Digital Archiving and eDiscovery.
Delivering Evidence in an age of Overload

Dr Geert-Jan van Bussel, Dr Hans Henseler
HvA Amsterdam University of Applied Sciences, Amsterdam, The Netherlands
g.j.van.bussel@hva.nl
j.henseler@hva.nl

Abstract: Within eGovernment, trust in electronic stored information (ESI) is a necessity, not only when communicating with citizens, but also for organizational transparency and accountability. In the last decades, most organizations underwent substantial reorganization. The integration of structured data in relational databases has improved documentation of business transactions and increased data quality. That integration has improved accountability as well. Almost 90% of the information that organizations manage is unstructured (e.g., e-mail, documents, multimedia files, etc.). Those files cannot be integrated into a traditional database in an easy way. Like structured data, unstructured ESI in organizations can be denoted as records, when it is meant to be (and used as) evidence for organizational policies, decisions, products, actions and transactions. Stakeholders in eGovernment, like citizens, governments and courts, are making increasing demands for the trustworthiness of this ESI for privacy, evidential and transparency reasons. A theoretical analysis of literature of information, organization and archival science illustrates that for delivering evidence, reconstruction of the past is essential, even in this age of information overload. We want to analyse how Digital Archiving and eDiscovery contribute to the realization of trusted ESI, to the reconstruction of the past and to delivering evidence. *Digital Archiving* ensures (by implementing and managing the 'information value chain') that: [1] ESI can be trusted, that it meets the necessary three dimensions of information: quality, context and relevance, and that [2] trusted ESI meets the remaining fourth dimension of information: survival, so that it is preserved for as long as is necessary (even indefinitely) to comply to privacy, accountability and transparency regulations. *EDiscovery* is any process (or series of processes) in which (trusted) ESI is sought, located, secured and searched with the intent of using it as evidence in a civil or criminal legal case. A difference between the two mechanisms is that Digital Archiving is implemented *ex ante* and eDiscovery *ex post* legal proceedings. The combination of both mechanisms ensures that organizations have a documented understanding of [1] the processing of policies, decisions, products, actions and transactions within (inter-) organizational processes; [2] the way organizations account for those policies, decisions, products, actions and transactions within their business processes; and [3] the reconstruction of policies, decisions, products, actions and transactions from business processes over time. This understanding is extremely important for the realization of eGovernment, for which reconstruction of the past is an essential functionality. Both mechanisms are illustrated with references to practical examples.

Keywords: E-government, Accountability, Evidence, Digital Archiving, eDiscovery

1. Neglected challenges of eGovernment

In the 1980s and 1990s, in view of a transformation into an information society, in many organizations the increase of business process efficiency was a predominant item. In the process of environmental adaptation, organizations re-engineered their business processes and exchanged their standalone applications for more standard, integrated solutions. The subsequent integration of structured data in relational database management systems (RDBMSs) improved the documentation of business processes, increased the quality of structured data and improved accountability. But almost 90% of the electronic stored information (ESI) that organizations manage is unstructured (Scholtes 2009), and cannot easily be integrated into traditional RDBMSs (Van Bussel 2011). Unstructured ESI is stored in text, spreadsheets, emails and/or multimedia files.

IDC estimates a 40 % annual growth rate of ESI until 2020 (IDC 2011). From 1986 until 2007, Hilbert and López (2011) concluded, ESI had an average annual growth rate of 23 %. From 2000 until 2007, this growth rate was especially high: in 2000 25 % of all data were stored as ESI; by 2007, that percentage had risen to 94 %. The amount of unstructured ESI is not likely to diminish: within public and private organizations the use of collaborative Web 2.0. technologies to streamline business processes creates 'big data' (Jacobs 2009). Social media tools generate large amounts of unstructured ESI, created by knowledge workers who engage in peer-to-peer knowledge sharing across organizational boundaries. Using all different sorts of (social) media, they participate in organization-wide and inter-organizational collaboration. The storage, dissemination and processing of unstructured ESI require

complex ICT systems.

