Miscellanea Archivistica Studia 224

Internationales Archivsymposion/International Archives Symposium in Arnhem (2024)

Sind wir vorbereitet ? Zukunftsfähigkeit unserer Archive
Are we prepared ? Are our Archives future-proof ?

ANNALEN/PROCEEDINGS

Redaktion – Rédaction – Redactie
Els Herrebout

INTERNATIONALES ARCHIVSYMPOSION / INTERNATIONAL ARCHIVES SYMPOSIUM
IN ARNHEM (2024)

SIND WIR VORBEREITET? ZUKUNFTSFÄHIGKEIT UNSERER ARCHIVE
ARE WE PREPARED? ARE OUR ARCHIVES FUTURE-PROOF?

–

ANNALEN / PROCEEDINGS

ALGEMEEN RIJKSARCHIEF
EN
RIJKSARCHIEF IN DE PROVINCIËN

ARCHIVES GÉNÉRALES DU ROYAUME
ET
ARCHIVES DE L'ÉTAT DANS LES PROVINCES

GENERALSTAATSARCHIV
UND
STAATSARCHIVE IN DER PROVINZ

MISCELLANEA ARCHIVISTICA
STUDIA

224

ISBN : 978 94 6391 548 9

Generalstaatsarchiv – Algemeen Rijksarchief – Archives générales du Royaume
Ruisbroekstraat 2 rue de Ruysbroeck
1000 Brüssel – Brussel – Bruxelles

D/2025/531/024

Publ. 6565

Die vollständige Liste der Publikationen finden Sie im Internet (http://arch.arch.be)

De volledige lijst van onze publicaties kan U raadplegen op internet (http://arch.arch.be)

La liste complète de nos publications est consultable sur notre page électronique (http://arch.arch.be)

E-mail, the Capture Challenge
Vincent HOOLT and Geert LELOUP[1]

This article examines the different methods used for the archiving of e-mails. For decades different methods have been tested, mostly depending on the user to archive. Only recently has the focus shifted toward automation, taking the decision away from the user. But if we want to make automatic archiving easier, we must also reconsider how we appraise and select records. This can be done using the so-called Capstone method. As this method has become the new preferred method in the Netherlands a deeper analysis of its principles and consequences is a key focus of this article.

Importance of E-mail and E-mail archiving

In 1971, the first e-mail was sent by Ray Tomlinson, an engineer at BBN Technologies. He used the @ symbol to address E-mails to different computers within the ARPANET network of the United States Department of Defense. The following years the first E-mail standard, RFC, was developed. In 1982, the MIME protocol (Multipurpose Internet Mail Extensions) was introduced, allowing attachments to be sent with E-mails.

During the 1990's E-mail became popular among the general public and governmental organizations due to the increasing accessibility of the internet. It gradually replaced paper correspondence and fax in practice. From the 2010's onwards, innovations such as mobile e-mail and cloud solutions have made e-mail even more accessible and flexible. E-mail has dramatically transformed communication and continues to evolve. For quite some time, it was regarded as informal and unofficial communication. However, this perception changed in the 2000s. Today, e-mail is legally accepted and trusted as a form of communication The amount of e-mails send and received, especially in the workplace has grown exponentially during all this time. And although newer methods of communication like Whatsapp exist, e-mail still holds its place in the information landscape.

Governmental organizations in the Netherlands must comply with the Archival Law, which mandates the preservation of e-mails as they are considered official records. Archiving e-mail ensures a transparent and accountable record of all communications and tells the story behind the

[1] Drs. Vincent Hoolt and Dr. Geert Leloup, National Archives of the Netherlands.

scenes. This is essential for governmental organizations themselves, but also for maintaining public trust.

In the Netherlands there have been several court cases affirming that e-mail are indeed records under archival and public access laws (Both share the same definition). However, this has not yet changed how users see e-mail. Not as an official means of communication, but a personal, informal form, provided by the organization, but not subject to the same controls. Still, as this does not conform with the law, the archiving and preservation of e-mails is required.

This preservation started with printing e-mails to complete paper files. Considering the growth of e-mail, it is impossible to contemplate continue printing e-mails. Only the amount of e-mails send and received within Dutch national government is estimated to be about 5 billion a year. Who can imagine printing all those? Clearly other methods are required.

Methods of E-mail Archiving

There are no less than five methods to archive e-mail messages. Two of them manual and three of them automated. Each has its own advantages and disadvantages.

The first method is currently the most widespread: the manual selection and export of e-mails to a Document Management System or - less than ideal - a file server. This option means that individual employees select e-mails that they deem are relevant to work processes from their mailbox and save them in their organization's DMS or file server. Once exported, the e-mail is removed from their own account. The retention of the e-mail follows that of the file the e-mail has become part of. This method, however, is highly dependent on employee cooperation; it requires extra action from them and clear process agreements. Who is responsible for saving the e-mail? Is each e-mail saved separately or the whole thread after the conversation has ended? And of course, where is the e-mail stored to? Often e-mails are not just related to one file but several. If that is the case, what is then the correct file, or should the e-mail be duplicated? All complex questions that require agreements, and that all employees are aware of these agreements and know how to act.

