ARTICLE

Attributes of Personal Electronic Records

Matt Balogh[1]*, William Billingsley[1], David Paul[1], and Mary Anne Kennan[2]

[1]Computer-Science – School of Science and Technology, University of New England, Armidale, NSW, Australia; [2]School of Information and Communication Studies, Charles Sturt University, Wagga Wagga, NSW, Australia

Abstract

The purpose of this article is to identify the key attributes of personal electronic records in order to develop systems that may enable people to manage them in the home. As more personal information becomes electronic, this is increasingly necessary. Personal electronic records were identified and categorised using interviews and virtual guided tours. Three main attributes were identified: primary user-subjective categories; attributes which identify the circumstances that give rise to the records; and attributes which describe the legal validity of each record. In addition to providing an improved understanding of personal electronic records in the home, these attributes are developed into a set of potential metadata fields.

Keywords: *Archives*; *Metadata*; *Personal electronic records*; *Personal information management*; *Records.*

I n this article we define what we mean by personal electronic records and explore their attributes. Our aim is to better understand the categories that people may use to sort their personal records with the purpose of re-finding them when required. To achieve this, people need to know what records they have created or received, and how to prioritise the task of keeping them. We draw on research from different fields to establish our background knowledge, prior to conducting a study using interviews with the guided tour method to establish a database of personal records retained by a sample of participants. The database was then analysed to address our questions about personal records.

For reasons explained below, in this article we adopt the term 'personal records' to describe the personal information and documents that people deal with in the home relating to their personal affairs, such as bills, insurance documents, receipts and so forth, as opposed to information and documents that people may deal with for the purpose of work or study tasks. We refer to those records that people retain in an electronic form in the home as personal electronic records. In this article we focus specifically on the nature of the personal information and documents themselves, rather than the practices adopted in managing these records.

*Correspondence: Matt Balogh Email: mbalogh@myune.edu.au

It is important that people keep and manage some personal records to ensure that they pay bills on time, maintain active insurance policies for concerns such as their motor vehicles and have the information they need for tax reporting when required. A notable example of the need to retain documentation for unforeseen events is illustrated by the Australian Government's '*robo-debt*' scheme that operated in 2018 and 2019. The scheme compared customer records of a social services department, Centrelink, to the same person's income records at the Australian Taxation Office (ATO) and automatically issued debt recovery proceedings if the information appeared to indicate an over-claim on Centrelink benefits.[1] A major problem with the scheme was that Centrelink benefits were paid based on the recipients' financial status at a given time during the financial year, such as when they were unemployed. The tax records relied on a full-year summary, so that if the recipient got a job later in the year, the figures could make them appear to have been ineligible for the Centrelink benefit based on income averaging through the year. The burden was placed on the recipient to disprove the debt. The recipient had to have week-by-week income records from up to 6 years prior to argue their case. This debt collection process was later curtailed by the Australian Federal Court, but highlighted significant record-keeping issues for people, including the need to keep pay slips and their own bank statements. Banks in Australia are only required to keep records for 5 years.[2]

The habit of keeping paper records of bills, payments, and similar records will be familiar to people whose personal record-keeping pre-dates the prevalence of electronic transactions. The more recent ubiquity of the option to receive electronic bills, pay bills online, shop online, or conducting one's correspondence by email and text messages raises the question of the validity of personal electronic records in providing a comparable level of evidence of transactions, as previously accepted for paper records. While it is understood that people need to retain personal records, there are no clear rules that describe which records need to be retained, nor guidelines that predict which records may need to be re-found in the future, perhaps long after they have been forgotten about. As people increasingly rely on electronic versions of their records, a further challenge lies in ensuring that the electronic versions are sufficient for future needs. By studying the nature of personal electronic records in the home, we contribute to improved understanding of the value of personal records and an ability to predict and identify records that need to be retained and how to best manage those. These may comprise records of transactions or purchases, life documents such as certificates, or a myriad of other records or notes of interest to the owner. Each type of record may need to be retained for a different amount of time, officially 5 years for tax documents in Australia, records of purchases may need to be kept for different periods depending on the item purchased, while other records may need to be kept for a lifetime or longer.[3] However, there are currently no disposal schedules for an individual's personal records, such as those created by state and federal authorities for public and private organisations.

The aim of this research was to better understand the nature of personal electronic records in the home in a way that helps identify, foresee, and prioritise the records that need to be retained. By achieving a better understanding of what gives rise to personal records (including those received and/or retained electronically) and the role they play in our lives, we may be able to lay down the groundwork for a more effective management system for personal electronic records.

We explore the question of whether it is possible to predict the creation of personal records in the future. We also explore the legal validity of different kinds of records and investigate the components of personal records and how they can be categorised. To address these questions, we have surveyed the literature for the types of items to which we are referring and have selected to call them personal electronic records for the reasons described in the next section. In addition, research was conducted with 30 participants to establish a database of personal

electronic records. We then analysed this database of records elicited from participants to develop three perspectives on personal electronic records reported below.

Defining personal electronic records

We have selected to use the term 'personal electronic records' to describe the combination of personal electronic information and documents maintained in the home. In part, this terminology is intended to distinguish discussion of such records in the home, as opposed to the information discussed in the Personal Information Management (PIM) literature, which largely examines the personal information of individual students and people in the workplace, particularly knowledge workers.[4] There is extensive PIM literature, the examples of which include Bergman, Israeli, and Whittaker,[5] Bruce, Jones, and Dumais,[6] Henderson,[7] Kwasnik,[8] and Oh.[9] Some aspects of PIM are quite applicable to the study of personal electronic records in the home, such as the importance of taking control of our records and information. As Jones[10] explained:

> …PIM is about finding, keeping, organizing, and maintaining information. PIM is also about managing privacy and the flow of information. We need to keep other people from getting at our information without our permission. We need to protect our time and attention against an onslaught of information from telephone calls, email messages, the television, radio, and the Web. PIM is also about measurement and evaluation: Is this new tool worth the trouble?[11]

The study of personal electronic records in the home also has roots in records management, which is the study of information in records that are an account of something of enduring value.[12] Documents on a computer are often synonymous with files; however, a file can include multiple documents, or a document may comprise several files. For reasons discussed in the related literature section of this article, we have adopted the word 'record' to embrace a wide variety of content and formats and avoid confusion with the use of the word 'documents' in computing to describe computer files. In support of this choice of terminology, we draw on Finnell's definition of records, which includes memos, e-mails, and instant messages as examples of formats of records.[13] Additionally, the Society of American Archivists defines records as 'data or information stored on a medium and used as an extension of human memory or to support accountability', providing more than 50 examples of types of records, including 'graphic records', 'narrative record', and 'housekeeping records'.[14] McKemmish used the words 'personal recordkeeping' to describe records that are 'evidence of me'.[15] Bass also used the word 'records' to describe the lifetime 'day after day' accumulation of records in a digital format as 'personal digital records'.[16] In practice, as in PIM, a significant proportion of research and published literature relating to records management 'has had a governmental or large organizational focus'.[17] The Sedona Guidelines for Managing Information and Records in the Electronic Age were specifically framed for organisations.[18] The study of records management alerts us to the breadth of personal information and documents that people may deal with in their day-to-day lives and suggests the adoption of the word 'records' to describe these, at least in the electronic context.

The study of personal electronic records in the home is also informed by personal archiving, which explores continual storage of personal documents through one's life, and particularly those in an electronic format.[19] Personal digital archiving focuses on long-term heritage, rather than on the short-term task of document management. Kim[20] observed that personal digital archiving offers a way of preserving records of sentimental value, historical value, the value of self-identity and personal legacy, and value in sharing useful or interesting documents with

others. However, these attributes differ from those required for the classification of current and active records.

By contrast, this study of personal electronic records includes and focuses on the active stage of records dealt with in the home, such as bills pending payment, records relating to acquisitions in progress, reminder notes, and shopping lists – in other words, records with a currency that does not associate comfortably with the word 'archiving'. Another point of difference between the study of personal electronic records and other areas of study that reference digital archives or digital records is the adoption of the broader term 'electronic'. We use the word 'electronic' to embrace all forms of information and documents that are not in a physical paper or engraved format. Digital records comprise those electronic records that are in the form of combinations of numbers and letters that are machine readable, such as a Word document.[21] 'Electronic records' includes records in more formats, such as photographs and image scans of documents which are not easily machine readable.

Hence, the term 'personal electronic records' describes the mix of records that people deal with on a daily basis related to their home affairs. We consider the content and the function of the records, which may manifest in a variety of formats including emails, text messages, and photographs of records such as receipts, recipes, and notes. Personal electronic records may also include some work-related records, such as payslips or pay stubs, but exclude work documents or correspondence that are not personal to the user.

Some of the attributes of personal records in the home are well understood. Whether they are in a hardcopy or electronic format, they typically have a creation date. If they have been sent to someone, there is a sender and a receiver: a 'from' and 'to' in digital terms. Electronic documents often have a most recent modification date in lieu of, or in addition to, the creation date as well as an indication of size. Electronic documents have a file name. But in addition to these elements, there are attributes of records that do not have a standardised format, such as the subject and purpose of the record. Prior to conducting this research, we explored what might already be known about personal electronic records in the home focussing on the nature of the records themselves, rather than related behaviours.

