Identifying digital records in business systems: the definition of a problem

Julia Kastenhofer
Information Manager
Inter-American Development Bank
Email: Kastenhoferju@gmail.com

Abstract
Before digital records can be preserved or managed, they need to be identified first. However, records identification is not a clearly defined process. Given the multi-faceted information system environment in organisations, large quantities of potential records are created and stored in systems not designed for records management. This leads to liabilities and risks on par with the non-preservation of digital records because those potential records are not identified and managed properly to begin with. A possible solution is the identification of records via business process analysis coupled with a thorough understanding of the information system environment.

Key words: Business process, digital records, information systems, record identification, recordkeeping systems

1. Identifying the problem
Preserving digital records is one of the biggest challenges the profession faces today. Decades’ worth of research went and still goes into the question of how to make digital records future-proof (Duranti and Rogers, 2012, Duranti and Thibodeau, 2006, InterPARES Trust, 2013, Xie, 2012, p. 299). This issue certainly needs to be addressed. But the problem of how to preserve digital records, as it is pursued now, pre-supposes that in the mass of diverse information created daily, records have already been identified. But this is the ideal scenario. And we are far from the ideal.

The premise of this paper is that records identification should not be equated with records creation or records capture, though these three activities might be conflated into the same work step. The identification activity entails the intellectual output necessary to determine the records value of an information aggregation. Records creation or declaration entails the work needed to make the information aggregation a record. Records capture is the work needed to manage the information as a record and might entail the transfer of the record to an appropriate environment. Therefore, creating a record and then capturing it are actions that take place after the record has been identified already. In this respect records identification can occur without records creation and capture, but records creation and capture cannot happen without identification.

In this paper I will discuss the challenge of record identification in a multi-faceted IT environment in an organisational context, and argue that records identification hinges on three parameters: first, an understanding of systems that potentially create records; second, an analysis of the business processes in the course of which records are created; and third, the grasp of what makes information a record. The three parameters are reflected in Figure 1, which depicts the four-part records identification process which is proposed in this article. To close the article I will propose a way to address the problem of practical records identification from the perspective of a records manager.
The four steps are reflected in the structure of the second segment of the article. The first segment is an attempt to define the problem related to records identification in business systems. First, the disengagement of the records identification activity in the execution of business processes. And second, the challenge of concisely explaining to complexities related to records creation to users. The second segment attempts to delineate a means of addressing the difficulties related to records identification in business systems via the four-part process in Figure 1.

Even though this article references findings resulting from Anglophone academic debates on that nature of a record and IT infrastructure and makes use of these sources, it is written from the perspective of records professionals working outside of academia. In terms of the literature available, interestingly, there seems to be a gap of academic research done on records identification and capture activities themselves. However, practical advice for information professionals gets published on a number of archives websites (Charles Sturt University, 2015, Lemieux, 1999, New South Wales State Records [Australia], 2015b, Tasmanian Archive and Heritage Office, 2014), most notably the National Archives of Australia (National Archives of Australia, 2015), which were used to fill gaps in the literature.

The scope of the article does not allow me to cover the divergent opinions related to the concept and definition of a record in the breadth they deserve. I will skim over the issues most relevant to records identification and reference articles which cover records definitions in full. Out of the wealth of research I picked out some of the definitions which I think are the most used.

1.1 **The technological environment and business processes**

Instead of one overall information system, organisations use a variety of systems and sub-systems from a variety of vendors to create and manage their business information (Alter, 2008, p. 448-449). As Xie (2012), who researched the foundations...
for digital records management in fourteen testbed organisations, acknowledges, the records of seven of those case studies “were typically in massive quantities and in dispersed repositories without systematic control over their creation, usage or maintenance” (Xie, 2012, p. 229).

For recordkeeping purposes those systems can be sub-divided into two categories: recordkeeping systems - systems that were specifically designed for the maintenance and storage of records and their related metadata. This includes all Electronic Document and Records Management (EDRMS) software and Enterprise Content Management (ECM) systems with records management functionality (New South Wales State Records [Australia], 2015a) and the more common transactional or business systems - systems that help staff do their work by supporting “single or multiple business transactions” (Cumming and Findlay, 2010, p. 269), but do not necessarily have out-of-the-box records management capabilities built-in (International Council on Archives, 2008, p. 4).

