respect to email communication. The CARAmèls project led to categorizing both mailboxes—as “functional,” “operational,” “managerial”—and individual classification behaviors—thematic

¹⁴ Here we should note that in France the use of professional email is governed by usage charts spelling out authorized and forbidden usages, although jurisprudence allows for some limited personal use of professional email.

and chronological. The combination of these two types of profiles shifts the appraisal framework proposed by Capstone by distinguishing the mailbox from its owner. In parallel, a model for a questionnaire study emerged out of our interview study and is open to reuse at the local level (Lejeune 2022).

The results of such studies should be combined with visualizations, part-of-speech tagging, term and named entities extraction, and classification, all of which are available with the Pêle-mél interface. Such functions are key for archivists to appraising mailboxes at the level of the network, an approach that requires the creation of decision-assistance tools to be of practical use. The data gathered from the field study and from mailboxes and emails can be fed back into research tools in order to access not only the content of messages but also their context. Although our NLP approach was used within a specialized environment (health, welfare), many elements were validated and can be reproduced within a different environment: named entity extractions, term extractions, acronym definitions, etc. In general, this approach enriches records managers and archivists' reflections and gives them new tools. Such decision-making aids can be combined with other entry points if so desired. The classification and visualization modules are still in the experimental phase and must be improved to facilitate usage. Various hypotheses must still be tested on larger corpora, both specialized and non-specialized. Nonetheless, and beyond appraisal and acquisition, this opens a path to improving access to digital archives.

Funding Open access funding provided by Université d'Angers. Angers Loire Métropole

Declarations

Conflict of interest There is no conflict of interest attached to this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ait El Mekki T, Grailles B, Randriatsitohaina T (2022) Création d'une base de connaissances à partir de messageries spécialisées pour améliorer l'exploitation et l'archivage des méls [Creation of a knowledge base from specialized messaging systems to improve email management and archiving]. In: Proc of the 16th Int Conf on Statistical Analysis of Textual Data, JADT 2022, Vadistat press and Editzioni Erranti, pp 52-59
- Alberts I (2013) Challenges of information system use by knowledge workers: the email productivity paradox. Proc Am Soc Inf Sci Technol 50(1):1–10. <https://doi.org/10.1002/meet.14505001089>

- Baker F, Butler P, Green B (2012) Carcanet press email preservation project. Research report of the Digital Preservation Strand Project. University of Manchester, United Kingdom. Accessed 24 July 2025 at: https://pure.manchester.ac.uk/ws/portalfiles/portal/33713849/FULL_TEXT.PDF
- Ballet P, Rietsch JM (2008) Conserver les courriers électroniques ? Ou comment résoudre la problématique de l'archivage des e-mails. [Should we record emails? Or how to solve the problem of email archiving]. FedISA Fédération. France, Paris
- Bälter O (1997) Strategies for organising mail. In: Thimbleby H, O'Conaill B, Thomas PJ (eds) *People and Comput XII: Proc of HCI'97*. Springer, pp 21–38
- Barreau D, Nardi BA (1995) Finding and reminding: file organization from the desktop. *ACM SigChi Bull* 27(3):39–43
- Bretesché S, de Corbière F, Geffroy B (2012) La messagerie électronique, principal métronome des activités de cadre. [Email, the main metronome of executive activities]. *La Nouv Rev Du Trav*. <https://doi.org/10.4000/nrt.262>
- Bretesché S, de Corbière F, Geffroy B (2014) Frameworks and messaging. *Réseaux* 187(5):135–162
- Bretesché S (2018) De l'apparente accélération à la permanence bureaucratique : l'usage du mail par les cadres territoriaux. [From apparent acceleration to bureaucratic permanence: the use of email by local government executives]. In: Aubert N (ed) @ la recherche du temps. *Individus hyperconnectés, société accélérée : tensions et transformations* [In search of time: Hyperconnected individuals, accelerated society: tensions and transformations]. érès, Toulouse, pp 351–370
- Burleigh N (2016) The George W. Bush White House 'lost' 22 million emails. *Newsweek* (12 September). <https://www.newsweek.com/2016/09/23/george-w-bush-white-house-lost-22-million-emails-497373.html>. Accessed 25 June 2025
- Capra R, Khanova J, Ramdeen S (2013) Work and personal e-mail use by university employees: PIM practices across domain boundaries. *J Am Soc Inf Sci Technol* 64(5):1029–1044. <https://doi.org/10.1002/asi.22815>
- Carlson JR, Zmud RW (1999) Channel expansion theory and the experiential nature of media richness perceptions. *Acad Manag J* 42(2):153–170
- Cocciolo A (2016) Email as cultural heritage resource: appraisal solutions from an art museum context. *Rec Manag J* 26(1):68–82. <https://doi.org/10.1108/RMJ-04-2015-0014>
- Cole I (1982) Human aspects of office filing: implications for the electronic office. *Proc Hum Factors Soc Annu Meet* 26:59–63. <https://doi.org/10.1177/154193128202600115>
- Creamer G, Rowe R, Hershkop S, Stolfo SJ (2009) Segmentation and automated social hierarchy detection through email network analysis. In: Zhang H et al (eds) *Advances in Web Mining and Web Usage Analysis: SNAKDD 2007*, 12–15 August 2007, San Jose, USA. Springer, New York, pp 40–58
- Dabbish LA, Kraut RE, Fussell S, Kiesler S (2005) Understanding email use: predicting action on a message. *Proc of the SIGCHI Conf on Hum Factors in Comput Syst CHI 2005*, 2–5 April 2005, Portland, USA. Association for Computing Machinery, New York, pp 691–700
- Decker S, Kirsch DA, Venkata SK, Nix A (2022) Finding light in dark archives: using AI to connect context and content in email. *AI SOC* 37(3):859–872. <https://doi.org/10.1007/s00146-021-01369-9>
- Ducheneaut N, Bellotti V (2001) E-mail as habitat: an exploration of embedded personal information management. *Interact* 8(5):30–38. <https://doi.org/10.1145/382899.383305>
- Dutch National Archives. (2003) The digitale bewaring Testbed: from digital volatility to digital permanence. *Preserving Email*. The Netherlands, The Hague. Accessed 25 July at: <https://pages.ischool.utexas.edu/galloway/spring2019/INF392K2019/volatility-permanence-email-en.pdf>
- Gagnon-Arguin L, Mas S, Maurel D (2015) Les genres de documents dans les organisations : analyse théorique et pratique. [Types of documents in organizations: theoretical and practical analysis.]. Presses de l'Université du Québec, Québec
- Grailles B, Aït El Mekki T, Vasseur E (2022) Improving the archiving and contextualization of electronic messaging in French. In: *Proc of the Int Conf on Digit Preserv iPres 2022*, 12–16 September 2022, Glasgow, Scotland. Digital Preservation Coalition, pp 374–77
- Johnston JA, Wallace DA, Punzalan RL (2019) Messages sent, and received? Changing perspectives and policies on US federal email as record and the limits of archival accountability. *Arch Sci* 19(4):309–329. <https://doi.org/10.1007/s10502-019-09318-8>
- Karagiannis T, Vojnovic M (2009) Behavioral profiles for advanced email features. *Proc of the 18th Int Conf on World Wide Web WWW' 09*, 20–24 April 2009, Madrid, Spain. Association for Computing Machinery, New York, pp 711–720