In this dynamic environment, ICT systems present security and durability challenges that pose a major threat for information quality (Bearman 2006). These threats are, together with the exponentially rising amount of ESI, the most neglected challenges of eGovernment. Those challenges threaten the trustworthiness of organizational records, that ESI that is meant to be (and used as) evidence for policies, decisions, products, actions and transactions. In this paper, we will denote those organizational records as ESI. Stakeholders in eGovernment, like citizens, governments and courts, are making increasing demands for the trustworthiness of this ESI, mostly for privacy, accountability and transparency reasons. In this context, the audit reports of the Dutch Algemene Rekenkamer (Court of Auditors) present an interesting case. Those reports are very critical of the design and deployment of ICT systems within the Dutch government. The Dutch Court of Auditors identifies several very serious consequences for information security, data quality, transparancy and privacy. According to these audit reports, those consequences should be tackled forthwith (AR 2007-2008, 2011).

2. Research Question

A theoretical analysis of literature of information, organization and archival science illustrates that for delivering evidence, reconstruction of the past is essential (as cases presented in, for instance, Barata and Cain (2001), the Sedona Conference (2004), Allman (2007), Paul and Baron (2007), and Toebak (2010) illustrate). This literature suggests two mechanisms to reconstruct the past and to realize trusted ESI: Digital Archiving (DA) (or, as it is sometimes called, enterprise records management), and eDiscovery. In this paper, we will analyse how those two mechanisms are contributing to the realization of trusted ESI, to the reconstruction of the past and to delivering evidence.

3. Accountability, evidence and the reconstruction of the past

Accountability is the acknowledgement of responsibility for policies, decisions, products, actions and transactions, and the obligation to report and be answerable for consequences. It is a social relation between an actor and a forum. When the *actor* is an organization (as it is here), we talk about 'organizational accountability'. The *forum* is a designated forum (shareholders, citizens, courts, etc.) or a virtual entity ('society', 'the people'). A forum will ask an actor to provide insight in its process effectiveness and the lawfulness or unlawfulness of its actions and transactions. It will ask an actor to provide evidence for all assertions made in this audit. The forum passes judgement on the actor's conduct and (if necessary) imposes sanctions (Bovens 2006). Evidence includes everything that is used to determine or demonstrate the truth of an assertion. Evidence must be: authentic, accurate, complete, reliable, and in conformity with legislative rules. It must be tied to a fact to prove something. The evidence collection and analysis procedures must not cast doubt on the evidence's authenticity and veracity (Kozushko 2003).

Barata and Cain (2001) prove conclusively (based on several practical cases) that accountability without trusted ESI as evidence of the organizational past is not possible. This means that organizations need an accountability function to safeguard that evidence. To improve accountability, new ICT systems, concepts and methods have to be implemented to identify, structure, organize, process and retain that ESI that is used within organizational processes, as well as all the ESI that is used to document how actions and transactions have been performed. Ensuring the quality of this ESI is difficult. Redman (2004) states that information quality is a 'disaster'. ESI is inaccessible, incomplete, inconsistent, irrelevant, untimely, inaccurate, and/or not understandable. Its provenance and contextual environment are unknown (Epler 2006). ESI is generated in a large variety of formats without adding something new (Paul and Baron 2007). It is stored in different forms and places and in various business processes. Organizations tend to facilitate this by allowing users to maintain this business related ESI within personal archives. In addition, ICT creates technological obsolescence. ESI has a longer lifespan than the configurations in which it is created or managed (Boudrez, Dekeyser and Dumortier 2005). Trusted ESI is indispensable as evidence and necessary for making reconstructions of the organizational past. For eGovernment to succeed, ensuring trusted ESI is a necessity, as an Singaporean empirical study suggests (Thompson, Srivastava and Jiang, 2008-2009).

4. Digital Archiving

4.1. 'Information value chain'

ESI used for informational and productive purposes in business processes, and meant to be or to be used as evidence, can be text, multimedia files, database records, technical drawings, or combinati-

ons thereof. This ESI is critical for business process performance (Marchand, Kettinger and Rollins 2001, ch. 6). Besides that, ESI documents how policies, decisions, products, actions and transactions have been performed. ESI's business importance means that it needs to be identified and controlled. According to Van Bussel (2012), DA organizes the *'information value chain'*, the chain that ensures that the informational and evidential 'value' of ESI is utilized in business processes to improve performance. This chain includes all information processes: generation or receipt, identify, capture, storage, processing, distribution, structuring, publication, (re-)use, appraisal, selection, disposal, retention, security, auditing and preservation. DA is oriented [1] on the effects of this chain on business processes, and [2] on the realization of the four dimensions of information. That way, it ensures the provision of trusted ESI for informational and evidential reasons (Van Bussel and Ector 2009).

