Digital Archiving, Green IT and Environment. Deleting Data to Manage Critical Effects of the Data Deluge

Author(s)
van Bussel, Geert-Jan ; Smit, Nikki; van de Pas, John

Creating Tomorrow

Digital Archiving, Green IT and Environment. Deleting Data to Manage Critical Effects of the Data Deluge

Geert-Jan van Bussel[1], Nikki Smit[2] and John van de Pas[3]
[1] HvA Amsterdam University of Applied Sciences, School of Economics and Management, Amsterdam, The Netherlands
g.j.van.bussel@hva.nl
[2] BECIS Information Management, Amsterdam, The Netherlands

[3] Saxion University of Applied Sciences, School of Creative Technology, Deventer, The Netherlands
j.h.vandepas@saxion.nl

Abstract: The development of the World Wide Web, the emergence of social media and Big Data have led to a rising amount of data. Information and Communication Technologies (ICTs) affect the environment in various ways. Their energy consumption is growing exponentially, with and without the use of 'green' energy. Increasing environmental awareness has led to discussions on sustainable development. The data deluge makes it not only necessary to pay attention to the hard- and software dimensions of ICTs but also to the 'value' of the data stored. In this paper, we study the possibility to methodically reduce the amount of stored data and records in organizations based on the 'value' of information, using the Green Archiving Model we have developed. Reducing the amount of data and records in organizations helps in allowing organizations to fight the data deluge and to realize the objectives of both Digital Archiving and Green IT. At the same time, methodically deleting data and records should reduce the consumption of electricity for data storage. As a consequence, the organizational cost for electricity use should be reduced. Our research showed that the model can be used to reduce [1] the amount of data (45 percent, using Archival Retention Levels and Retention Schedules) and [2] the electricity consumption for data storage (resulting in a cost reduction of 35 percent). Our research indicates that the Green Archiving Model is a viable model to reduce the amount of stored data and records and to curb electricity use for storage in organizations. This paper is the result of the first stage of a research project that is aimed at developing low power ICTs that will automatically appraise, select, preserve or permanently delete data based on their 'value'. Such an ICT will automatically reduce storage capacity and reduce electricity consumption used for data storage. At the same time, data disposal will reduce overload caused by storing the same data in different formats, it will lower costs and it reduces the potential for liability.

Keywords: data deluge, digital archiving, archival retention levels, information value chain, green archiving, green IT

1. Setting the stage: Data Deluge, Digital Archiving and Green IT

1.1 Data Deluge

The development of the World Wide Web, the emergence of social media and Big Data have led to a rising amount of data (Armitage and Roberts, 2002; Segaran and Hammerbacher, 2009; Manyika, 2011). The seemingly infinite opportunities to process and publish data, global electronic communications, an explosion in devices located at the periphery of the network, including embedded sensors, smartphones, and tablet computers, aerial sensory technologies, software logs, cameras, microphones, radio-frequency identification readers, wireless sensor networks, and a large-scale digitization of cultural heritage such as film, music, art, images, maps, and text, have caused an unprecedented global growth in the amount of data. This growth has been analysed in several research projects, but the comparison of their results is difficult because of the different definitions and research methods used (Lyman and Varian, 2003; Hilbert and López, 2011; Gantz and Reinsel, 2012). This research agrees on one basic fact: the astonishing growth rate in the amount of data in the world. To summarize the results, it is confirmed [1] that the data storage capacity doubles every 40 months, and [2] that the annual growth rate in the amount of data is almost 40 %, creating a 'data deluge'. This data creates new opportunities for analytics in human genomics, health care, oil and gas, search, surveillance, finance, and many other areas (Golden, 2010), but is also putting great pressure on the infrastructures of information and communication technologies (ICTs) (Van Bussel and Henseler, 2013).