Ideally at the point of record identification and declaration, records should therefore be captured from the business system they were created in into a secure records management system:

To ensure the document is managed as a record, you have to “declare” it as a record, which means storing the record appropriately in an official recordkeeping system (United Nations Archives and Records Management Section, 2015, p. 2).

By capturing the document in a designated records management system and transferring it to its designated system, the document becomes a record (National Archives of Australia, 2015). The assumption is that the document will only be managed as an authentic, reliable and usable record if it and its related metadata are stored in an appropriate recordkeeping system (Kastenhofer, 2015).

But records and potential records are not only created, kept and managed in record keeping systems, but are also created and kept in great quantities in a wide range of transactional systems. These systems are not equipped for their long-term preservation, or even medium-term management (McLeod, 2012, p. 191). They are the systems that support the business process, and unless the business process requires the capture of those records in designated record keeping systems, those records will most likely stay undeclared in their home environment (Charles Darwin University, 2014, p. 2, Johnston and Bowen, 2005, p. 132).

In Lemieux’s(2001) analysis of the concept of a record in the context of the Jamaican financial crisis, the interviewees attribute weaknesses in the bank accounts to the mismanagement of ‘information systems’:

In the words of one interview subject:

“I would have to say that one of the big problems that faced Jamaica is [that] the management information systems ... were absolutely atrocious; nobody could get anything out of them and if I was a manager I would not have a clue as to what was going on in my bank” (Lemieux, 2001, p. 89).

Related to difficulties of managing information flows in systems is the problematic adoption of recordkeeping systems themselves. Cunningham attributes this challenge to the fact that recordkeeping systems are mostly “disconnected from core business processes” (Cunningham, 2011, p. 27) external to daily workflows and are therefore impractical and unwieldy. On top of that, recordkeeping systems and practices are perceived as user unfriendly. As Reed (2008, p. 124) notes, “there is a perception that we are narrowly focused and somewhat
obsessed with implementing recordkeeping according to a prescriptive framework, which users do not like”. Given the mass of potential records stored in transactional systems it remains questionable how practical an information transfer to a records management system really is.

Thus the challenge of managing digital records does not start with the records maintenance in systems not designed for their preservation, but that records identification and capture are at best an afterthought to staff members with other duties. The ideal would be the seamless integration of not just business and records systems but also of records capture into the business process (Cunningham, 2011, p. 27, Duranti, 1995, p. 9). As reasoned by the University of British Columbia Project “the procedures of making records should be integrated with those of the related business activities […] however, the procedures of making records and even those of conducting business activities were not documented, and the documented ones were incomplete and outdated” (Duranti et al., 2002, Xie, 2012, p. 303).

The activity of capturing a record also presupposes that the user is aware of what it is he or she is supposed to be capturing and knows how to practically go about it. One way to circumvent the necessity of having users identify and create records is to automate the process of records creation by having metadata-based retention rules running in the back of the business system. This would relieve users of having to decide which documents are records. The work the users would need to perform would be to capture metadata at the point of document creation or upload on which the rules are based. But that is the optimum from the user perspective. The task of identifying records would be deferred to document and records managers and again does not solve the problem of how to identify records in the first place.

The focus of business users is not on the creation of records as an information asset, but on the creation of information, documents and data. Records are an afterthought. Lemieux (2001, p. 91) summarizes the ambiguity of records identification:

Likewise, archivists and accountants have different labels and ideas about this recorded information. Even within the archival profession, there are many differences among archivists’ conceptualizations of records. Perhaps the argument is best summed up with the observation that one person’s ‘management information’ is another person’s ‘record’.

There is little awareness within the business community of what a record is, what it could be and how to identify potential records. Thibodeau (2009, p. 27-28) for example, notes that due to its nature, digital data cannot be easily divided into discrete units which makes it difficult to distinguish where one potential record ends and the next begins. On top of that Cunningham (2011, p. 23) acknowledges that “the problem is that agencies are incapable of identifying the important and valuable records within these mostly low-value digital holdings. Picking through the digital slag heaps to find the occasional gem takes too long and costs too much”.

1.2 Records management guidance

One way to address user’s unawareness of what a record is to educate them. McLeod (2012, p. 168) argues that “everyone is a records manager (with a lower case “r” and “m”), not just in principle but in practice, and that the role of the records manager (professional) is strategic and enabling, horizon scanning and focused on our role in solving big challenges“. If records managers want to enable users to manage records, as McLeod writes, a more thorough understanding of the technological and
procedural context records are created is supremely beneficial. This means understanding the systems that create information aggregations and the business processes they support. But that alone does not suffice.