One crucial question remains: what happens to e-mails that are not selected and exported and thus remain in the mailboxes? From different points of view – the organization, the citizen and heritage – this could lead to an uncontrolled loss or retention of e-mails.

This widespread method has been in use ever since printing seized to be an option but has not led to more preservation and management of e-mails by the organization.

The second method is the Capstone method, as originally conceived and developed by the National Archives and Records Administration (NARA) in the United States. This method categorizes e-mails based on the role or position of the e-mail account holder, rather than the content of each individual e-mail and its link with a work process. E-mails are thus managed at the account level, not individually. E-mails from designated "Capstone accounts" (usually higher-level officials) can quite easily be scheduled for permanent retention while all other accounts can be scheduled for temporary retention. An important advantage is that it reduces the burden on individual users to manually categorize and file each e-mail. However in a European and Dutch context it also raises discussions regarding the fundamental premise, privacy, retention, technology and retrieval which we will discuss later on in this article.

Thirdly there is the archiving of e-mail through a workflow based information system. These systems often include an e-mail functionality. The e-mail that the server receives or sends is forwarded to the workflow based information system, quite similar to how an e-mail is sent to a regular mailbox. The mail functionality in websites works similarly. Think of a filled-in web form: these are also forwarded as an e-mail via the server. Another possibility to classify and categorize e-mails is the use of business rules, for example through the use of an unique case number. An important advantage of this method is that e-mail can be automatically linked (without human interaction) to a Document Management System. A disadvantage is that it requires clearly defined workflows and is less suited for e-mails sent internally in an organization and for those linked to policy related processes.

A new method is auto-classification of e-mail. This is an algorithmic- or AI-driven process that automatically categorizes e-mails based on their content, context, and other criteria. This aims to route e-mails to the appropriate folders, teams, or individuals without manual intervention. The advantages are clear: it reduces the time spent manually sorting e-mails and ensures that e-mails are categorized according to predefined rules. However, it is a new technology that raises questions concerning accuracy and privacy. Training and checks by humans are therefore necessary.

A final possibility is manual appraisal and retention by the archival institution. In this approach no e-mail can be deleted until the archival institution has appraised the e-mail, either as an account, or individually. This method fits best when archival institutions receive private archives, e.g. from a specific person. This method does, in the end, also rely on the methods mentioned above, now however undertaken by the archivist, not by the individual. This raises questions regarding privacy, authenticity and the practicality of this approach. Appraising all e-mail records of an organization by an archivist, considering the volumes, is nigh to impossible.

Introduction of the Capstone Method in the Netherlands

Manual categorization and archiving is currently the most used method for e-mail archiving in the Netherlands but has been proven to be too time-consuming and prone to error. Government information is difficult to find and - even worse - gets lost. One telltale sign is that on the national level government officials spend an average of 37 minutes per day searching for the right information. Another sign is the long and very persistent delays in the handling of Freedom of Information requests of citizens and journalists, often resulting in incomplete answers, leading to more issues and loss of trust in the government down the line.

The basic principles of the Dutch approach remain the same as those in the United States. E-mails are categorized based on the role or position of the account holder, rather than the content of each e-mail. E-mails from Capstone officials - representing approximately 1 to 1,5% of all officials - are scheduled for permanent retention and transfer to an archival institution, while e-mails from other roles are scheduled for temporary retention. The Dutch term for Capstone official is ‘sleutelfunctie’, which translates as key function. Therefore in the Netherlands the Capstone method is also known as the ‘sleutelfunctiemethodiek’, meaning key function method.

The Capstone method for e-mail archiving is or will be adopted by different governments in the Netherlands. A large number of Dutch municipalities have already implemented it in 2024 with the support of the VNG (Association of Dutch Municipalities), while others are expected to follow. The national government had promised to implement the Capstone method in 2021, but still has not yet achieved this goal by 2024. The reason is that discussions keep arising in new and various areas.

Discussions Regarding the Capstone Method in the Netherlands

Fundamental Premise

The discussion among professionals centers around the fundamental premise of the Capstone approach. That the context of an e-mail can be defined by the function and account of the user, rather than the process and specific transaction (file). In other words, with this approach we deviate from Schellenberg's, especially the appraisal of records according to the evidential value that document the processes of the organization. This has been, and mostly still is, the basis for appraisal and the creation of retention schedules in the Netherlands. For a lot of practitioners, it still seems heretical to go against the grain. Organizing information according to process and transaction provides valuable context and ensures retrieval of information only within the context, the particular transaction (file) to which it belongs. Any deviation from this fundamental principle is non-negotiable.

The premise behind Capstone is however also well-known, that of original order. The records are created and received in the context of the e-mail account and are best understood within that context. This organization of information is not new, either. Paper correspondence files have existed for a long time. The unique nature of e-mail, where any topic can be discussed, combined with the presence of e-mail chains and the sheer amount of e-mails, requires a solution that does not overly burden the e-mail user but ensures the capture of all e-mail records. Still this discussion has legs. Looking at the forms of secondary use, as Schellenberg suggests, may help in coming together to focus on making the records accessible to those users in all the ways we can.