4.2. The four dimensions of information

In this age of organizational chains, inter-organizational data warehouses, cloud computing, authentic registrations, and computer mediated exchange, the four dimensions of information allow for the reliable reconstruction of policies, decisions, products, actions and transactions. Those four dimensions are quality, context, relevance and survival (Van Bussel 2012) (Figure 1).

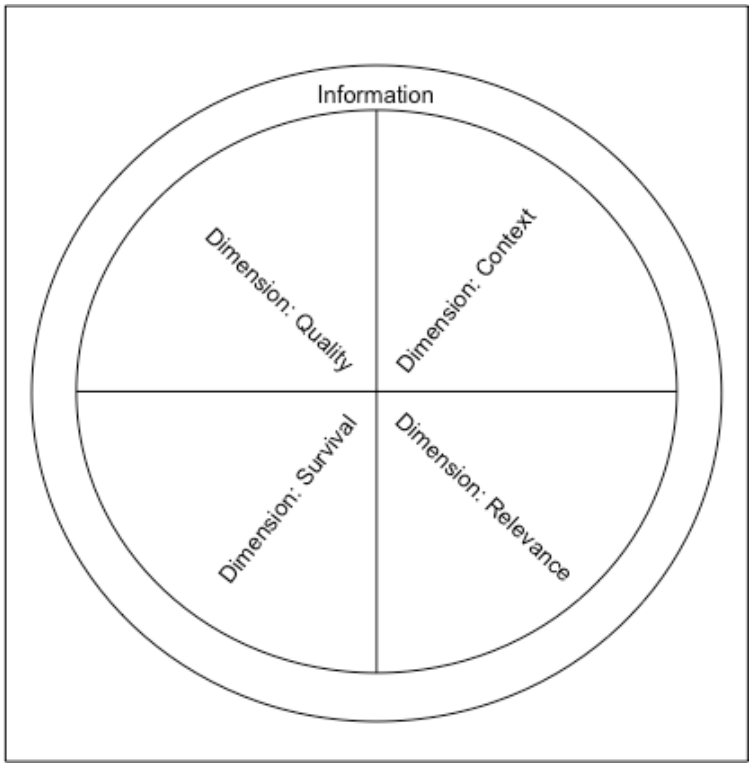

Figure 1: The four dimensions of information

For *Quality*, DA is focused on the quality requirements of [1] (structured or unstructured) ESI and [2] the 'information value chain'. For ESI, four quality requirements are recognized: integrity (ESI cannot be manipulated), authenticity (ESI presents the required (and original) content and lay-out), controllability (ESI can be tested on reliability) and historicity (ESI can be reconstructed). Those requirements realize the fixity of ESI. This means that ESI is (or can be reconstructed as) an 'immutable mobile' (Latour 1990). Fixity is necessary because ESI is [1] recorded for *later* consultation and [2] used repeatedly for the reconstruction of *past* happenings. Fixity enables users to trust ESI and to use it as evidence. Levy (2001, ch. 2), gives several practical examples. The 'information value chain' ensures that ESI is correct and complete in spite of all necessary handling. The requirements for this value chain are identical to those for organizational business processes. They are well-known, namely reliable time of delivery, effectiveness, efficiency, product quality, alignment of needs, product management, and compliance (Van Bussel and Ector 2009).

The second dimension, *Context,* provides meaning (Duranti 1997). Knowledge of the (environment of the) policies, decisions, products, actions or transactions for which ESI was generated is necessary

for extracting meaning out of ESI. This knowledge applies to the existing juridical system, the organizational structure, the procedures by which ESI is generated and the ESI collection to which it belongs. Groth (2007) suggests three characteristics of context: [1] it needs accurate documentation, [2] it is in the past, and [3] it is necessary for the tracking and the reconstruction of business processes. Pipek, Wulf and Johri (2012) describe the case of a German steel mill to elaborate on the importance of this dimension of information.

