Managing and Preserving E-mails

Abstract

The successful management of e-mail messages requires effort across the entire life-cycle to ensure that e-mail records are identified, captured, managed, preserved, and made accessible in a manner that safeguards the authenticity and integrity of the records and any associated attachments. At the same time, e-mails that are not records must also be properly managed. This involves not only technical challenges, but also those of a legal, organisational and cultural nature. All must be identified, understood, and addressed. Collaboration and communication with all stakeholders that play a role in the life cycle, from creators and recipients to management, IT staff, and data re-users, is essential. Ultimately, an approach that integrates e-mail management into overall digital records management is likely to be most successful, although some explicit measures in the areas of policy and education are required to ensure that e-mails records are afforded records status and treated appropriately by users.

This brief article summarises a much larger work, *'Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages'* by the same author. *'Curating E-mails'* was published by the UK Digital Curation Centre (DCC) in July 2006.

Introduction

E-mail has revolutionised communications in the modern working environment. Thousands of valuable e-mail records are created and exchanged by e-mail systems in offices around the world every day. Yet despite this prolific level of usage, the process of managing and preserving - or curating - e-mail records remains a significant challenge for many organisations. This is not least because of the technical challenge involved in e-mail curation, but also the accompanying organisational and cultural challenges of raising awareness, achieving compliance, and integrating e-mail management into organisational infrastructures. E-mail has thus become both a solution and a problem.

Failure to properly address these challenges has already resulted in billion-dollar financial penalties, negative public perception, and corporate embarrassment for a number of high-profile organisations. Such repercussions can only be avoided if action is taken across the life-cycle of not only e-mail records but also other, more transient e-mail messages.

Digital Curation & E-mails

Digital curation is the active management and appraisal of any type of digital information across its entire life-cycle, from creation and active use, to appraisal & selection, disposal, preservation, archiving, access, and re-use. This involves a number of tasks across the life-cycle that require attention by an array of different stakeholders, from records and data creators or originators, to records managers, archivists, librarians, IT specialists, and even organisational management. Addressing these tasks as part of a continuous chain of events enables the right actions to be taken from the very beginning, i.e. from the point of creation, to ensure that:

- transient information or data with little or expired value is managed, identified, and reliably destroyed;
- important information is created in a sustainable manner with the right metadata and

managed properly so that it can be reliably preserved and re-used despite changes to the technological environment that threaten its authenticity.

The ongoing management and preservation of authentic digital records is a challenge into which research is taking place at a (supra)-national level. Digital records are highly environmentally dependent. Their content and value is accessible only through the processing of file(s) on hardware and software, and they often rely upon on a particular technological and organisational environment for context and proper rendition of their content. As technology continues to develop and advance, curators must find ways to preserve records with a particular technological dependency in new technological environments. This is an issue not just for records, but also for other types of non-record digital materials that have value outside of a record-keeping context.

E-mails are just like these other types of digital records and materials in that they too require curation if they are to remain accessible, usable and authentic for future users, despite changes to their environment. Unfortunately, the curation of e-mail messages is often overlooked as:

- it is a type of record for which there is no parallel in the analogue world and which is often not properly integrated into an overall record-keeping infrastructure;

- responsibility for e-mail curation is frequently not explicitly allocated and is distributed across a wide array of different staff and roles.

Furthermore, the value of some e-mail messages as records, whilst recognised by records management staff, is often undermined by an lack of official organisational responsibility for e-mail and a sense of personal ownership by users over their 'own' inboxes. Despite this, e-mails, both records and non-records, must be curated if legal obligations are to be met, technological obsolescence – the loss of access to content because of a change in technology - avoided, and e-mails are to assume their rightful place in organisational memory.

Legal issues

One of the biggest risks is using e-mail arises from the impact and applicability of legislation on e-mail messages and systems. All stored e-mail messages, whether they be designated record or non-record, and stored in a users inbox, on a shared drive, or in an Electronic Records Management System (ERMS), are subject to a number of legal obligations. This paper draws specifically on UK and EU legislation to illustrate this point.

In the UK, the **Data Protection Act** and the **Freedom of Information (FoI) Act** are probably the most well-known examples and relate particularly to accessibility and proper management of information. The issue of deletion is also significant, for deleted messages can still form part of an organisation's electronic information corpus if they are not properly purged from recycle bins, deleted folders, or back up tapes. To ensure full compliance with the principles of both Acts, a data cleansing or destruction strategy that will ensure inappropriate or unnecessary data is destroyed is therefore just as essential as a data retention strategy.