If the records professional only understands the context without being able to effectively convey what a record is, he or she cannot provide users with the necessary expertise in records management. In order to manage the information quantity, users need to identify their records themselves. But the record creators cannot do that if they do not know what they need to identify, or how to practically go about it.

So apart from understanding the technological context and business process, another challenge related to records identification records managers face is to explain what a record is regardless of its format or storage environment and to do so in a manner that is practical and easily applicable. At first glance this task might seem straightforward, but it is deceptively complicated to strive for something approaching accuracy, simplicity and practicality all at once.

The classic solution at this point is to turn to standards. ISO 15489, the international standard for records management, defines a record as “information created, received, and maintained as evidence and information by an organization or person, in pursuance of legal obligations or in the transaction of business” (International Standards Organization, 2001, p. 3). ISO 30300 extends the original version by adding three explanatory notes:

“NOTE 1 Adapted from ISO 15489-1:2001, definition 3.15.
NOTE 2 The term “evidence” is not limited to the legal sense (see 3.1.5).
NOTE 3 ‘This applies to information in any medium, form or format.” (International Standards Organization, 2011, p. 9)

Even though ISO 30300 extends the original definition by clarifying that records can be in any format or medium, it reveals nothing about the systems records are stored in. Basically, what the ISO definitions tell their readers is that records show that something happened.

Apart from the standards, records professionals also extensively deliberated the different concepts of a record and with more finesse than I can do in this space (Brothman, 2002, Duranti and Thibodeau, 2006, Finnell, 2011, Yeo, 2007). Duranti and Rogers (2012) for example argue that records are evidence of actions and transactions. Yeo (2007, p. 337) defines a record as “persistent representations of activities, created by participants or observers of those activities or by their authorized proxies”. With this definition he evades pitfalls related to the concepts of information, transactions and evidence, but sacrifices usability for the sake of an approximation to accuracy. Then again, how persistent a digital record can really be is debatable.

The crux of the records definition problem is not so much the definition of an indescribable essence, but the purpose of record keeping, which is future use (Van Garderen, 2007). Records are not records because of their own essential ‘recordness’ (Brothman, 2002, p. 315). It would make things easier if it were so. Records are records because someone made the purposeful decision to keep them. Because they show that something significant, whatever that means, to the person or the organisation happened, which will be revisited in the future. The content of the information in a specific context is the trigger for records identification, but it is that intention of future use for whichever reason that makes something a record.

Definitions by themselves do not really help. They are not practical for non-professionals. They are not clear-cut decision-making aids. Not because the definitions as such are bad, but because the act of record identification is
deeply ambiguous. No information aggregation is intrinsically a record. It is made a record by a thinking, feeling person. So even though definitions by themselves are not the silver bullet, they can be the starting point of a guidance strategy.

As a result records managers are hard-pressed to explain the records identification process for information with not necessarily stable content for two reasons: First, because of a limited understanding of the business process that created the information or potential record in the first place. And secondly, because of an inadequate knowledge-base of the transactional system. So how can records identification be made clearer?

2. Identifying records

One of the big challenges records managers in organisations face is how to guide and manage the records identification process with a multitude of different information systems in use. It is the lack of strategic vision for practical digital record keeping that results in grave risks to the information. Leaving these potential records in the host environment with no clear strategy in place is ill-advised.

The United Nations Archives and Records Management Section (2015, p. 3) provides practical guidance on records identification online and suggests that “a document becomes a record when you decide you need it as evidence of a decision or action”.

Thibodeau (2009, p. 28) expands on this explanation by defining two more parameters with which digital records can be identified – persistency and reproducibility:

We can, then, formulate a simple set of three criteria for identifying a stored electronic record: it must be (1) a persistent digital object that (2) contains fixed information about an activity or the state of affairs at the time when the action was done, and that (3) can be used to produce one or more manifested records. The stored record may be a single information chunk or digital component, but it might just as well contain many thousands of such elements.

While Thibodeau and the UN Archives and Records Management section focus on the would-be records themselves, Rogers (2015, p. 18) advocates the use of metadata for records identification. She argues that “digital diplomatics, based on a foundation of traditional diplomatic principles, can help identify digital records through their metadata”. She writes further:

To identify records in digital systems and assess their authenticity, we must return to the principles of archival diplomacy. We must locate the extrinsic and intrinsic elements of form in the identity and integrity metadata that may be found at various abstraction layers of the conceptual and logical record (Rogers, 2015, p. 14).