Privacy

Privacy issues have been an interesting discussion. Is Capstone compliant with GDPR? As the appraisal is on function and role, not of individual e-mails, this might lead to more privacy-related information to be retained for longer than necessary or even allowed. At least that's the perception. Interestingly, research shows that using the Capstone methodology less e-mails are retained permanently, approximately 3 percent, than if all e-mails would be retained according to the process, which results in about 5 percent. However, that is not enough. Additional measures are required to ensure that e-mails are not accessible without the correct controls. Since e-mail can be used, in a limited way, for private use, measures must be taken to ensure these can be excluded from archiving by the user. For that purpose a ten-week period is introduced in the current policy: this gives the user enough time – even after a prolonged absence due to illness or leave – and the opportunity to delete any private e-mails or exclude them from archiving via a separate folder. The perception is however still that e-mail belongs to the user, not the organization. This, however, is not the case if you look at the Archival and Public Access laws, as was confirmed by the highest court in the Netherlands.

However, privacy remains a crucial consideration when implementing Capstone.

Retention

Appraising e-mails using Capstone is easy, having a retention period for e-mails based on the function of the user of the e-mail account as well. Especially when this is based on the date an e-mail was received or sent. The discussion therefore centers around the retention period itself. The model retention scheme provides a retention period of 10 years for non-Capstone officials. Is this the right period? Shouldn't it be shorter or longer? And what if by law, there's a required shorter retention period?

The first is always easy, any proposed retention period, any proposed Capstone official, that is in the model can be changed by the organization, if there's a clear reasoning behind it. It follows a risk management assessment, taking into account and balancing the different interests in the information. And is also submitted to and decided upon by an archivist.

With regards to the information that should be destroyed according to other laws, well, this should be possible in the technical solution, those periods should be recognized. However, the fundamental question is more, why is e-mail used in these cases? It's information that is so sensitive that it should only be shared in controlled environments, which e-mail patently isn't. So how can we change the behaviour? That's the first discussion to have.

Technology

The technology to implement Capstone exists. However, with more exemptions and rules it will become harder and require more discussion. In general we can see three different methods:

1. Management in place. The e-mail stays in the e-mail account, however retention periods are added, preventing the user from permanently deleting the e-mail. This method is easy to implement, except when you also require a period upfront to exempt the e-mail from this rule. Like the 10 weeks highlighted earlier. And it still depends on the search options within the e-mail system used, which may not be the best. Still this is an excellent option.

2. Capture when receiving or sending e-mail and archiving them in a separate system. This option, often called journaling, is easy to implement, it will indeed capture all e-mail, however an option must then be build to allow the deletion of personal (private) e-mails. It does have the benefit of making the search easier. However there still needs to be a retention period implemented in the e-mail account.

3. Capture e-mails from the e-mail account after 10 weeks and archiving the e-mails in a separate system. This option is a bit more complex, as it requires an agent to run the retrieve of the e-mail and copy or move the e-mail to another system. When you copy the e-mail the same applies here as for option 2. You must still implement retention periods in the e-mail accounts. However, the privacy aspect is better ensured and therefore possibly a better fit.

Retrieval

The retrieval of information needs to fit different needs. As stated earlier, here the secondary use is important. Knowing this usage and providing a solution that delivers the information in a manner required by a user is essential. In general, e-mail is the best source for finding out who did what and when. The metadata that comes standard with each e-mail makes this an easy question to answer. More proper process and topic related questions are harder and require a full-text search option. The discussion therefore often centers around this usage, seeing as this is more complex. Additionally, the nature of e-mail, with the same messages across multiple accounts, poses a potential issue for retrieval. You can retrieve the same e-mail multiple times. De-duplication of search results is therefore required from any technical solution.

And finally the discussion here often centers around just the sheer amount of e-mails, how can you retrieve information when there are five billion e-mails send or retrieved each year?

The End?

This is only the beginning. And it is possibly the end. How long will e-mail retain its position as the main form of communication in society and government? That is unclear.

Other forms of communication are creeping up. Especially all forms of messaging apps, like WhatsApp and Signal. And platforms like MS teams and Slack offer other ways to communicate and share information with a group of colleagues. It’s clear these will (at least partially) take over the role of e-mail. However the questions that are asked of e-mail archiving are also there with these solutions. It is all based on free-flowing communication and sharing of information instead of the organized, process-controlled way of working that allowed for the capture and archiving of information in a more structured way.

Which means that we should look at ways of providing structured views together with the view of the original order and usage. Using AI might make it possible to also appraise the information in different ways, on the level of the individual record. This might also be needed to ensure compliance with GDPR. This will have an even greater impact, especially on the retrieval of information. How can we restrict access to information within e-mails, but still provide enough information for the user of the archives?

Finally, you can see generative AI already providing summaries of e-mails, which begs the question on how we should appraise this kind of information.

It is therefore important to think ahead, look at what might be possible now and in the future and prepare ourselves. We've taken already more than fifty years, and still aren't ready to solve the question of e-mail archiving. It's high time we came to grips with it, and then with other and upcoming means of communication and working.