Other legal Acts have an impact on the way that an organisation may monitor or intercept e-mail communications addressed to individual employees: the UK **Regulation of Investigatory Powers Act** and the EU **Human Rights Act** both impact upon the procedures and methods that an organisation can use to make sure it captures essential e-mails and complies with the UK Data Protection and FoI Acts. Other pertinent legislation includes the **Law of Confidentiality**, **Copyright Law**, and **Database Right**, all of which can be applied to the use and re-use of e-mail messages.

Raising awareness of the legal consequences of failing to properly manage and curate e-mail messages can be an excellent driver for achieving organisational compliance. A risk management analysis can identity legal and other types of risks, their likelihood and the consequences, and can play a significant role in raising awareness and obtaining sufficient resources to fund and support an e-mail management and curation programme.

Roles and responsibilities

For Creators and Recipients

Creators and recipients of e-mail messages are the first in the chain of stakeholders with a role to play in the management and curation of e-mail messages. Decisions that are taken at the creation and active use stage of the records life-cycle can have a significant impact on options for records management and preservation at a later date. Creation and receipt activities must be carefully controlled, for example regarding message transmission format, storage format, use of attachments, and storage location, to ensure that these options remain open. Guidelines and training for users on the issues below can help achieve this.

Message creation

Written guidance and training sessions can help ensure that staff are aware of the status of many e-mail messages as organisational records and are an opportunity to communicate and encourage good creation practices. Three particularly useful creation practices that require very little extra effort from the user and help identify message context both now and for the future are:

- Using the address books embedded within many e-mail applications. Address books enable many e-mail applications to link a anonymous e-mail address with the name and position of a person, e.g. mp281@bath.ac.uk becomes Maureen Pennock of UKOLN at the University of Bath.

- Including a message signature that identifies the sender. This may be a digital and authenticated signature, or simply a plain text insertion that identifies the name and position of the message sender, the organisation they represent, and further contact details.

- Inserting meaningful subject lines.

Guidelines that encourage a consistent approach to message creation have additional value where large collections are concerned, as messages that are created in a consistent manner are much easier and more economical to preserve in bulk than a collection of disparate messages. Guidelines on how to communicate certain types of content over e-mail, i.e. whether to put the content in the body of the message or in an attachment, can have a similar impact. Such guidelines can also help reinforce corporate or brand identity by ensuring that messages conform to a particular house-style.

Inbox management

The average office worker handles seventy five message each day.[1] With message exchange taking place on this scale, inboxes can quickly become unmanageable if filtering and filing or classification is not implemented. Training and guidelines regarding the storage and management of e-mails in users inboxes can help limit inbox sizes, assist in compliance with legal requirements,

---

1   Special briefing on E-mail management, Aslib magazine, March 2006.

and increase retrievability. Furthermore, developing and communicating guidelines on implementing a useful filing structure may also ensure that meaningful and appropriate folder names and naming conventions are used. This increases efficiency and helps other users to understand the inbox structure if wider access becomes necessary.

Advice on filing mails and managing mails should include guidelines on identifying records with value that should be retained, from messages of a more transient nature that should either be retained only for a very short period or which should be immediately deleted. Ensuring that users are aware of their legal responsibilities may help achieve further compliance in this.

For Curators

Curatorial responsibilities permeate across the entire life-cycle of digital objects, not just the creation phase, and curation of digital resources required collaboration and input from staff across the entire breadth of institutions. This section details those activities that are often driven by staff with a curatorial role (such as records managers and archivists), although many also require input from others in the organisation, including IT staff and senior management.

Policy

Policies indicate the extent to which activities are embedded in an organisation and the importance attributed to them. E-mail policies are essential to address the common misconception that e-mail is part of the users' personal domain and to reinforce organisational ownership of official e-mail records.

Many organisations have e-mail policies. These have historically related to the use of e-mail, although they are increasingly being expanded to address interception conditions and basic archiving and retention periods. The JISC-funded project on Institutional Records Management and E-mail that was carried out by the University of Loughborough resulted in a useful generic policy on e-mail retention and disposal. This is available from the project website.[2]

Very few institutions have broader e-mail policies in place that cover curation or preservation, although interest in the development of such policies is slowly growing.

Appraisal & Selection

Appraisal and selection decisions must be taken to determine what will be added to a collection. This is, to an extent, a 'weeding out' of messages saved by users, who will generally retain more than is necessary. Organisational records retention schedules can be applied to e-mail records just as with other types of records.

Decisions must also be taken on whether entire message threads should be retained or whether it is possible to clearly identify context from a limited number of messages in a thread.