Relevance is a fundamental concept in human communication. As Saracevic (2007) explained, ESI is only relevant for users if it fits the context in which it is used, managed and retrieved. For organizations, ESI is only relevant if it fits the organizational objectives of performance and accountability (Van Bussel 2012). A special kind of relevance is *appraisal*. It is developed within archival theory to determine the 'value', relevance, of information over time. It means the complex evaluation of ESI to determine its economic, organizational, financial, fiscal, juridical, legal, societal and historical relevance and to develop a retention schedule for ESI. In such a schedule the periods ot time that ESI should be kept or 'retained' (as, for instance, stated in (privacy) law and regulations) are documented, including indefinite retention for ESI of 'enduring value' (Cox and Samuels 1988). When the retention period has expired, ESI has no organizational (or legal) relevance any more and should be irreparably deleted (Van Bussel 2012). Disposing of irrelevant ESI saves high costs for retention and accessibility. Besides that, irrelevant ESI makes organizations vulnerable to legal proceedings, for instance in the context of privacy law, fraud or corruption. The much disputed 'right to be forgotten' is an essential part of the discussion on the relevance of ESI (Blanchette and Johnson 2002; Rosen 2012). Appraisal ensures that within the 'information value chain' (privacy) laws and regulations are respected. Pederson, Routledge and Thurston (1999) have collected 24 case studies on this kind of relevance.

The fourth dimension, *Survival*, concerns the many security and durability challenges, which have to be overcome to realize access, retrieval and preservation over time (Bearman 2006). It stresses the importance of a reliable and durable ICT infrastructure (which is not self-evident) to enable the continuous and secure storage and manipulation of ESI. This infrastructure assists organizations in seeking competitive advantages, developing the organization learning concept, increasing autonomy and being accountable (Croasdell 2001). The infrastructure needs to safeguard the first three dimensions of information over time, but its features are fragile and continuously influenced by the restructuring of organizations (Boudrez, Dekeyser and Dumortier 2005). First, hard- and software configurations are needed for accessing, retrieving and viewing ESI, which means that a solution for technological obsolescence should be available. ESI has a longer lifespan than the configurations in which it is created or managed. Secondly, the large influx of information requires automated archiving and retrieval functionalities. The ICT infrastructure needs to adapt, transform, renew and grow. Thirdly, ESI is of a diverse nature. There is a diversity of object types, operating systems and applications. The handling of this diversity is not self-evident, while at the same time information can be continuously modified. This endangers the trust in reliable ESI. Fourthly, ESI can only be reliably used, when it can be interpreted by users in its original organizational context. Context and ESI need to be forever linked (Van Bussel 2011). A case-based review of this dimension has been offered, among others, by Hockx-Ju (2006).

4.3. The benefits of Digital Archiving

DA [1] safeguards the 'information value chain', [2] ensures that ESI meets the four dimensions of information, [3] realizes a reliable, effective and efficient use of trusted ESI in business, [4] provides an ICT infrastructure to (indefinitely) store (identified and trusted) ESI and keep it accessible, [5] ensures that (privacy) laws and regulations are respected within the 'information value chain', and [6] audits periodically the possibility to reliably reconstruct the past and to use ESI as evidence. In DA, it is emphasized that the failure to realize the dimensions of information is a threat to the possibilities to reliably reconstruct the past. Because of that, the organizational accountability function can not be successful. In managing the 'information value chain', DA ensures that ESI can be used as evidence. Its purpose is to reduce the costs of transactions, to enlarge the speed of access to past experiences, to help in decision-making, to share knowledge and to realize accountability. DA, therefore, is a necessity for realizing organizational accountability. DA is always implemented *ex ante* the beginning of possible legal proceedings.

5. eDiscovery

EDiscovery can be seen as a process or a series of processes in which (trusted) ESI is sought, located, secured and searched with the intent of using it as evidence in a civil or criminal legal case (Conrad 2010). It tries to reconstruct the past to learn the informational context of ESI, because only then

ESI can be used as evidence. Forensic accountants, lawyers, internal and external auditors, they all resemble archaeologists trying to reconstruct events in an almost forgotten past. Similar to archaeology, there may not be much left to investigate, making advanced techniques to reconstruct the past necessary. eDiscovery is an essential tool for investigators to reconstruct past actions and transactions and to support findings in order to prosecute organizations that are in violation with (inter-)national regulations. For instance, the European Union aims to reduce all forms of corruption at every level. It also aims to prevent fraud by setting up anti-fraud offices and actively investigates and prosecutes violations of competition regulations. EDiscovery has to be attuned to (inter-)national laws and regulations. This is a challenge, especially as far as privacy law is concerned, for there are worldwide different interpretations of privacy and divergent privacy law traditions (Berman 2010). EDiscovery itself does not contribute to the realization of trusted ESI in organizations, because it is implemented *ex post* the beginning of legal proceedings. It uses ESI as it is at the time of those proceedings.