Related literature

As we have noted, the study of personal electronic records management draws on several different disciplines, including PIM (which is a branch of computer science), records management, and personal archiving. PIM also provides a range of relevant case studies and findings discussed below.

Documents and records

In the glossary to 'Keeping Archives', Acland defined a document as 'recorded information regardless of medium or form'.[22] Documents may contain information to which they relate or may form a record of that information in a required format. Documents can encompass any form of records if they are compiled in a 'collection, indexed, cross-referenced, etc'.[23] Yeo, citing Oliver and Foscarini, described records as 'information as evidence'.[24] Buckland also described one purpose of documents as 'storing…evidence of some assertion'.[25] Conversely, Roberts observed that 'Where the essentially evidential quality of a record is not accepted, that is, where records are simply equated with recorded information, the distinction between records and documents tends to disappear'.[26] We concur with this interpretation and do not concern ourselves with a definitional distinction between documents and records. We use the word 'records' in reference to items in the home, partly to draw a distinction between personal records management and PIM, which as we have previously observed tends to investigate

information and document management in workplaces,[27] and to avoid confusion with the use of the word 'documents' in reference to individual computer files.

Records as evidence

Records management contributes to the study of personal records management in the home as both fields explore the nature and use of documents and records. The notion of records as evidence is useful in that it explains why many records need to be retained. For instance, a collection of telephone bills and receipts provides evidence of how much one has been charged and paid. In the same vein, an entry for the Records Continuum Model in the Encyclopaedia of Library and Information Sciences observes that all transactions 'can leave archival traces',[28] as they become records and hence subject to the unified process for record-keeping including archiving. Myburgh described records as 'documents which provide evidence of business transactions that have taken place'.[29] Existing literature clearly shows a relationship between personal records and their use as evidence of transactions. Zacklad[30] suggested that if a digital document is a record of a transaction between two parties, then the transaction and terms described in the electronic record are partially verified by the fact that a co-operative transaction occurred. To be clear, where we are talking about records as evidence, we are talking about a subset of all personal records that perform this function. We do not suggest that all records are evidentiary; rather we adopt McKemmish and colleagues' broader definition that records '…are vehicles of communication and interaction, facilitators of decision-making, enablers of continuity, consistency and effectiveness in human action, memory stores, identity shapers, repositories of experience, evidence of rights and obligations'.[31]

Many records in hardcopy comprise letters or agreements endorsed by one or more signatures. A signature created by hand is sometimes referred to as a wet signature,[32] as opposed to electronic signatures, often referred to as E-signatures. Financial transactions and trade occur readily through the Internet with no wet signatures, relying on the expectation that if there is a problem, transactions and terms can be verified.[33] The legal validity of electronic records in lieu of paper documents with wet signatures for a range of purposes has been re-affirmed by a wave of legislation around the turn of the century. There is little indication of legal recognition of E-signatures; rather, Australian federal law adopted the notion that signatures on documents are not required if authentication of a document is:

> … as reliable as appropriate for the purpose for which the electronic communication was generated or communicated, in the light of all the circumstances, including any relevant agreement.[34]

Similar laws came in around the developed world enabling paperless transactions.[35] The general acceptability of E-signatures on contracts and documents in the United States and Canada increased significantly in 2020.[36] Laws relating to signatures were further relaxed during the coronavirus disease 2019 (COVID-19) pandemic in 2020. For instance, in New South Wales, documents requiring witnessed signatures were amended to provide for witnessing of signatures by means of video calls.[37]

In combining the interpretations of Myburgh,[38] Zacklad,[39] and Alba[40] with the legislation on electronic transactions, it is apparent that, for day-to-day transactions, personal electronic records are sufficient as evidence and neither paper records nor signatures in any form are generally required. In the case that there is a matching transaction, the electronic record and the transaction work together. The electronic record describes the transaction, typically a product or service in exchange for money. The transaction provides verification that the record describes something that really occurred. Electronic records verified by a matching transaction

are accepted as valid, without the necessity of paper evidence. For example, an email regarding the details of an online purchase is verified by the simultaneous transfer of a matching payment from one party to the other.

Personal records

Taking a broad perspective on personal electronic records, Jones[41] described everyone as having a '*Personal Space of Information (PSI)*' which we inhabit in a similar way to the habitation of a physical space. A PSI contains information that is personal in any of six senses:

(1) Owned by me
(2) About me
(3) Directed to me
(4) Sent by me
(5) Already experienced by me
(6) Useful to me[42]

Jones' observation describes the relationship between a person and their electronic records but does not explore the nature of the records themselves. The most specific definition of personal records was provided by Smith,[43] comprising a list of 36 common documents, of which the personal documents were:

(1) Accident report
(2) Bank or pension records
(3) Bankruptcy certificate
(4) Birth certificate
(5) Death certificate
(6) Diploma
(7) Divorce decree
(8) Insurance/health
(9) Insurance card/cert.
(10) Insurance claim form
(11) Insurance plan
(12) Marriage licence
(13) Medical records
(14) Mortgage agreement
(15) Passport/visa
(16) Payslip
(17) Rental records
(18) Stock certificate
(19) Tax form (some categories have been combined, indicated by a slash '/')[44]

Smith also suggested, somewhat counter-intuitively, that documents that are used for a consequential purpose, such as a contract, licence, completed form, or a certificate, are deemed 'creative', while documents with fewer consequences, such as a novel, textbooks, or painting, are deemed 'non-creative'.[45] Smith further distinguished between allographic and autographic documents – allographic being those documents for which there is no particular original form, as opposed to autographic documents that have a clear original, and therefore other possible versions that are inevitably copies. Smith cited books, recipes, advertising fliers, and a completed tax form as examples of allographic documents and a painting, birth certificate, and

will as examples of autographic documents.[46] Bergman, Whittaker, and Tish[47] studied personal music collections, commenting: 'Surprisingly there is very little research on music collections within the PIM research community… Nevertheless, there are similarities between the organization of music collections and prior PIM literature'.[48] The similarity observed here is that software tools used to manage music collections help people to search for and find music in a variety of ways such as by the title or genre of music, but the music files remain independent of the software used to manage them. The same is true of software used to manage photographs and other collections. In our research we do not consider the management of personal records by systems that are already in existence and specific to a particular kind of collection such as photograph or music collection management software.

Smith's detailed list of personal documents provides useful examples, but leaves open the question of how these documents come about, and why these documents, and not others, are retained. The literature does not answer the question of what gives rise to personal documents, but it informs our perspective on this question by telling us that documents and records are the outcome of how users manage and retain them. The question of why people keep certain records has been addressed to some extent by Furner's extension of Smith's framework, taking the concept of a document defined by what one can do with it to include actions such as 'finding it, identifying it, selecting it, and obtaining access to it … as well as organizing it, classifying it, and indexing it, and reading it, interpreting it, citing it, and using it, in many and various ways'.[49]

Oh observed that personal electronic items often transition through a temporary categorisation and storage stage, such as the desktop or downloads folders.[50] Oh noted that items go through an 'active' stage, as also observed by Barreau and Nardi.[51] Oh also noticed a reluctance to add new categories, with people preferring to force items into existing categories.[52] She also observed that categories have blurred boundaries.[53] Oh found that, over time categories, and therefore folders, may be merged. For example, when users found they only had one or a few items in a particular category, those items may later be moved to be incorporated in another category. Similarly, categories can be subdivided (or new categories created), not necessarily at inception of the first item in the category, but at the point that the new category had sufficient items to merit its own creation.[54] Oh described 'purpose' and 'use' as being the most influential factors on how items are categorised,[55] and, to a lesser extent, 'accessibility', 'topic', 'format', 'source', and 'time'.[56] Oh also identified the challenge of filing items that could equally fit into more than one category. Such items either need to be allocated to only one category or duplicated. Some items don't fit into any existing categories and people are reluctant to create a new category for just one item.[57]

Categorising records

In Cognition and Categorisation, Rosch observed multiple ways that people might approach categorisation, depending on their perspective and purpose.[58] For example, consideration must also be given to the subjective priorities of the user. Bergman and colleagues also noted that documents that one user might consider important may not be considered important to another.[59] This can be contextual to the circumstances. For example, a document may be vital in the context of a short-term situation, but unimportant in the long term. A user may consider that a file belongs to one category on one occasion, but that the same file belongs in a different category on a different occasion.[60] Similarly, Oh drew on Rosch's notion of 'prototype' categories for given items[61] selected by the users, and some items clearly fit into one category or another, while the categories of others may be more blurry.[62] This begs the question as to whether and why an individual record needs to fit into only one category. Could a record not be classified

into several categories? Jones observed that 'Folders – file folders, in particular – can be regarded as an expression... of a person's internal categories'.[63] Jones cautioned that people are not very good at creating clear definitions of categories. He suggested an alternative approach wherein the computer '*learns*' the definition of a folder through the items saved within it.[64]