- Kefi H, Kalika M, Saidani N (2021) Dépendance au courrier électronique : effets sur le technostress et la surcharge informationnelle et répercussions sur la performance. [Email addiction: effects on technostress and information overload, and repercussions on performance]. *Syst D'inf Manag* 26(1):45–83
- Lejeune E (2022) Rapport final du projet CARAMèls (Comprendre les administrateurs et leurs relations à leurs mèls). [Final report of the project CARAMèls (Understanding administrators and their relationships with their emails)]. Université d'Angers & TEMOS - Temps, Mondes, Sociétés, France. Accessed 25 July 2025 at: <https://shs.hal.science/halshs-04546948>
- Lockerd A, Selker T (2003) DriftCatcher: the implicit social context of email. In: Rauterberg GWM, Menozzi M, Wesson J (eds) *Proc of the Hum-Comput Interact INTERACT' 03*, 1–5 September 2003, Zurich, Switzerland. IOS Press, Amsterdam, pp 813–816
- Mackay WE (1988) More than just a communication system: diversity in the use of electronic mail. *Proc of the 1988 ACM Conf on Computer-supporter Coop Work CSCW' 88*, 26–28 September 1988. Portland USA. Association for Computing Machinery, New York, pp 344–353
- Malone TW (1983) How do people organize their desks? Implications for the design of office information systems. *ACM Trans on Inf Syst (TOIS)* 1(1):99–112
- Messaoui A (2020) Pratiques de gestion personnelle de l'information (GPI) des enseignants. [Teachers' personal information management (PIM) practices]. *I2D-Inf. Données Documents* 3(3):149–166
- Meynaud V (2021) L'archivage numérique des courriels au Conseil d'État. Réflexions fonctionnelles et techniques. [Digital archiving of emails at the Council of State. Functional and technical considerations]. Master thesis, École Nationale des Chartes
- Mission des Archives auprès du Premier ministre (2018) La gestion et la collecte d'archives électroniques. FAQ à destination des référents-archives des Services du Premier ministre. [Management and collection of electronic archives. FAQ for archive administrators in the Prime Minister's Office]. France, Paris. Accessed 25 July 2025 at: <https://www.documentation-administrative.gouv.fr/adm-01858798v1/document>
- U.S. National Archives and Records Administration (2013) Guidance on a new approach to managing email records. *Bulletin* 2013–02, 29 August. <https://www.archives.gov/records-mgmt/bulletins/2013/2013-02.html>. Accessed 20 May 2018
- Naud D (2022) Retour d'expérience sur la stratégie de préservation des Archives nationales. [Feedback on the National Archives' preservation strategy]. *Modernisation Et Archives (blog)*. <https://doi.org/10.58079/u5yq>
- Paczkowski WF, Kuruzovich J (2016) Checking email in the bathroom: monitoring email responsiveness behavior in the workplace. *Am J Manage* 16(2):23
- Pennock M (2006) Curating e-mails; a life-cycle approach to the management and preservation of e-mail messages. *DCC Digital Curation Manual*, United Kingdom, University of Bath. Accessed 25 July 2025 at: <https://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/curating-e-mails/curating-e-mails.pdf>
- Prom C, Murray K, Baker F, Connelly M, Gogel W (2018) The future of email archives. A report from the task force on technical approaches for email archives. Council on Library and Information Resources, USA, Washington. Accessed 25 July 2025 at: <https://www.clir.org/wp-content/uploads/sites/6/2018/08/CLIR-pub175.pdf>
- Schneider J (2016) ePADD: supporting archival appraisal, processing, and research for e-mail collections. *Midwest Arch Conf Newsl* 43(3):170
- Schneider J, Adams C, DeBauche S, Echols R, McKean C, Moran J, Waugh D (2019) Appraising, processing, and providing access to email in contemporary literary archives. *Arch and Manuscr* 47(3):305–326. <https://doi.org/10.1080/01576895.2019.1622138>
- Tang G, Pei J, Luk WS (2014) Email mining: tasks, common techniques, and tools. *Knowl Inf Syst* 41(1):1–31. <https://doi.org/10.1007/s10115-013-0658-2>
- Texier B (2020) Archivage électronique : quand le Quai d'Orsay archive ses emails. [Electronic archiving: when the French Foreign Ministry archives its emails]. *Archimag*, 21 August 2020. <https://www.archimag.com/archives-patrimoine/2020/08/21/archivage-electronique-quai-orsay-archive-emails>. Accessed 25 Jul. 2025
- Tran S (2010) Quand les TIC réussissent trop bien dans les organisations : le cas du courrier électronique chez les managers. [When ICTs are too successful in organizations: the case of email among managers]. *Manag Avenir* 34(4):200–215