The eDiscovery process (or series of processes) is broadly illustrated by the eDiscovery Reference Model (EDRM). This reference model was developed to address the lack of standards and guidelines in eDiscovery identified in the 2003 and 2004 Socha-Gelbmann Electronic Discovery surveys (Socha and Gelbmann 2004). The EDRM diagram presents an overview of eDiscovery and symbolizes the transformation of a large volume of *general* ESI into a small volume of *relevant* ESI that can be used as evidence in legal procedures. The automated part of this process, 'culling', is aimed at searching, selecting and filtering ESI without manual review. The model divides the process into six stages: [1] Information Management; [2] Identification; [3] Preservation and Collection; [4] Processing, Review and Analysis; [5] Production; and [6] Presentation (figure 2). Those stages are described in detail on the EDRM Website (EDRM 2006). The first stage, Information Management, is related to DA. It symbolizes the discovery of all possible sources of trusted ESI in an organization. They may be business systems, such as fileservers, email servers, ERP systems, (partly implemented) enterprise document and records management systems, (forgotten) backup tapes or personal archives. eDiscovery specialists like to collect data from personal archives and combine this with data recovered from old data backups that were maintained for disaster recovery purposes. They can disclose ESI that has been deleted on purpose (outside of a 'defensible disposal') (Henseler 2010a), as has been demonstrated in the Enron and Lehman Brothers cases (Keena 2002; Valukas 2010). This stage is difficult to implement within eDiscovery for it cannot be controlled. It requires knowledge about how and where trusted ESI has been stored. Most information management efforts within eDiscovery are crippled by insufficient collaboration among key stakeholders: [1] the business users who need ESI to operate an organization, [2] the IT departments who must implement the mechanics of information management, and [3] the legal, risk, and regulatory departments who understand the duty to preserve ESI.

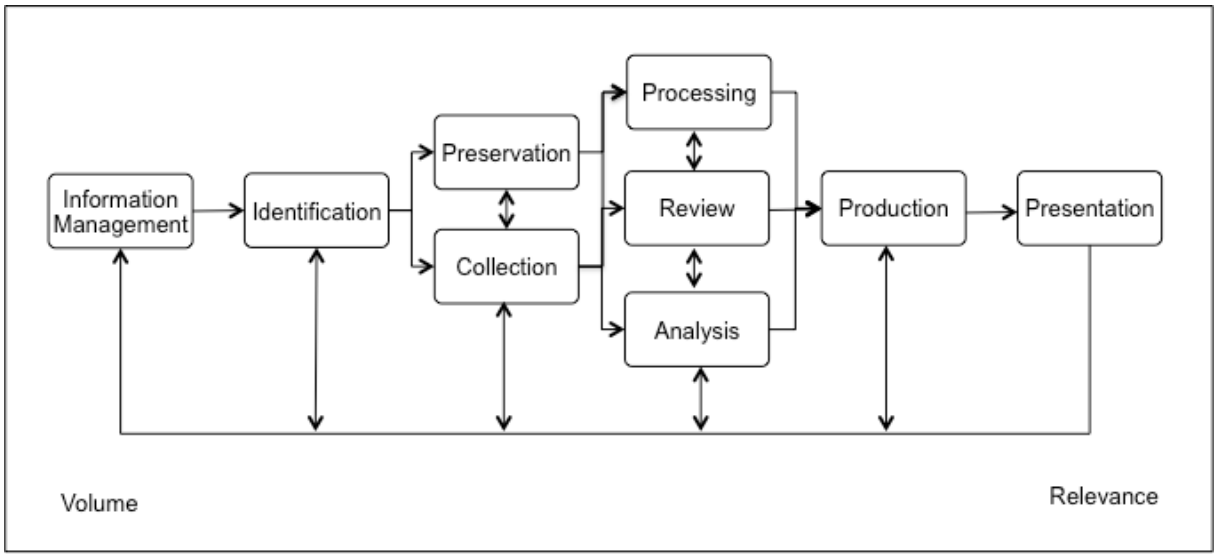

Figure 2: The EDRM Model

Information overload makes finding and reviewing ESI expensive and increases the risk of failure. The problem is amplified as ESI is more easily generated in a large variety of formats without adding new information, as Paul and Baron (2007) have illustrated. The solution is to collect as much ESI as can