A common topic in the PIM literature addresses how files are saved within a hierarchical folder structure where folders are named in order to categorise their contents.[65] For a file to be categorised under two different categories, it requires that the file is either duplicated or a link or shortcut is created to represent the document in additional folders. Multi-categorisation also tends to be associated with a preference for searching for files rather than navigating them through the hierarchical folder structure.[66] Despite many experiments with tools designed to make multi-categorisation of files easier, hierarchical structures still dominate local computer systems and people often prefer to navigate to find items rather than to search.[67]

Many people use their email software or host as the repository for personal records, either because they leave email and associated records in their inbox, or sort email into folders or tag email rather than saving items to an alternative location.[68] Research in 2018 found that over 80% of respondents received at least some of their personal records electronically.[69] Crawford and colleagues developed a system, 'i-ems', that took into account the sender's email address and keywords in the subject and body of the email to predict how the email should be categorised. The predicted category was provided to the user as a suggestion, which the user could accept or amend. This system stored the user's decisions in order to improve future categorisation.[70] The paper concluded that using only the sender's email address 'achieved high precision' in predicting the correct category for email.[71] In 2014, a team at Yahoo proposed a system to automatically categorise emails in terms of category names and optimising the number of email folders.[72] The research of Grbovic and colleagues has noted that navigating to find emails was more effective than searching all email folders when users had up to 20 email folders, and that search was more effective for finding email if there were more than 20 folders. The Yahoo team sampled 600 email senders to find six latent email categories:

(1) Shopping
(2) Financial
(3) Travel
(4) Career
(5) Social
(6) Human

The Yahoo research comprised an experiment in automated classification of emails and claimed a success rate of more than 70%.[73] Nevertheless, the research did not address the nature of the emails, such as whether they indicated that a payment was required or comprised a receipt – or indicated a future appointment that might require a calendar or diary entry. This latter aspect of personal records was studied in research conducted in order to develop '*common sense*' task reminders based on calendar entries that identified 25 fields from an electronic calendar entry.[74] Comparing these two approaches indicates that personal records can be perceived from different angles – the Yahoo study classified emails into broad topics, whereas the Lieberman et al. study did not pre-empt categories, and instead searched multiple fields in order to identify the most likely matches for the required information.[75]

Metadata

Amidst the discussion of how personal records are saved and the role of folders or other sorting systems is consideration of where the categories or other descriptive information are

stored. Are they part of the record, such as a file name, or maintained in a database outside or encompassing the records, such as an operating system or folder? Metadata are information that describes other information, such as a catalogue.[76] For example, email software stores and displays the date and time of when an email arrived, as well as who it was from and to. That information is not part of the content of the email, but is the metadata about the email. The aspects of the information that the metadata describes are referred to as the 'data elements'.[77] Examples of data elements include the 'ownership and authenticity' of the information.[78] The US National Science Foundation describes metadata as a subset of data: 'Metadata summarize data content, context, structure, inter-relationships and provenance… They add relevance and purpose to data, and enable the identification of similar data in different data collections'.[79] An important aspect of metadata are that it usually comprises 'controlled vocabularies applied to a digital object to classify or index its content…'[80]; however, there can be exceptions when 'non-specialists' catalogue their own information.[81] Oliver and Harvey[82] emphasised the importance of the design of a metadata system to ensure that necessary information is collected from the outset and stored in a suitable format. They included consideration of the file names used and the structure of folders. In addition to ensuring that all the necessary metadata are captured from the outset, it is also highly beneficial for that data to be stored with interoperability in mind, so 'that digital objects can be successfully exchanged between computer systems…'.[83] An example of a metadata system is Dublin Core,[84] a set of 15 core elements used to describe resources but there are many other such schema.[85] Perhaps one challenge to using metadata to catalogue or describe personal information has been the lack of a single standard[86] relevant to personal information.

There is also an important lesson to be learnt from Dourish and colleagues who proposed to avoid the use of folders and eliminate the file duplication that occurs in a typical distributed system such as a home or work computer or network.[87] *Placeless* used file properties, such as the topic (*active properties*) in a metadata database. *Placeless* cached the contents of active files in order to make it faster, and allowed for Application Program Interfaces (APIs) so that it could communicate with a variety of other operating systems and programs.[88] *Placeless* also permitted different users to apply different properties to the same file. *Placeless Documents* was intended to cater for work flow, with the notion of active document properties, a form of metadata.[89] However, there was little uptake of *Placeless*. The failure of Microsoft's experiment with *Placeless Documents* was attributed to a lack of collaboration and its inability to interact with other systems.[90]

Karger discussed a range of file formats and approaches for managing personal records and proposed a unified database structure.[91] However, he cautioned that:

> The database community has argued for decades that we would all be better off storing all our personal information in (personal) databases. This clearly has not happened, most likely due to the apparent complexity of interacting with a database. No one has yet come forward with applications that hide the complexity of installing and maintaining a database, designing the schemas for the data to be stored and creating the queries that will return the desired information. And people seem generally allergic to having all their information presented to them as lists of tuples.[92]

Karger considered the issue of an address book format for a PIM system comprising a single file, and the challenge at the time of sharing single records within that with another application, something that can now often be resolved with APIs. Karger made the key observation that 'Agreeing on names for particular fields seems less demanding than agreeing byte-for-byte on file formats for all applications' data'.[93] Karger proposed the use of metadata to group and link

files, thereby removing the requirement for each file name to comprise a thorough description of the purpose and content of that file, nor for the database to interpret the file's content – while advocating that the database offers 'click-to-open' accessibility to the file. Metadata describing personal records may not be confined to a single flat database. IBM has been structuring databases into sets of inter-related data in order to be more succinct for more than half a century.[94] Kelly also described a file system that automatically maintained a limited range of metadata about each file, such as the creating software, size, and modification date; however, this was limited to attributes that the system could determine automatically.[95] One solution Jones suggested was to use long descriptive file names.[96] A key question is whether people are willing to put the effort in from the outset to label records or create metadata consistently and from the outset for their own personal records. As Marshall observed, disorder occurs over a long period of 'benign neglect',[97] and may only become apparent at the point that labelling or categorising records becomes an unappealing task. A study among Croatian university students found that nearly all of them organised their files into folders (97.4%), approximately half (53.3%) added metadata, and less than 1% (0.9%) used a tool to organise their digital information.[98]

In summary, the existing literature describes personal electronic records as records people choose to retain, often because they relate to transactions and the electronic records act as evidence of those transactions both for personal and legal use. Personal electronic records include records sent to or sent by people, about those people, or otherwise owned by those people. Oh's references to the purpose and use of documents raise a similar question of validity. Might an electronic document have the purpose of being the legal record of something? How is the document to be used? As noted above, a list of example documents has been published[99] but it falls short of fully embracing all possible forms of personal electronic records. Despite the volume of research in PIM, as we have noted, this primarily relates to information management in the workplace or among knowledge workers or students and does not inform us regarding the attributes of personal records in the home. More research is required to understand when and how personal electronic records are created and used, which records need to be retained, and why and how they can be categorised, so that we can develop systems to improve personal electronic records management. The purpose of this article is to identify the key attributes of personal electronic records which may be used in future to develop systems that may enable people to manage their personal records in the home. While there is extensive work on similar topics in other archives and records environments, especially workplaces and organisations, this work may be applicable in more niche environments staffed by volunteers or extremely small staffs where full archives and records management systems are not an option.

Method

The method for this research was drawn from the field of PIM using a 'guided-tour' method wherein the participant leads the researcher on a tour of their hardcopy personal records, showing a physical desktop or any other tools, as well as their electronic desktop and electronic tools with the researcher.[100] In this research we used an online virtual adaptation of the guided tour due to the requirement of maintaining social distance during the outbreak of COVID-19 in 2020.[101] Interviews and guided tours of participants records were conducted using Facebook Messenger, selected for both its audio and visual capability, and its availability to participants who were recruited through Facebook.

Data collection method

Twenty-two interviews were conducted, which involved 30 participants, as eight of the interviews were conducted with couples. As participants were recruited online via social media

(Facebook), they were therefore known to the researcher. A selection procedure screening potential interviewees was used to include a balance of people by gender and broad age group until saturation was reached, in that no new information was being contributed by participants. The sample size of participants was insufficient for quantitative analysis of their comments or observed behaviours. Nineteen of the interviews were conducted with people in Australia, two in the United Kingdom, and one in the United States, the latter interviews providing insight into alternative terminology used outside of Australia, such as 'pay-check' instead of 'pay slip', as well as some other minor variations. Nevertheless, the number of international participants was insufficient to draw conclusions as to any significant behavioural or practical differences from one country to another. Ethical approval was granted by the ethics panel of the first author's institution, subject to ensuring the privacy of the research participants and the confidentiality of the information they shared. Participants were invited to talk about their records including bank and credit card statements, receipts, insurance policies, vehicle registration documents, or any other records they received or retained. Where necessary, participants were prompted with examples of common records such as utility bills. Each participant was asked about how they received each of their regular records such as bills and statements, and the steps they took with each record type. They then shared video stream of their physical and electronic files and filing systems with the first author.