- Ville M (2023) 2013–2023: a review of ten years of email archiving in France. In: Proc of the Int Conf on Digit Preserv iPres 2023, 19–22 September 2023, Urbana-Champaign, USA. Digital Preservation Coalition, pp 19–24
- Vitam (2013) L'archivage des messageries électroniques. [Email archiving]. Proof of concept. Ministère de la Culture et de la Communication, Ministère de la Défense, Ministère des Affaires étrangères et du Développement international. France, Paris.
- Whittaker S, Sidner C (1996) Email overload: exploring personal information management of email. In: Proc of the SIGCHI Conf on Hum Factors in Comput Syst CHI '96, 13–18 April 1996, Vancouver, Canada. Association for Computing Machinery, New York, USA, pp. 276–83

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Edgar Lejeune is historian of science, and currently a post-doctoral researcher at the Chaire d'Excellence en Édition Numérique at the University of Rouen Normandie. In 2021, he received a PhD from the University Paris-Cité (France). He has been a postdoctoral researcher at TEMOS (University of Angers), at the Centre Alexandre-Koyré (CNRS, EHESS, MNHN), and a research fellow at the Max Planck Institute for the History of Science (Berlin, Germany), and at the Vossius Center (University of Amsterdam). His research aims to tell a history of the transformation of textual cultures at the digital age.

Bénédictte Grailles is a professor of archival science and head of a course specializing in records and archives management. Her research focuses on professional and spontaneous practices developed around archives, and on social archival studies. She is also interested in archival appraisal, particularly of electronic data and records.

Patrice Marcilloux graduate of the École nationale des chartes, was director of several archives service from 1993 to 2013, before becoming a professor at the University of Angers. His current research focuses on new forms of social demand for archives. He has notably published *Les ego-archives, traces documentaires et recherche de soi* (Presses universitaires de Rennes, 2013).

Authors and Affiliations

Edgar Lejeune^{1,2}  · Bénédictte Grailles¹  · Touria Aït el Mekki³ · Patrice Marcilloux¹

✉ Edgar Lejeune
edgarlejeune@hotmail.fr

Bénédictte Grailles
benedicte.grailles@univ-angers.fr

Touria Aït el Mekki
touria.aitemekki@univ-angers.fr

Patrice Marcilloux
patrice.marcilloux@univ-angers.fr

¹ Laboratoire Temps, Mondes, Sociétés (Temos) Unité Mixte de Recherche CNRS, Université d'Angers, Angers, France

² Laboratoire Sciences, Philosophie, Histoire (Sphere) Unité Mixte de Recherche CNRS, Université Paris-Cité, Paris, France

³ Laboratoire d'étude Et de Recherche en Informatique d'Angers (LERIA), Université d'Angers, Angers, France