be found. eDiscovery tools use a variety of culling techniques for removing duplicates, filtering based on file extension, date time and / or keywords in order to reduce the volume of ESI. Keyword filtering relies on a set of keywords that is designed based on the context of the investigation, for instance names of persons, projects, places, companies, etc. The retrieval of information from large amounts of data is an important part of this process. This step involves a lot of of manual work done by lawyers searching for evidence in e-mail which may include millions of documents (Sedona Conference 2004). This situation is rapidly changing. Experiments have shown that finding relevant ESI using keyword search is far from perfect (Krause 2009) and that it is expensive and very inefficient (Oard, et al. 2010). Experiments have shown that technology-assisted review can outperform manual review (Grossman and Cormack 2011). The need for better search tools and methods is reflected in the growth of the eDiscovery market (Gartner 2009) and in the growing research interest (Ashley and Engers 2011). New techniques are being introduced that allow a faster review of ESI by using tools such as conceptual search (Chaplin 2008), detection of near duplicates and visual analysis (Görg, Stasko and Liu 2008), predictive coding, that is used to automatically identify ESI using statistical pattern recognition (Peck 2011) and social network analysis (Henseler 2010b).

As a consequence of these developments eDiscovery experts are getting better in recovering ESI that has never been archived and that was assumed to be deleted. Organizations should realize that it is almost impossible to delete information that is stored in many different forms and places in various business processes without implementing DA. The introduction of cloud-based storage is likely to worsen this situation because ESI is stored redundantly and beyond the control of an organization, as the Los Angeles case demonstrates (Murphy 2011). This creates a liability because ESI might be discovered while it should have been deleted according to internal accountability rules. eDiscovery demands the four dimensions of information. It demands DA to retain what should be remembered and to forget what should be forgotten.

6. Conclusion

In this paper, our objective was to analyse how DA and eDiscovery contribute to the realization of trusted ESI, to the reconstruction of the past and to delivering evidence in a time of information overload. We describe the core objectives of DA and eDiscovery. DA's core objectives are improving business process performance and realizing accountability. DA assists organizations in using the 'information value chain' and the four dimensions of information to reconstruct past policies, decisions, products, actions and transactions. It enables the use of ESI as evidence and enables business processes to improve performance. DA is a necessity for realizing organizational accountability and delivering evidence. It is implemented *ex ante* possible legal proceedings. The objective of eDiscovery is delivering evidence. To find evidence, it uses complex information retrieval and culling technologies. It safeguards its evidence collection and analysis procedures, to ensure the authenticity and veracity of the evidence. It is implemented *ex post* the beginning of legal proceedings.

DA and eDiscovery are two sides of the same coin. The first is implemented to assist an organization in improving business process performance and in preparing for litigation. The latter is used when an organization is asked to be accountable and is subject of litigation. DA identifies trusted ESI and manages it with the 'information value chain'. ESI is identified, structured, stored, preserved, appraised, deduplicated and irreparably deleted (according to legal rules and regulations). DA will not be able to eliminate the explosion of ESI, but it will reduce additional overload caused by storing the same ESI in different formats. It will ensure that ESI is deleted, when necessary. eDiscovery will use its information retrieval and culling technologies to locate the needed evidence in the enormous sources of trusted ESI. By deleting ESI the cost of eDiscovery will be (much) lower, while simultaneously reducing the potential for liability. In combination both mechanisms ensure that organizations have a documented understanding of the past. This is important for eGovernment, for reconstruction of the past is an essential functionality for delivering evidence, and for improving organizational transparency and accountability.