Analysis method

A database was created to analyse the data collected regarding records that participants retained. The database included a field for the method of delivery where applicable (electronic or hardcopy); a field for each of the steps that each record went through; and a field to record where the record was finally saved, if at all. In some cases, the information that populated the fields of the database were not verbally expressed but they were implied. For example, participants did not need to explain that a gas bill came from the gas company. Bills necessarily included an amount and a date due. Emails always have a sender, receiver, and usually a subject line, as well as their contents. Using this information, relevant fields were added to the database to represent the owner of the record, the sender (when applicable), the subject, and so forth. Fields were also created for each step or action applied to the record, such as saving the contents, forwarding the record, making a payment, or any other action.

The description of each record was 'cleaned' and a consistent set of labels was created for synonyms. For example, the phrases 'documents for tax', 'tax stuff', and 'tax documents' were standardised to 'tax documents'. Additional fields were then added to the database to describe (1) the practical event that caused the creation of the record so that future records could be anticipated and (2) the validity of the record in terms of legal standing, so that records management tasks could be prioritised. All categorisation fields were developed using an iterative process for categorising text responses.[102] To categorise the records according to causation (by which we mean what caused them to be created), we looked at the first record in our list. For example, in the case of a regular bill, the practical creation of the record was the arrival of the date for the issue of that bill. Using the iterative method, we looked at the next record. Was it also created by the arrival of a date? And if not, what gave rise to it: a decision to take an action? Each record was evaluated to decide whether it fitted the categories already established or required a new category. Our test of the reliability of this approach was the expectation that someone else tackling the same task would replicate the results.

To categorise the validity of each record, we asked a set of evaluative questions: is this the only version of the record? Does the record stand up in its own right or does it need to be validated in another way? If it needs to be validated, what was required to validate it? As the

30 participants in our research collectively described 489 records, we were able to analyse the database quantitatively, as has been done in previous studies with comparable samples.[103] In this article, we focused exclusively on the question of categories of electronic records described by participants in the study.

Analysis and findings

We began the analysis by addressing the elements that comprise a personal electronic record. We then explored the categorisation of personal electronic records, what caused the records to come into existence at inception, and, finally, the legal validity of personal electronic records.

What is in a personal electronic record?

By examining the contents of the database created during the guided tour interviews, we determined the key elements that form personal records, thus creating the metadata fields that may be most useful in describing personal records. These metadata fields were then tested against all the records in our database to ensure their relevance and check for omissions. Six fields were required to describe the core elements of every record and they are:

All records
  (1)   The record owner, and who the record relates to (may be several people)
  (2)   Records categories (there may be several)
  (3)   A record subject, such as an account, dwelling, vehicle, or person
  (4)   A creation date
  (5)   Content of the record – what is the record about?
  (6)   If the metadata are not stored with the record itself, then the location of the record and a hyperlink to that record are required.

Additional fields are required for other records, for example, transactions, appointments, and the documentation of journeys, which are listed below.

Transactions
  (1)   The record creator, which may be an account, supplier, or the author/sender of an email
  (2)   An account number or reservation number
  (3)   Causal event, such as a bill cycle, if and when necessary
  (4)   Additional dates and times, such as a due day for bills, reminder date or check-in time for a flight, expiry date of a membership, etc.
  (5)   Location details for appointments, events, and start point for journeys
  (6)   A transaction amount and currency
  (7)   Related tax amount
  (8)   Whether an item is paid, how paid and when paid
  (9)   A receipt number
  (10) Other notes or relevant hyperlinks

Additional for records of an event
  (1)   A start and end date (and possibly time and time zone) related to the record, such as a billing period, or the start and end of a journey

Additional for records related to a journey
  (1)   Destination location

Additional for transmitted records such as emails and text messages
  (1) Sender and receiver
  (2) Date and time sent and received
  (3) Other recipients (CCs), delivery receipt, whether the record was read, what was attached, and other metadata, such as flags and subject line.

User-subjective categorisation of personal electronic records

The outcome of the analysis of the records gave rise to 88 detailed categories based on the purpose and content of the records, such as phone bills, rental documents, warranties, and travel tickets. These categories were further combined into 13 (overlapping) broader groups of categories, suggested by the way in which participants described the records they discussed. These broader categories comprised:

  (1) Advertising/brochures (that people wish to keep)
  (2) Music, for example, downloaded music
  (3) Interests, for example, hobbies, personal development, recommendations, restaurants, movies, books, and volunteering
  (4) Personal budget, for example, a spreadsheet or budgeting application
  (5) Photos
  (6) Travel and tickets
  (7) Receipts
  (8) To do lists/reminders, for example, addresses, appointments, birthdays list, change of address notification list, credit card numbers, exercises, packing list, passwords, recommendations, restaurants, movies, books, shopping lists, and sports registration numbers
  (9) Study documents
  (10) Bills and statements, for example, bank statements, council rates, credit card statements, electricity bill, statements for toll road transponders (called etag in Australia, equivalent to US E-ZPass), garbage bills, household bills (no further information), Internet subscription bills, online subscriptions, bills (no further information), phone bills, store cards, natural gas bills, subscriptions, water and sewage rates
  (11) Documents that 'need to be retained' (user-subjective definition) , for example, bicycle insurance, car loan, correspondence – non personal, deeds, health insurance documents, house sale/renovations, insurance (no further information), investment documents, job-seeking documents, lease on house, liability insurance, manuals, motor vehicle documents including insurance, pet documents, property acquisition documents, recipes, renovations, rental documents, strata/managing agent documents, superannuation statement / documents, television licence, warranties, work documents/registration
  (12) Income and tax documents, for example, pay slips/pay stubs, payment notices, property tax, rental statements, work invoices, shares-related documents, pension documents, tax documents
  (13) Personal documentation, for example, baptism certificates, birth certificates, personal correspondence, certificates/licences, driver licences, family history, health records, identity documents, children's documents, marriage certificates, memorabilia, motoring association memberships, passports, resumes, school and sports reports, sports club records, visa documents and wills.

Many records fitted into several categories with the same functionality. For example, personal documents may include scans of passports, birth certificates and educational certificates.

For another example, consider the case of households with more than one person, notionally person 'A' and person 'B'. In some households, documents were sorted in a shared folder by category, including similar documents for person A and person B in the same folder. In others, documents were stored separately for person A and person B. In other words, the structure of categorisation that participants used varied. Further, when people look for documents in a hierarchical structure, a different approach is used and needed, depending on the structure that each person adopts.

The causal inception of personal information and documents

An additional form of categorisation was applied to the record codes identified in this research: what is it that gives rise to the existence of the record? For instance, regular bills are generated on a specific date. When that date occurs, the bill is created. Irregular bills are caused by a decision to purchase something, or an unexpected occurrence that created a cost. It was found that every personal record is generated by one or more of five circumstances: either an event occurs; an event is anticipated; a transaction occurs; a date is arrived at; or the item is generated by interest. From these another set of categories was created, which describes the determinant circumstances. Just as the record categories are not mutually exclusive, accordingly, the inception categories overlap. Records can be initiated by several antecedent circumstances:

(1) Anticipated future events, for example, travel, a planned purchase such as a car, appliance, or home
(2) Dates, for example, the billing date of an account, a due date for payment, or a birthday
(3) Events, for example, an accident, school re-union or receiving an award
(4) Interests, for example, cooking, destination, or artistic pursuit
(5) Transactions, for example, a purchase, or execution of a contract

We refer to these categories of events that cause the creations of the records as the causal categories of personal records. Understanding causal categories helps us to recognise what occurrences are likely to generate personal electronic records, how to begin to categorise them, and to improve how people deal with them.

The validity of personal electronic records

A second iterative categorisation process was used to classify each of the records in the database with regard to its purpose or legal validity. Categories were established in consideration of the reviewed literature to determine the legal status of the record. Did the record relate to a transaction? Was the record an electronic copy of a paper record? And if not, what was the legal validity of each record? How could the record be authenticated? Six categories were established from this process that described the legal validity of the records in our database. These categories were:

(1) Electronic assertions of legal validity: a digital document embedded in an application (usually in lieu of a physical equivalent) making an assertion (usually about a person's identity or rights) such as a driver's licence, Medicare card, or electronic credit card.
(2) Electronic records matching a transaction, such as an electronic receipt, pay slip, or insurance certificate. This includes scans or photographs of original paper documents (which may then be discarded). The electronic versions of these generally have transaction numbers and are authenticated by a matching transaction, such as a payment. These electronic records comprise the official record of the transaction.

(3)   Original electronic documents that make assertions not related to a transaction, such as academic, literary, or artistic works. These items are not copies of hardcopy. Their content is intellectual or artistic, rather than representing a transaction or a legal status.

(4)   Electronic records reflecting an agreement taken on trust, such as a digital document making an assertion that does not have a matching transaction because it comprises an agreement or promise, typically relating to a foreseen future event. The electronic form of these documents is accepted in lieu of a paper document, such as an unsigned contract or a letter of offer, quotation, or acceptance of a quotation. These are accepted on trust that, if necessary, a paper version can be created and, in some cases, signed in ink. They may not be substantiated by a transaction, usually because the transaction is yet to occur.

(5)   Electronic copy of a hardcopy document, either comprising the unsigned version of an agreement or a scan of a signed agreement.

(6)   Records that make no assertions and do not relate to transactions. Examples include recipes, manuals, and letters (although depending on the content of the letter, it may belong into the second category above). Many of these may function as a memory aid.