What was most likely meant with those statements is that a digital record may be identified given a set of metadata known about the record that determines whether the potential record is a record. I will take the message out of its intended interpretation and argue that metadata collection of the systems and business processes that produce records, instead of the records themselves, is a decisive step towards manageable record identification.

2.1 Analyse the system

Therefore, an alternate method is a four-step records identification process via business process analysis and system analysis. The first step to identifying and capturing records in business systems is to collect metadata about the systems. Metadata can after all be collected about systems, just as it can be collected about records (ISAD (G)) and functions (ISDF) and record creators (ISAAR (CPF)). Figure 2 depicts the possible components of a systematic description of information systems, which
should be completed to get the thorough understanding of the information environment necessary to identify records in the business system.

Figure 2: System analysis components

The focus is not on determining the recordkeeping capabilities of the systems in question, though that is a side product, but on collecting data on systems in order to assess whether they produce records at all. That includes the collection of basic information which facilitates system identification – type, name, vendor, version, business owner, the hardware they depend on, the software used, its information architecture, which units and how many users use it, which business functions and processes it supports, contingency plans and so forth etc.

Furthermore, the systems can be categorized according to their longevity, which can be defined as the estimated future use of the system. Some systems might be in use for the next five years, while some will be phased out and subsumed by another system. Others are already phased out and are in read-only mode or lost altogether.

While metadata of all systems has to be collected, prioritization and pragmatism dictates that it is crucial to address the systems that will be or already have been phased out first, as here the risk of information loss is the gravest. If emergency records capture at the point of system decommission is done, valuable records might be saved. To address this problem, records capture activities should also be integrated into the system decommissioning process. A key factor in this regard is the effective collaboration with IT professionals (Rogers, 2015, p. 9).

Parts of the InterPARES 1 Template for Analysis can aid the system analysis process because the Template’s intention for use is to analyse digital records and collect metadata on their components, the technological infrastructure is of secondary importance (InterPARES 1 Project: Authenticity Task Force, 2002, p. 1). As the Template covers the technological context of the system the record is stored in, it can however be re-purposed, cut in some sections, extended in others and re-used to
analyse information systems, especially business systems.

The pertinent part of the Template for Analysis is the ‘Technological Context’ section, because “the characteristics of the technical components of the electronic system in which the record is created” (InterPARES 1 Project: Authenticity Task Force, 2002, p. 6) are under scrutiny here. As it is now, the Template collects information on six aspects of the technological environment – hardware, software, data, information system, system models and system administration.

This part could be extended to include information on system ownership, licensing and the system lifecycle. Metadata collection on the system lifecycle as a whole could include information on the longevity of the system, vendor support, development and implementation year, predecessor system, updates and the backup strategy.

2.2 Analyse the process

The second step is to define and outline the business processes done by the organisation via functional and business process analysis in order to determine how the organisation works and for which ends information is produced. Records professionals already have ample experience applying these methods, as business process and functional analysis is already widely used to determine file plans and retention schedules. For records identification purposes business process analysis is vital, because knowing how staff members execute their functions ensures that the records manager also knows what and where information is produced and stored.

An example of a business process is the Manage Payables process from the Guideline on Common Financial Management, published by the Treasury Board of Canada Secretariat. Within the financial management function in most institutions, there are various business processes and one would be managing payables. The Treasury Board of Canada Secretariat issued guidelines describing a business process entitled Manage Payables. The business process, which was drawn by the Treasury Board as seen in Figure 3, describes “handling invoices, completing the account verification, and providing certification authority” (Treasury Board of Canada Secretariat, 2012).
The process starts at the point the invoice is handled up to the point where payment is performed. Figure 3 shows that there are ten activities undertaken in this business process with four decision points and three outcomes. As can be seen in part 2.3, the number of outcomes of the business process does not necessarily correspond with the records that can be identified.

2.3 Identify where the information is produced

The third step is to locate steps in the process where information is produced in order to determine which records are produced and re-used in the business process. In the case of Treasury Board of Canada’s Manage Payables process, the information which is created in the course of the process includes the invoice or credit memo, its supporting documentation, correspondence on matters like discrepancy resolutions, the confirmation that contract terms have been met, the confirmation that the payee is entitled to payment, the master data file, relevant regulations, policies, directives, the certification that all the information is complete and the request for payment.