References

Allman, T.I. (2007) "Managing preservation obligations after the 2006 federal e-discovery amendments", *Richmond Journal of Law & Technology*, Vol 13, No. 3, pp 1-35.
AR Algemene Rekenkamer (2007-2008) *Lessen uit ICT-projecten bij de Overheid, Deel A, Deel B*, Algemene Rekenkamer, 's-Gravenhage.
AR Algemene Rekenkamer (2011) *ICT bij de Politie. Rapport 2010*, Algemene Rekenkamer, 's-Gra-

venhage.
Ashley, K.D. and Engers, T.M. van (eds.) (2011) *The 13th International Conference on Artificial Intelligence and Law. Proceedings of the Conference June 6-10, 201*, ACM, Pittsburgh.
Barata, K., and Cain, P. (2001) "Information, not technology, is essential to accountability: electronic records and public-sector financial management", *The Information Society,* Vol 17, pp 247-258.
Bearman, D. (2006) "Moments of risk. Identifying threats to electronic records", *Archivaria*, Fall, Vol 62, pp 15-46.
Berman,S. (2010) "Cross-border challenges for e-Discovery", *Business Law International*, Vol 2, No. 2, pp. 123-131.
Blanchette, J.F. and Johnson, D.G. (2002) "Data Retention and the Panoptic Society. The social benefits of forgetfulness", *The Information Society*, Vol 18, pp. 33-45.
Boudrez, F., Dekeyser, H., and Dumortier, J. (2005) *Digital Archiving. The new challenge*, IRIS, Mont Saint Guibert.
Bovens, M. (2006) "Analysing and assessing public accountability. A conceptual framework", [online], http://www.connex-network.org/eurogov/.../egp-connex-C-06-01.pdf.
Chaplin, D.T. (2008) "Conceptual Search. ESI, Litigation and the issue of Language", Paper read at DESI II. Second International Workshop on Supporting Search and Sensemaking for Electronically Stored Information in Discovery Proceedings, London, June.
Conrad, J. (2010) "E-Discovery revisited: the need for artificial intelligence beyond information retrieval", *Artificial Intelligence and Law*, Vol 18, No. 4, pp 321-345.
Cox, R.J. and Samuels, H.W. (1988) "The archivist's first responsibility. A research agenda to improve the identification and retention of records of enduring value", *The American Archivist*, Winter/Spring, Vol 51, pp. 28-42.
Croasdell, D.T. (2001) "IT's role in Organizational Memory and learning", *Information Systems Management,* Vol 18, No. 1, pp 8-11.
Duranti, L. (1997) "The preservation of the integrity of electronic records", *Proceedings of the DLM-Forum on electronic records. Brussels, 18-20 December 1996*, EU, Luxemburg, pp 60-65.
EDRM (2006) "EDRM E-Discovery Reference Model", [online], http://www.edrm.net.
Epler, M.J. (2006) *Managing information quality. Increasing the value of information in knowledge-intensive products and processes*, Springer, Berlin, 3rd edition.
Gartner (2009) "MarketScope for eDiscovery Software Product Vendors report 2009", [online], http://www.gartner.com/DisplayDocument?id=1262421.
Görg, C., Stasko, J., and Liu, Z. (2008) "Jigsaw. Supporting investigative analysis through interactive visualization", *Information Visualization,* Vol 7, No. 2, pp 118-132.
Grossman, M.R. and Cormack, G.V. (2011) "Technology-assisted review in E-Discovery can be more effective and more efficient than exhaustive manual review", *Richmond Journal of Law and Technology*, Vol 17, No. 3, Article 11.
Groth, P.T. (2007) *The origin of data. Enabling the determination of provenance in multi-institutional scientific systems through the documentation of processes*, Southampton.
Henseler, J. (2010a) *E-Discovery. Op zoek naar de digitale waarheid*, Amsterdam University Press, Amsterdam.
Henseler, J. (2010b) "Network-based filtering for large email collections in E-Discovery", *International Journal of Artificial Intelligence and Law*, Vol 18, No. 4, pp. 413-430.
Hilbert, M. and López, P. (2011) "The world's technological capacity to store, communicate, and compute information", *Science*, Vol 332, No. 6025, pp. 60-65.
Hockx-Yu, H. (2006) "Digital Preservation in the Context of Institutional Repositories", *Program: Electronic Library & Information Systems,* Vol 40, No. 3 pp. 232-243.
IDC (2011) "Extracting value from chaos", [online], http://www.wmc.com/leadership/programs/digital-universe.htm.
Jacobs, A. (2009) "The pathologies of Big Data", *Communications of the ACM,* Vol 52, No. 8, pp 36-44.
Keena, J.R. (2002) "E-Discovery. Unearthing Documents Byte by Byte", [online], http://www.mnbar.org/benchandbar/2002/mar02/ediscovery.htm.
Kozushko, H. (2003) "Digital Evidence", [online], http://infohost.nmt.edu/~sfs/Students/HarleyKozushko/Papers/DigitalEvidencePaper.pdf.
Krause, J. (2009) "In search of the perfect search", [online], http://www.abajournal.com/magazine/in_search_of_the_perfect_search.
Latour, B. (1990) "Postmodern? No, simply amodern! Steps towards an anthropology of science", *Studies In History and Philosophy of Science,* Vol 21, No. 1, pp 145-171.
Levy, D.M. (2001) *Scrolling forward. Making sense of documents in the digital age.* Arcade Publish-