The above list can be interpreted as a hierarchy of legal validity for electronic documents, from the first item which has the highest level of legal validity to the sixth item that has the least legal validity. As in the cases of the previously discussed categories of personal electronic documents, there may be overlaps and grey areas among validity categories. For instance, a record of a deposit in part payment of an item may be both a record of a transaction and a future agreement to complete the purchase. E-signatures on documents may be recognised as adding legal validity in certain circumstances but may add little legal validity in other jurisdictions. Wet signatures may be required on certain kinds of documents, but not others. The legal validity of an electronic record is unlikely to be recorded in a metadata field despite the importance of understanding and being aware of the legal validity of each personal document. Once the notion of legal validity for each type of record is understood, it is not necessary to document this with each record.

In summary, we find that every personal electronic record can be categorised in several ways, each of which adds depth to our understanding of the nature of that record. The primary categorisation of personal records relates to the subjective categories that the research participants used to describe records. These categories essentially describe what the personal electronic records are about – for example, whether the records are bills, travel documents, income and tax documents, or documents that people feel that they need to retain. These user-subjective categories have blurred boundaries and records can fit into more than one category.

The second form of categorisation relates to the inception of personal electronic records. Personal electronic records are born of either the anticipation of a future event, the arrival of a date or a certain event, the adoption of an interest, or as documentation of a transaction.

A third categorisation of personal electronic records groups them for the purpose of prioritisation according to legal validity. Some personal electronic records can make a legal assertion without a hardcopy version ever being created. Other records comprise the documentation of a transaction, and are verified by the existence of that transaction, such as a payment. Still others may make an assertion that is taken on trust or make no assertion. Collectively, categorising personal electronic records using these three attributes describe what the record is about, how it came about, and the nature of what it says in terms of its legal validity. It is also

possible to identify a set of fields that describe records, and which could be used to establish a metadata database that describes personal electronic records.

Discussion

In this research we have identified three attributes of personal electronic records that assist in understanding the nature of the record, namely, the subjective categories that people apply to their personal information and documents; the circumstances that caused the record to come into existence; and the legal validity that can be attributed to the personal electronic record. We use the term 'attributes' as opposed to Hider's 'data elements'[104] because attributes such as the cause of the creation of the record may not be in the records themselves. Collectively, understanding these attributes could help people to anticipate the creation of future records, determine the importance of keeping the record and choosing how to categorise the record, and thus re-finding the record if necessary.

It has been established that there is a need for people to retain personal information and documents, including their personal electronic records. The long-term curation of personal electronic archives is an established area of study, but there is also a need for the disciplined management of current active personal electronic records, in order to ensure effective day-to-day management of people's personal affairs – not the least because many of these electronic records have acquired legal recognition. Our study provides the foundation for analysis to determine a greater understanding of personal electronic records in the home.

Analysis of the 489 records that participants in the research described found that, in addition to the well-understood attributes of personal electronic records, such as the date, form, and who they are from or to (if applicable), records can be categorised into several ways. Firstly, there is a user-subjective categorisation that builds on Smith,[105] adopting the terminology that the item's owners used to describe their records, such as bills, receipts, study documents, and so forth. Keeping personal records in user-defined category groups can improve people's ability to re-find their own records, but conflicts with the benefit of having a standardised category vocabulary for all users.[106] A single major category can help people to save and re-find records in a hierarchical folder structure and may lend itself to standardisation more readily than a multi-tagging system. Additional categorisation may provide improved search results for people who choose to find records using search applications.

An additional, novel, level of categorisation can also be applied to personal electronic records as each record has an identifiable causal circumstance, comprising an anticipated future event, a date, the occurrence of an event, an interest, or the execution of a transaction or registration. For example, if a person decides to take a vacation, they may start collecting brochures about potential destinations. They may purchase a ticket that needs to be retained until the time of travel. Shortly before departure they may obtain a boarding pass for a flight. These are foreseeable records instigated by the anticipation of a vacation, the transaction of booking a flight, and the arrival of the date of that journey. Another example consists of regular bills usually generated on a regular specific date, such as each month. An interest in a hobby or club typically results in a collection of records related to that interest or activity. The key benefit of this form of categorisation is that it gives rise to valuable predictions about the generation of future records. Regular bills can be expected on certain dates. Specific actions will generate a predictable set of records. Knowing what records to expect in the future is useful for budgeting and other purposes – but most particularly for identifying missing records. By predicting what records to expect, people can be notified when a required record does not occur. Such a prediction would significantly minimise overlooked bills and late bill payment fees and missing documentation. A reminder system to retain a complete set of pay slips or

pay stubs, as well as other documentation may have significantly reduced the burden placed on individuals by Australia's 'robo-debt' scheme. The recognition and systemisation of predictable personal electronic record generation lends significant value to the secondary categorisation of personal electronic records in terms of the circumstances of their creation.

This research identified a further categorisation of personal electronic records also inspired by Smith,[107] in terms of the form of the legal validity of the record. This differs from the 'authenticity' of records described by Hider[108] because many records may be entirely authentic but lack legal validity. This form of record categorisation is particularly useful for identifying the importance of specific records. For instance, original electronic documents need to be retained and backed up very thoroughly, as they are not substantiated by any other records or in other forms. If such documents are lost, they may be irreplaceable. The particularly interesting category comprises electronic records taken on trust. This group of documents is increasingly replacing paper documents both in practical terms and in legal recognition. Understanding the legal validity of personal electronic records contributes significantly to understanding which records need to be maintained in various formats and helps identify important records that need to be backed up or secured. To re-visit the example of the Australian Robo-debt incident, it is apparent that a well-managed application of these categories could have increased peoples' ability to prove their eligibility to receive social security at the time.

Metadata fields for personal records

In addition to attributes of personal records described above, each record contains information. Unless those records are in a structured database, there are elements of information in the records that people need to be able to easily extract. As Karger[109] pointed out, there are considerable benefits to ordering personal information in a structured metadata format. Due to the unlimited range of information that personal records may contain, a definitive list of personal records metadata fields could never be complete. By examining our database of personal records, we have been able to identify a set of fields that would likely be most useful. Nevertheless, a metadata database for personal electronic records may also have related objects, such as an index of categories, or sub-records. For instance, a single invoice or income statement may have several items on it that could form sub-records. A travel reservation record may list several flights, each meriting a record in its own right. Equally, there is also the issue of records that exist in several different files. For instance, an insurance renewal may comprise an insurance document, a separate invoice and separate product disclosure document, or terms and conditions. These parts of the record need to be linked so that someone who searches for the record will find all the components. Determining the fields (or data elements) that would comprise a personal records metadata system would be an endless task. However, our research indicates that a limited set of fields would be almost ubiquitous for the most common types of records, such as owner, date, subject, and some description of the contents of the record, some of which are described by Hider.[110] A common labelling of such fields would allow records to be shared between personal records management systems, thereby providing some standardisation which would address some of the problems in saving and re-finding records as well as other possible personal records management functions. This differs from the adoption of a standardised vocabulary for categories and records within the fields. Standardising categories or labels for records would be more difficult because our research showed that different people use different terminology for the same types of records. People also re-categorise records from time to time unpredictably. Hence, the field (or data element) benefits from a standardised field label and format, even if the content of the field is not from a standardised vocabulary.

In summary, comparing the types of documents described by Smith,[111] our research found three ways of categorising personal electronic records, each of which adds depth to Smith's examples: firstly, the user-subjective categories into which people sort their records; secondly, categorisation in terms of the causal events that gave rise to the records; and, thirdly, the different levels of legal validity of personal electronic records. We also propose that some form of metadata could be used to provide a level of consistency in personal records databases. In combination, these findings dove-tail with Smith's observations, adding another layer of depth, and a further step towards improving the management of personal electronic records in the home.

Conclusion

In this research we identified a set of metadata elements that we can usually expect to find in relation to personal electronic records, such as the record owner, its creation date, and in the case of records of transactions, various details of those transactions. Some of this information is stored within the record, and some is stored in descriptive metadata about the record.

This research has also identified three attributes of personal electronic records that contribute to our understanding of these records and provide for improvements in personal electronic records management in the home. The most recognisable form of categorisation of personal electronic records comprises user-subjective topics, such as bills and receipts, records related to income and taxation, travel records, interests, and music. These categories are organic and sometimes overlapping categories are the primary way that people categorise their personal electronic records. A second attribute of personal electronic records comprises the circumstances of their creation, such as an event or the occurrence of a date. Understanding the conditions that give rise to new electronic records is useful for predicting when new records will be created and invaluable for ensuring that records are not overlooked. A third attribute of personal electronic records comprises the records' validity. Many electronic records have substantial legal validity and may comprise the only record that describes certain transactions. Further records are taken on trust, assuming that a paper copy with a wet signature could be provided if required and still other records have little legal validity, such as notes and memory aids. Recognising the legal validity of personal electronic records helps identify the importance of keeping certain kinds of records, as well as ensuring that irreplaceable records are backed up.