Storage and preservation of authentic e-mail records

Storage and preservation is not necessarily as simple as saving messages on a server or in an ERMS and leaving them there for as long as they are required. Depending on the retention period, preservation actions may be required to ensure that records remain accessible and authentic despite changes to the supporting technology. These actions will often involve changes to the format in

---

2   http://www.lboro.ac.uk/computing/irm/index.html

which the e-mail records are stored, particularly if they have been saved as proprietary files. Similar action is required for e-mail attachments.

Preservation often involves converting messages to a more sustainable format. For e-mail messages, the selected format must suitable and capable of representing the message in an authentic manner. This means that the context, content, structure, and appearance of the messages must be preserved. Appearance does not have to be preserved in a manner identical to its original rendition as e-mails are often represented quite differently on different PCs even when first exchanged. In most circumstances, printing to paper or converting messages to TIFF images or PDF will not be suitable because these formats will not preserve all of the important characteristics of the messages. Complete plain text or XML files with all message headers intact are preferred. For attachments, the preservation and authenticity requirements will depend upon each type of attachment, for example, PDF/A, ODT or TIFF may be suitable for images or some documents. Documents structured in XML (eXtensible Markup language) may also be a strong candidate.

The technical storage infrastructure should also be sustainable. It must be possible to move records through time, possibly to other storage infrastructures. Open source solutions should be considered for this purpose. Furthermore, specifications concerning security arrangements must be drafted and implemented to ensure that stored records are not accidentally or maliciously altered.

A number of commercial management and storage systems are available that exclusively address e-mail messages. However, the ability of such systems to manage, preserve, and provide access to e-mail messages and attachments over extended periods of time has yet to be proven. Many are proprietary systems, which may hinder future access to the contents if the supplier goes out of business and support for the system is no longer available. It may also be more difficult to integrate the system with other digital object storage systems. Should organisations decide to use such systems, these issues should be addressed early in the system life-cycle to minimise possible disruptions at a later date.

Re-use of stored records

Facilitating and enabling the re-use of stored data is one of the primary objectives in digital curation. Re-use requirements and activities may differ from the manner in which original records were used and curators must ensure that appropriate facilities are provided to meet these requirements as much as possible.

Access provisions must be implemented that protect the integrity of the stored records whilst still providing users with freedom to manipulate retrieved records. This can be done by, for example, storing a preservation master and responding to access requests by creating on-the-fly copies. Resource discovery tools must be provided so that users and curators can locate and retrieve records. Other access and re-use requirements should be identified according to the needs of each organisation.

An Integrated Approach

Integrating e-mail storage and management into the wider organisational records management strategy is recommended as the most efficient way to manage all organisational records and respond to legal obligations. This minimises the costs when responding to an FoI, Data Protection or other legal discovery notice is received as the number of sites to be combed is minimised. Costs are also lessened as multiple systems do not have to be managed through time. Investment can instead be concentrated on the main records repository. An integrated approach helps reinforce the role of e-

mails as formal records that should be treated in the same manner as other organisational records. It also enables transfer activities to be coordinated and initiated from a single site when transfer to external repositories (for example, those run by longer-term archival bodies) is required.

Conclusion & future developments

This article has offered a brief overview of the main challenges, roles, and responsibilities involved in the successful management and preservation of e-mails, a vastly prolific yet rarely curated type of electronic record. Yet technology continues to evolve, and e-mail is already being supplemented by other, newer types of technologies for which curation solutions must also be sought. These include Instant Messaging (IM), RSS (Really Simple Syndication/RDF Site Summary) news feed technology, wikis, discussion forums, and web blogs. If these newer channels of communication create records when used, then they must also be managed as a records creating medium. However, it is unlikely that these will signal the imminent death of e-mail – e-mail has too many advantages that these other technologies cannot yet match, and e-mail is so widely embedded in modern working practices that, despite ongoing advances in technology, it will undoubtedly persist in the office environment for many years to come. Newer technologies are only, as yet, supplementing e-mail; they are not supplanting it. The challenges of successfully managing, storing and preserving e-mails will therefore remain relevant for some time yet.

'Curating E-Mails: A life-cycle approach to the management and preservation of e-mail messages' can be downloaded free of charge from the DCC website at http://www.dcc.ac.uk/resource/curation-manual/chapters/curating-e-mails/

Author Details:

Maureen Pennock
Research Officer
Digital Curation Centre
UKOLN, University of Bath

E-mail: m.pennock@ukoln.ac.uk
Web site: http://www.dcc.ac.uk/ http://www.ukoln.ac.uk/