ing, New York.
Marchand, D.A., Kettinger, W.J. and Rollins, J.D. (2001) *Information Orientation: The Link to Business Performance*, Oxford University Press, New York (NY).
Murphy, B. (2011) "e-Discovery in the cloud not as simple as you think", [online], http://www.forbes.com/sites/jasonvelasco/2011/11/29/e-discovery-in-the-cloud-not-as-simple-as-you-think/.
Oard, D.W., Baron, J.R., Hedin, B., Lewis, D.D., Tomlinson, S. (2010) "Evaluation of information retrieval for eDiscovery", *Journal of Artificial Intelligence and Law*, Vol 18, No. 4, pp 347-386.
Paul, G. and Baron, J. (2007) "Information Inflation. Can the legal system adapt ?", *Richmond Journal of Law & Technology* Vol 13, No. 3, pp 1-41.
Peck, Judge A. (2011) "Search, Forward: Will manual document review and keyword searches be replaced by computer-assisted coding?", *Law Technology News*, October 2011.
Pederson, A., Routledge, D. and Thurston, A. (1999) *Managing Public Sector Records: Case Studies, Vol. 1, Vol. 2*, International Records Management Trust, London.
Pipek, V., Wulf, V., and Johri, A. (2012) "Bridging artifacts and actors: expertise sharing in organizational ecosystems", *Computer Supported Cooperative Work*, Vol 21, No. 2-3, pp. 261-282.
Redman, T.C. (2004) "Data: an unfolding quality disaster", *DM Review*, Vol 8, No. 8, pp 22-23, 57.
Saracevic, T. (2007) "Relevance: a review of the literature and a framework for thinking on the notion in information science. II.", *Journal of the American Society for Information Science and Technology*, Vol 58, No. 3, pp. 755-776.
Rosen, J. (2012). "The Right to be Forgotten", [online], http://www.stanfordlawreview.org/online/privacy-paradox/right-to-be-forgotten.
Scholtes, J. C. (2009) *Text mining: de volgende stap in zoektechnologie*, Maastricht University, Maastricht.
Sedona Conference (2004) "The Sedona Principles. Best practices, recommendations & principles for addressing electronic document production", [online], http://www.thesedonaconference.org/dltForm?did=SedonaPrinciples200401.pdf.
Socha, G., and Gelbmann, T. (2004) "Socha Gelbmann Electronic Discovery Survey", [online], http://www.sochaconsulting.com/Publications/LawTechnologyNews Article 8.04.pdf.
Thompson, S.H.T., Srivastava, S.C., and Jiang, L. (2008-2009) "Trust and Electronic Government Success: An Empirical Study", *Journal of Management Information Systems*, Vol 25, No. 3, pp. 99-132.
Toebak, P.M. (2010) *Records Management. Gestaltung und Umsetzung,* Hier + Jetzt, Baden.
Valukas, A.R. (2010) "Report of Anton R. Valukas in the Lehman Brothers Holdings, Inc.", [online], http://www.scribd.com/doc/28228424/Lehman-Brothers-Examiner-s-Report-Vol-1.
Van Bussel, G.J. (2011) "Accountability and the reconstruction of the past", Castelnova, W. and Ferrari, E. (eds.), *Proceedings of the 5th European Conference of Information Management and Evaluation (ECIME 2011)*, Academic Publishing Limited, Como, pp 483-490.
Van Bussel, G.J. (2012) *'Archiving should be just like an Apple™' en acht andere, nuttige (?) stellingen*, Amsterdam University Press, Amsterdam.
Van Bussel, G.J. and Ector, F.F.M. (2009) *Op zoek naar de herinnering. Verantwoordingssystemen, content-intensieve organisaties en performance,* VBDS, Helmond.