Considering these types of categorisations in combination and knowing what records to expect (such as regular pay slips) would alert people to ensure that they received each payslip to which they were entitled. Understanding that the electronic payslips were the only record of payment and that these were of legal significance and validity could encourage people to save these. Together, these three attributes of personal electronic records, their primary user category, what causes their formation, and their legal validity lend considerable insight to the understanding of these records and provide a groundwork for future research and the development of improved systems for the management of personal electronic records.

Notes on contributors

Matt Balogh is currently a PhD candidate. He has 35 years' experience conducting research, including as general manager of Quadrant Research, 2 years as vice president of the Gallup Organisation, followed by 17 years as managing director of McNair yellowSquares. Matt is a fellow of The Research Society (TRS), was the Best Practice trainer for TRS, has been a member of the Professional Conduct Review Committee and is currently on the Qualified Practicing Researcher assessment committee. Matt has overseen the implementation of over

5,000 research projects, including the NSW Health survey, surveys for the Department of Prime Minister and the Cabinet (PM&C), Finance, Human Services, The Australian Taxation Office (ATO) and more.

Dr William Billingsley is an associate professor in Computational Science at the University of New England. William's research interests broadly include human–computer interaction, software design, and applied computer science: 'smart useful systems'. He has particular interests in computing education, smart and social educational technology, and computer-supported cooperative work. https://twitter.com/wbillingsley.

Dr David Paul is a senior lecturer in Computational Science at the University of New England, Australia. David's research interests include computer networks and distributed systems, security and privacy, and applied computer science in areas such as health and agriculture. He has experience with nation-wide research projects such as the Australian Schizophrenia Research Bank, the QuON Web survey system, and the ASKBILL farmer prediction system.

Dr Mary Anne Kennan is an adjunct associate professor in the School of Information and Communication Studies at Charles Sturt University. Mary Anne's research interests focus broadly on scholarly communication including open access and research data management; the education of, and roles for, librarians and information professionals; and the practices of information sharing and collaboration in various contexts. She is the editor of the *Journal of the Australian Library and Information Association* (JALIA). http://twitter.com/MaryAnneKennan.

ORCID
Matt Balogh
Dr William Billingsley
Dr David Paul
Dr Mary Anne Kennan

Notes

1. Community Affairs Reference Committee, The Senate of the Commonwealth of Australia, Centrelink's Compliance Program: Second Interim Report, 2020, available at https://parlinfo.aph.gov.au/parlInfo/download/committees/reportsen/024338/toc_pdf/Centrelink'scomplianceprogram.pdf;fileType=application%2Fpdf, accessed 12 March 2021.
2. Casey Tonkin, 'Court Finds Robodebt Unlawful', Information Age, 28 November 2019, available at https://ia.acs.org.au/article/2019/court-finds-robodebt-unlawful.html, accessed 21 December 2019.
3. Australian Taxation Office, 'Records You Need to Keep', 2022, available at https://www.ato.gov.au/uploadedFiles/Content/IAI/Downloads/Toolkits/TaxTimeToolkit_Recordsyouneedtokeep.pdf, accessed 5 May 2022.
4. Matt Balogh, Mary Anne Kennan, William Billingsley and David Paul, 'Understanding the Management of Personal Records at Home: A Virtual Guided Tour', Information Research, vol. 27, no. 2, 2022, available at http://informationr.net/ir/27-2/paper926.html, accessed 6 May 2022; Paris Buttfield-Addison, Understanding and Supporting Personal Information Management Across Multiple Platforms. (PhD). University of Tasmania, Hobart, Tasmania, Australia, 2014, available at https://eprints.utas.edu.au/18564/, accessed 1 May 2022; Bryan Kalms, 'Household Information Practices: How and Why Householders Process and Manage Information', Information Research, vol. 13, no. 1. 2018, available at http://informationr.net/ir/13-1/paper339.html, accessed 4 May 2022.
5. Ofer Bergman, Tamar Israeli and Steve Whittaker, 'Factors Hindering Shared Files Retrieval', Aslib Journal of Information Management, vol. 72, no. 1, 2020, pp. 130–147, doi: 10.1108/ajim-05-2019-0120.

6. Harry Bruce, William Jones and Susan Dumais, 'Information Behaviour that Keeps Found Things Found', Information Research: An International Electronic Journal, vol. 10, no. 1, October 2004, available at http://www.informationr.net/ir/10-1/paper207.html, accessed 3 April 2022.

7. Sarah Henderson, 'How Do People Organize Their Desktops?', Extended Abstracts on Human Factors in Computing Systems, 24 April 2004, doi: 10.1145/985921.985972; Sarah Henderson, 'Personal Digital Document Management', Paper Presented at the APICHI 2003 – Asia-Pacific Conference on Computer-Human Interaction, doi: 10.1007/978-3-540-27795-8_72; Sarah Henderson, 'Genre, Task, Topic and Time: Facets of Personal Digital Document Management', CHINZ '05: Proceedings of the 6th ACM SIGCHI (Association of Computing Machinery Special Interest Group on Computer-Human Interaction) New Zealand Chapter's International Conference on Computer-Human Interaction: Making CHI Natural, July 2005, doi: 10.1145/1073943.1073957; Sarah Henderson, 'Personal Document Management Strategies', CHINZ '09: Proceedings of the 10th International Conference NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction, July 2009, doi: 10.1145/1577782.1577795; Sarah Henderson and Ananth Srinivasan, 'An Empirical Analysis of Personal Digital Document Structures', Paper Presented at the Symposium on Human Interface Human Interface 2009: Human Interface and the Management of Information. Designing Information Environments, 2009, doi: 10.1007/978-3-642-02556-3_45.

8. Barbara Kwasnik, 'How a Personal Document's Intended Use or Purpose Affects Its Classification in an Office', Paper Presented at the ACM SIGIR Forum (Association of Computing Machinery) Special Interest Group on Information Retrieval), 1 May 1989, doi: 10.1145/75335.75356.

9. Kyong Eun Oh, 'What Happens Once You Categorize Files into Folders?', Proceedings of the American Society for Information Science and Technology, vol. 49, no. 1, 2012, pp. 1–4, doi: 10.1002/meet.14504901253; Kyong Eun Oh, 'The Process of Organizing Personal Information', ProQuest Dissertations Publishing, 2013, doi: 10.7282/T31N7Z5F; Kyong Eun Oh, 'Types of Personal Information Categorization: Rigid, Fuzzy, and Flexible', Journal of the Association for Information Science and Technology, vol. 68, no. 6, 2017, pp. 1491–1504, doi: 10.1002/asi.23787; Kyong Eun Oh, 'Personal Information Organization in Everyday Life: Modeling the Process', Journal of Documentation, vol. 75, no. 3, 2019, pp. 667–691, doi: 10.1108/JD-05-2018-0080; Kyong Eun Oh and Nicholas Belkin, 'Cross Analysis of Keeping Personal Information in Different Forms', ACM International Conference Proceeding Series, 8 February 2011, pp. 732–733, doi: 10.1145/1940761.1940888.

10. William Jones, Keeping Found Things Found: The Study and Practice of Personal Information Management, Morgan Kaufmann, Burlington, 2008, p. 2.

11. Ibid., p. 5.

12. Joshua Finnell, 'Records Management Theory's Dilemma: What Is a Record?', Library Philosophy and Practice, p. 2, 2011, available at http://hdl.handle.net/10760/18977, accessed 4 February 2022; Geoffrey Yeo, Records, Information and Data: Exploring The Role of Record-Keeping in An Information Culture, Facet Publishing, London, 2018, p. 73, doi: 10.29085/9781783302284 – referencing The Sedona Conference, 2007, p. 3.

13. Finnell, p. 2.

14. Society of American Archivists, 'Record', 2020, available at https://www2.archivists.org/glossary/terms/r/record, accessed 28 March 2020.

15. Sue McKemmish, 'Evidence of Me', Australian Library Journal, vol. 45, no. 3, pp. 178 and 184, doi: 10.1080/00049670.1996.10755757.

16. Jordan Bass, 'A PIM Perspective: Leveraging Personal Information Management Research in the Archiving of Personal Digital Records', Archivaria – The Journal of the Association of Canadian Archivists, vol. 75 (Spring 2013), pp. 49 and 54, available at https://archivaria.ca/index.php/archivaria/article/view/13433, accessed 13 February 2022.

17. Adrian Cunningham, 'Waiting for the Ghost Train: Strategies for Managing Electronic Personal Records Before It Is Too Late', Archival Issues, vol. 24, no. 1, 1999, p. 55, doi: 10.2307/41102007.

18. Charles Ragan, Jonathon Redgrave and Lori Ann Wagner, 'The Sedona Guidelines: Best Practice Guidelines & Commentary for Managing Information & Records in the Electronic Age', in Lori Ann Wagner (ed.), The Sedona Conference Second Edition November 2007, Arizona, 2007, available at https://thesedonaconference.org/node/7953, accessed 6 April 2022.

19. Sarah Kim, 'Landscape of Personal Digital Archiving Activities and Research', in Donald T. Hawkins (ed.), Personal Archiving: Preserving Our Digital Heritage, Information Today, Medford, available at https://www.worldcat.org/title/personal-archiving-preserving-our-digital-heritage/oclc/856579475, accessed 24 April 2022.