The information that reflects the decisions taken and products of the process, as well as any information that needs to be kept because of its operational, legal or historical value, is identified as a record. For the Manage Payables process this is the invoice or credit memo, which triggers the process, the master data file, the certification that all the information is complete and the request for payment.

The supporting information to the invoice would need to be appraised so as to determine if the information should be identified as a record, because it contributes to the understanding of the invoice. Routine correspondence would not need to be identified as a record unless irregular errors in the process execution occurred at which point the correspondence would need to be kept to maintain the accountability of the information. The various confirmations would not need to be kept separately, because this would be included in the certification that all information is complete. Regulations, policies and directives would not need to be kept, because they did not originate in this particular business process. Unless the policies have been lost and cannot be recovered, they would not need to be identified as records of this business process.

2.4 Identify where the records are

In the fourth and last step, step 1 (system analysis) and step 3 (records identification) are combined in order to determine in which systems the records are. Now that the records have been identified, after a strategy has been worked out about what should now be done with them, steps can be taken to create and capture them.

Once those steps have been completed and the records are identified, measures can be taken to preserve them. This is a tentative hypothesis for records identification itself, a topic which has not yet been exhaustively studied. The similarities between records identification, records creation and records capture should be analysed in more detail. The four step process could be tested via the comparative analysis of case studies in order, which would allow the identification of gaps in the theory and propose ways to address them. A simplified illustration of steps 2 to 4 of the records identification process, business process analysis, information identification and records identification, can be found in Figure 4.
Figure 4: From business process analysis to records identification

**Records Identification Step 2-4**

**Business Process**

- **Start of Process**
  - Step 1
  - Step 2a
  - Step 2b

- **Potential record:** Step 2a Outcome

- **Potential record:** Step 5a Outcome

- **Step 3**
- **Step 4**

- **Where are potential records?**

- **End of Process**
  - Potential record: Step 5d Outcome

**Step 2: Analyse the Business Process**

- Define the functions the organisational unit does
- Define the activities for the function (consult for example the file plan for a breakdown of functions and activities)
- Analyse the business process of each activity: for example the Treasury Board of Canada Secretariat "Manage Payables" Workflow - Subprocess 3.1.7 (Treasury Board of Canada Secretariat 2012)

**Step 3: Identify steps in the process where potential records are created**

- Identify steps of the business process where potential records are generated by
  - a) Analysing what documentation is produced in the business process.
  - For example: supplier lists, invoices, receipts, purchase orders, statement of funds, accounting entries, correspondence
  - b) Analysing what - of total number of documentation produced in the business process - has to be kept to prove that the work was done and decisions were made.
    - For example: The building is on fire and you can only take two documents from this business process. Which do you take?
  - c) determining what needs to be kept for legal reasons
  - d) determining what needs to be kept for ongoing business

**Step 4: Test whether the potential record is a record and identify where the records are**
3. Conclusion

This paper argues that before digital records can be preserved, they need to be identified first. Records identification is an activity that is related to, but ultimately a separate activity from records creation and records capture. In this respect, records identification is defined as the intellectual output necessary to determine the records value of an information aggregation. While records identification was already a deeply ambiguous process in a paper-based environment, it is further complicated given complex and evolving information environments in transactional or business systems.

Records creators need to know what they are supposed to identify before records can be created. Records managers cannot and should not achieve this task by themselves, which in turn creates the necessity on training record creators on what they should be creating. This task is further complicated through the ambiguity of the concept of the record in its different host environments. One way to address the need to clarify the records identification activities could be through a four step process which entails an analysis of the transactional systems, an analysis of the business processes that are supported by the various systems in use, a breakdown of the business process into actions where records are potentially created and lastly a combination of step 1 and 3 in order to determine where which potential records are stored.

References

xt=libphilprac (accessed 30th July 2015).


**Biography**

Julia Kastenhofer completed her undergraduate studies in History at the University of Vienna (Austria). She holds master's degrees in Archives and Records Management from University College London (UK) and in Global History from the University of Vienna. Her work experience in Archives and Records Management includes positions at the International Atomic Energy Agency in Vienna (Austria) and the Historical Archives of the European Union in Florence (Italy). She currently works as an Information Manager at the Inter-American Development Bank in Washington, DC (USA).