20. Ibid.

21. Bernadette Bosse, 'Electronic vs. Digital Data', 2015, available at https://www.linkedin.com/pulse/electronic-vs-digital-data-bernadette-bosse, accessed 15 April 2021.
22. Glenda Acland, 'Glossary', in Judith Ellis (ed.), Keeping Archives, 2nd Edition, Thorpe in Association with the Australian Society of Archivists, Port Melbourne, 1993, p. 469.
23. David Bawden and Lyn Robinson, 'Basic Concepts of Information Science', in David Bawden (ed.), Introduction to Information Science, Facet Publishing, London, 2012, p. 77.
24. Geoffrey Yeo, Records, Information and Data: Exploring the Role of Record-Keeping in an Information Culture, Facet Publishing, London, 2018, p. 73; Gillian Oliver and Fiorella Foscarini, Records Management and Information Culture: Tackling The People Problem, Facet Publishing, London, 2014, p. 20.
25. Michael Buckland, 'What Is a "Document"?', Journal of the American Society for Information Science and Technology, vol. 48, no. 9, 1997, p. 807, doi: 10.1002/(SICI)1097-4571(199709)48:9%3C804::AID-ASI5%3E3.0.CO;2-V.
26. David Roberts, 'Defining Electronic Records, Documents and Data', Archives and Manuscripts, vol. 22, May 1994, pp. 14–26, available at https://search.informit.org/doi/10.3316/ielapa.950605258, accessed 18 December 2021.
27. Balogh, Billingsley, Paul and Kennan.
28. Susan Marilyn McKemmish, Franklyn Herbert Upward and Barbara Reed, 'Records Continuum Model', in Marcia J. Bates and Mary Niles Maack (eds.), Encyclopedia of Library and Information Sciences, 3rd Edition, Taylor & Francis, London, 2010, pp. 4447–4459.
29. Sue Myburgh, 'Records Organization and Access', in Marcia J. Bates and Mary Niles Maack (eds.), Encyclopedia of Library and Information Sciences, 3rd Edition, CRC Press, Boca Raton, p. 4460, doi: 10.1081/E-ELIS3.
30. Manuel Zacklad, 'Information Design: Textualization, Documentarization, Auctorialization', Proceedings from the Document Academy, vol. 6, no. 1, 2019, p. 2, doi: 10.35492/docam/6/1/2.
31. Sue McKemmish, Glenda Acland, Nigel Ward and Barbara Reed, 'Describing Records in Context in the Continuum: The Australian Recordkeeping Metadata Schema', Archivaria, vol. 48, Fall 1999, pp. 3–37, available at https://archivaria.ca/index.php/archivaria/article/view/12715, accessed 15 March 2022.
32. Manuel Alba, 'Order Out of Chaos: Technology, Intermediation, Trust, and Reliability as the Basis for the Recognition of Legal Effects in Electronic Transactions', Creighton Law Review, vol. 47, no. 3, 2013, p. 390.
33. Mustafa Emre Civelek, Nagehan Uca and Murat Çemberci, 'eUCP and Electronic Commerce Investments: e-Signature and Paperless Foreign Trade', Eurasian Academy of Sciences, Eurasian Business & Economics Journal, vol. 3, no. 3, October 2015, p. 61, doi: 10.17740/eas.econ.2015.
34. The Office of Legislative Drafting and Publishing, Attorney-General's Department, Australia, 'Electronic Transactions Act', Canberra, 1999, available at https://www.legislation.gov.au/Details/C2011C00445, accessed 9 March 2022.
35. Electronic Commerce Act, 2000, S.O. 2000, clause. 17, 2020, available at https://www.ontario.ca/laws/statute/00e17; Electronic Communications Act – UK, 2000, available at http://www.legislation.gov.uk/ukpga/2000/7/contents accessed 2 February 2022; Electronic Records and Signatures in Commerce Act – USA, 2000, available at https://www.govinfo.gov/content/pkg/PLAW-106publ229/pdf/PLAW-106publ229.pdf, accessed 11 March 2022; The European Parliament and The Council of 23 July 2014, on Electronic Identification and Trust Services for Electronic Transactions in the Internal Market and Repealing Directive 1999/93/ EC, 'Regulation (EU) No 910/2014', 2014, available at https://eur-lex.europa.eu/legal-content/EN/ TXT/?uri=uriserv%3AOJ.L_.2014.257.01.0073.01.ENG, accessed 5 January 2022
36. Claudio Arruda, Mark Tibberts and Faye Williams, 'The Law of E-Signatures in the United States and Canada', 2020, available at https://www.bakermckenzie.com/en/insight/publications/2020/03/the-law-esignatures-us-canada, accessed 16 July 2020.
37. SuperCentral, 2021, available at https://www.supercentral.com.au/smsf-news/995/electronic-witnessing-of-documents/, accessed 31 March 2021.
38. Myburgh.
39. Zacklad.
40. Alba.
41. Jones, 2008, p. 43.
42. Ibid., p.43.
43. Barry Smith, 'The Ontology of Documents', Paper Presented at the Conference on Ontology and Analytical Metaphysics, Tokyo, 2011, available at http://ontology.buffalo.edu/smith, accessed 11 February 2022.
44. Ibid., p. 3.

45. Barry Smith, 'How to Do Things with Paper: The Ontology of Documents and the Technologies of Identification', Paper Presented at the Ontolog, virtual location, 2005, p. 19, available at http://ontolog. cim3.net/wiki/ConferenceCall_2005_10_13.html, accessed 20 February 2022.

46. Ibid., p. 38.

47. Bergman, Whittaker and Tish, 2021.

48. Ibid., p. 2.

49. Jonathan Furner, 'The Ontology of Documents, Revisited', *Proceedings from the Document Academy, vol. 6, no. 1, 2019, p. 20, doi:* 10.35492/docam/6/1/1.

50. Oh, 2013, p. 215.

51. Deborah Barreau and Bonnie A. Nardi, 'Finding and Reminding: File Organization from the Desktop', ACM (Association for Computing Machinery) SIGCHI (Special Interest Group on Computer Human Interaction) Bulletin, vol. 27, no. 3, 1995, pp. 39–43, doi: 10.1145/221296.221307.

52. Oh, 2013, p. 217.

53. Oh, 2013, p. 221.

54. Oh, 2013, p. 227.

55. Oh, 2013, p. 223.

56. Oh, 2013, p. 224.

57. Kyong Eun Oh and Nicholas Belkin, 'Understanding What Personal Information Items Make Categorization Difficult', Proceedings of the American Society for Information Science and Technology, vol. 51, no. 1, 2014, pp. 1–3, doi: 10.1002/meet.2014.14505101139.

58. Eleanor Rosch and Barbara Lloyd, Cognition and Categorization, Lawrence Erlbaum Associates, Hillsdale, 1978.

59. Ofer Bergman, Ruth Beyth-Marom and Rafi Nachmias, 'The User-Subjective Approach to Personal Information Management Systems Design: Evidence and Implementations', Journal of the American Society for Information Science and Technology, vol. 54, no. 9, 2003, pp. 872–878, doi: 10.1002/asi.10283.

60. Ibid.

61. Eleanor Rosch and Barbara Lloyd, Cognition and Categorization, Lawrence Erlbaum Associates, Hillsdale, 1978; Eleanor Rosch and Carolyn B. Mervis, 'Family Resemblances: Studies in the Internal Structure of Categories', Cognitive Psychology, vol. 7, no. 4, 1975, pp. 573–605, doi: 10.1016/0010-0285(75)90024-9.

62. Oh, 2013, p. 9.

63. William Jones, 'Finders, Keepers? The Present and Future Perfect in Support of Personal Information Management', First Monday, vol. 9, no. 3, 2004, p. 314, doi: 10.5210/fm.v9i3.1123.

64. Jones, 2008, p. 315.

65. Jesse David Dinneen and Charles-Antonie Julien, 'The Ubiquitous Digital File: A Review of File Management Research', Journal of the Association for Information Science and Technology, vol. 71, no. 1, 2020, doi: 10.1002/asi.24222; Sarah Henderson, 'Document Duplication: How Users (Struggle to) Manage File Copies and Versions', Paper Presented at the Proceedings of the American Society for Information Science and Technology (ASIS&T), New Orleans, 2011, doi: 10.1002/meet.2011.14504801013; Sarah Henderson and Ananth Srinivasan, 'Filing, Piling & Structuring: Strategies for Personal Document Management', Paper Presented at the 2011 44th Hawaii International Conference on System Sciences, doi: 10.1109/HICSS.2011.205; Oh, 2012.

66. Nehad Albadri, Richard Watson and Stijn Dekeyser, 'TreeTags: Bringing Tags to the Hierarchical File System', Paper Presented at the Proceedings of the Australasian Computer Science Week Multiconference, Canberra, 2016, pp. 1–10, doi: 10.1145/2843043.2843868; Ofer Bergman, Noa Gradovitch, Judit Bar-Ilan and Ruth Beyth- Marom, 'Folder Versus Tag Preference in Personal Information Management', Journal of the American Society for Information Science and Technology, vol. 64, no. 10, October 2013, pp. 1995–2012, doi: 10.1002/asi.22906; Ofer Bergman, Noa Gradovitch, Judit Bar-Ilan and Ruth Beyth-Marom, 'Tagging Personal Information: A Contrast between Attitudes and Behavior', Proceedings of the American Society for Information Science and Technology, vol. 50, no. 1, 2014, pp. 1–8, doi: 10.1002/meet.14505001029; Pierre Fastrez and Jerry Jacques, 'Managing References by Filing and Tagging', Paper Presented at the International Conference on Human Interface and the Management of Information, July 2015, pp. 291–300, doi: 10.1007/978-3-319-20612-7_28; Qin Gao, 'An Empirical Study of Tagging for Personal Information Organization: Performance, Workload, Memory, and Consistency', International Journal of Human-Computer Interaction, vol. 27, no. 927(9), 2011, pp. 821–863, doi: 10.1080/10447318.2011.55530.

67. Ofer Bergman, Tamar Israeli and Yael Benn, 'Why Do Some People Search for Their Files Much More than Others? A Preliminary Study', Aslib Journal of Information Management, April 2021, doi: 10.1108/

AJIM-08-2020-0250; Ofer Bergman, Tamar Israeli and Steve Whittaker, 'Search Is the Future? The Young Search Less for Files', Proceedings of the Association for Information Science and Technology, vol. 56, no. 1, January 2019, pp. 360–363, doi: 10.1002/pra2.29.

68. Mihajlo Grbovic, Guy Halawi, Zohar Karnin and Yoelle Maarek, 'How Many Folders Do You Really Need? Classifying Email into a Handful of Categories', Paper Presented at the Proceedings of the 23rd ACM International Conference on Information and Knowledge Management, Shanghai, China, 2014, doi: 10.1145/2661829.2662018; Jacek Gwizdka, 'Email Task Management Styles: The Cleaners and the Keepers', Paper Presented at the CHI'04 Extended Abstracts on Human Factors in Computing Systems, Vienna, Austria, 2004, doi: 10.1145/985921.986032; Steve Whittaker, Victoria Bellotti and Jacek Gwizdka, 'Email in Personal Information Management', Communications of the ACM (Association for Computing Machinery), vol. 49, no. 1, January 2006, pp. 68–73, doi: 10.1145/1107458.1107494; Steve Whittaker, Victoria Bellotti and Jacek Gwizdka, 'Email and PIM: Problems and Possibilities', available at https://www.researchgate.net/publication/246050153_Email_and_PIM_Problems_and_Possibilities, accessed 14 April 2021; Steve Whittaker, Tara Matthews, Julian Cerruti, Hernan Badenes and John Tang, 'Am I Wasting My Time Organizing Email?: A Study of Email Refinding', CHI '11:Proceedings of the SIGHI Conference on Human Factors in Computing Systems, May 2011, pp. 3449–3458, doi: 10.1145/1978942.1979457.

69. Matt Balogh, Taming the Paper Pile at Home: Adopting Personal Electronic Records, ArXiv, 2022, doi: 10.48550/arXiv.2204.13282.

70. Elisabeth Crawford, Judy Kay and Eric McCreath, 'Automatic Induction of Rules for E-mail Classification', Paper Presented at the Sixth Australasian Document Computing Symposium, December 2001, available at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.127.6781, accessed 10 February 2022.

71. Ibid., p. 7.

72. Grbovic, Halawi, Karnin and Maarek, p. 2.

73. Grbovic, Halawi, Karnin and Maarek, p. 10.

74. Henry Lieberman, Alexander Faaborg, Jose Espinosa and Chris Tsai, 'A Calendar and To-Do List with Common Sense', 2005, pp. 2–3, available at https://agents.media.mit.edu/projects/tasks/calendar_draft.pdf, accessed 12 March 2022.

75. Ibid., p. 8.

76. Brad Eden, 'Metadata and Its Application', American Library Association, vol. 38, no. 5, September 2002, p. 2, doi: 10.5860/ltr.38n5; Jeffrey Pomerantz, Metadata, MIT Press, Cambridge, 2015, p. 32.

77. Philip Hider, Information Resource Description: Creating and Managing Metadata, Facet Publishing, London, 2012, p. 4.

78. Ibid., p. 5.

79. National Science Foundation, Infrastructure Council, 'Cyber Infrastructure Vision for 21st Century Discovery', 2007, p. 22, available at https://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf, accessed 17 April 2021.

80. Gillian Oliver and Ross Harvey, Digital Curation, Facet Publishing, London, 2016, p. 68.

81. Hider, p. 190.

82. Oliver and Harvey, p. 107.

83. Ibid., p. 109.

84. Dublin Core. (2021, 7 July 2021). About DCMI, available at https://www.dublincore.org/about/, accessed 12 April 2021.

85. Julie Allinson, 'Describing Scholarly Works with Dublin Core: A Functional Approach', Library Trends, vol. 57, no. 2, 2012, pp. 221–243; Dublin Core, Dublin Core Metadata Initiative, 2012, available at https://www.dublincore.org/specifications/dublin-core/dces/, accessed 12 April 2021.

86. Eden.

87. Paul Dourish, W. Keith Edwards, Anthony Lamarca, John Lamping, Karin Petersen, Michael Salisbury, Douglas Terry and James Thornton, 'Extending Document Management Systems with User-Specific Active Properties', ACM Transactions on Information Systems (TOIS), vol. 18, no. 2, 2000, p. 10, doi: 10.1145/348751.348758.

88. Ibid., p. 12.

89. Paul Dourish, 'The Appropriation of Interactive Technologies: Some Lessons from Placeless Documents', Computer Supported Cooperative Work (CSCW), vol. 12, no. 4, December 2001, p. 12, doi: 10.1023/A:1026149119426.

90. Ibid.

91. David R. Karger, 'Unify Everything: It's All the Same to Me', in William P. Jones and Jaime Teevan (eds.), Personal Information Management, University of Washington Press, Washington, DC, 2007, pp. 138–139.

92. Karger, p. 139.

93. Ibid. p. 139.

94. Edgar Frank Codd, 'A Relational Model of Data for Large Shared Data Banks', Communications of the ACM, vol. 36, no. 6, 1970, pp. 377–387, doi: 10.1145/362384.362685, accessed 18 April 2021.

95. Mike Kelly, 'What Next for Tool Development', in William Jones (ed.), Keeping Found Things Found: The Study and Practice of Personal Information Management, Morgan Kaufmann, Burlington, 2008, p. 137.

96. Jones, 2008, p. 149.

97. Catherine Marshall, 'Challenges and Opportunities for Personal Digital Archiving', in C. Lee (ed.), I, Digital: Personal Collections in the Digital Era, Society of Archivists Chicago, Chicago, IL, 2011, pp. 90–114.

98. Maja Krtalić, Hana Marčetić and Milijana Mičunović, 'Personal Digital Information Archiving Among Students of Social Sciences and Humanities', Information Research, vol. 21, no. 2, June 2016, p. 7, available at https://files.eric.ed.gov/fulltext/EJ1104374.pdf, accessed 2 April 2022.

99. Smith, 2011.

100. Richard Boardman and M. Angela Sasse, '"Stuff Goes into the Computer and Doesn't Come Out" A Cross-Tool Study of Personal Information Management', Paper Presented at the Proceedings of the SIGCHI (Special Interest Group in Computer Huan Interaction) Conference on Human Factors in Computing Systems, Vienna, April 2004, doi: 10.1145/985692.985766; Michelle C. Everett and Margaret S. Barrett, '"Guided Tour": A Method for Deepening the Relational Quality in Narrative Research', Qualitative Research Journal, vol. 12, no. 1, 2012, pp. 32–46, doi: 10.1108/14439881211222714; Thomas W. Malone, 'How Do People Organize Their Desks? Implications for the Design of Office Information System', ACM (Association for Computing Machinery) Transactions on Information Systems, vol. 1, no. 1, 1983, pp. 99–112, doi: 10.1145/357423.357430; Leslie Thomson, 'The Guided Tour Technique in Information Science: Explained and Illustrated', Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, 2015, pp. 1–5, doi: 10.1002/pra2.2015.1450520100135.

101. Ofer Bergman, Steven Whittaker and Gideon Tish, 'Collecting Music in the Streaming Age', Personal and Ubiquitous Computing, vol. 26, pp. 121–129, doi: 10.1007/s00779-021-01593-6.

102. Joanne Neale, 'Iterative Categorisation (IC) (Part 2): Interpreting Qualitative Data', Addiction, vol. 116, no. 3, 2021, pp. 668–676, doi: 10.1111/add.15259.

103. Barreau; Henderson, 2003; Charlotte Massey, Sean Tenbrook, Chaconne Tatum and Steve Whittaker, 'PIM and Personality: What Do Our Personal File Systems Say About Us?', Paper Presented at the CHI (Computer-Human Interaction) 2014, Toronto, 2014, doi: 10.1145/2556288.2557023; Steve Whittaker and Candace Sidner, 'Email Overload: Exploring Personal Information Management of Email', Paper Presented at the CHI 96, Vancouver, April 1996, doi: 10.1145/238386.238530.

104. Hider.

105. Smith, 2011.

106. Hider.

107. Smith, 2005

108. Hider, p. 5.

109. Karger.

110. Hider.

111. Smith, 2011, p. 3.