1.2 Digital Archiving

The use of collaborative technologies in organizations to streamline business processes also creates huge amounts of data (Jacobs, 2009). These data are used and generated by knowledge workers who engage in peer-to-peer knowledge sharing across organizational boundaries. The storage, dissemination and processing of this data require complex ICT systems. These ICT systems present security and durability challenges that pose a major threat for information quality (Bearman, 2006). Digital data are fragile. They are easily altered without recognition. They require storage media that have relatively short life spans, and access technologies that are changing extremely fast. For some data types, such as multimedia, it is almost impossible to be used outside the proprietary environments in which they were generated (Hodge, 2000). These problematic challenges threaten the trustworthiness of organizational records, that data that are meant to be (and used as) evidence for policies, decisions, products, actions and transactions. Organizations have to respond to increasing societal demands for the trustworthiness of these records, mostly for privacy, accountability and transparency reasons. That is why Digital Archiving (DA) is important for organizations.

DA ensures that the informational and evidential 'value' of records is utilized in business processes to improve performance. It provides an ICT infrastructure to (indefinitely) store (identified and trusted) records and keep them accessible. It ensures that (privacy) laws and regulations are respected and audits periodically the possibility to reliably reconstruct the past. DA manages the four dimensions of information to allow for such a reconstruction. Those four dimensions are [1] quality, [2] context, [3] relevance, and [4] survival. The quality dimension is focused on the quality requirements of data and records to realize 'immutable mobiles' (Latour, 1990). 'Immutable mobiles' allow for the repeated use of data and records for consultation and for reconstruction of past happenings. Context provides meaning to the data and records: metadata are captured that give information about the organizational, technological, and societal environment in which the data and records were generated. Data and records are only relevant if they fit the organizational objectives of performance and accountability. The survival dimension concerns the security and durability challenges, which have to be overcome to realize access, retrieval, and preservation over time for all 'immutable mobiles' (Van Bussel 2012ab). DA's purposes are to reduce the costs of transactions, to enlarge the speed of access to organizational experiences, to help in decision-making, to share knowledge, and to realize accountability.

The deluge of data is threatening DA's possibilities to realize its purposes (Van Bussel, 2012ab; Van Bussel and Henseler, 2013).

1.3 Green IT

ICT has not always worked to the benefit of environmental sustainability, although there are many ICTs that have positive environmental effects, such as GPS systems and online mapping software, which lead to more efficient travel and, as a result, reduce emissions of carbon dioxide (Tomlinson, 2010). The origins of an environmental approach to ICTs can be traced back to the beginning of the 1990s, when the reduction of the use of hazardous materials, the maximization of energy efficiency, and the recyclability or biodegradability of defunct products and factory waste became hot items in computing (Jacob and K.G, 2012; Esfahani et al, 2015ab).

Green IT (Brooks et al, 2012) is defined by Murugesan (2008: 25-26) as 'the study and practice of designing, manufacturing, using, and disposing of computers, servers, and associated subsystems - such as monitors, printers, storage devices, and networking and communications systems - efficiently and effectively with minimal or no impact on the environment'. ICTs affect the environment in various ways. Its production requires electricity, raw materials, chemical materials and large amounts of water, and supplies (often toxic) waste (Robinson, 2009). Computers and peripherals are changed two or three years after purchase (Murugesan, 2008). In 2006, global production of E-waste was estimated at 20-50 million tonnes per year (UNEP, 2006). In rich countries, E-waste represents some 8 percent of municipal waste (Widmer et al, 2005). It is the fastest growing municipal waste stream (EPA, 2011). Most of this E-waste is not recycled, because those items tend to go out with the normal household waste and do not receive special treatment (Ladou and Lovegrove, 2008). Some 80 percent of collected E-waste is exported to poor countries and ends up in landfills and informal dumps (Schmidt, 2006). These dumping sites are poisoned and groundwater is polluted (Murugesan, 2008).

Green IT has been introduced to minimize environmental effects of ICTs, to save costs and for corporate social responsibility (CSR). There are four paths along which the environmental effects of ICTs should be addressed: green use (reducing the energy consumption of ICTs and use them in an environmentally sound manner), green disposal (refurbish and reuse old ICTs and properly recycle unwanted ones), green design (designing energy efficient and environmentally sound ICTs), and green manufacturing (manufacturing ICTs with minimal or

no impact on the environment) (Murugesan, 2008). Green IT can also develop, according to Donnellan, Sheridan, and Curry (2011), solutions that align IT processes with the principles of sustainability and stimulate innovative technologies to deliver green benefits across an organization. In that way, end user satisfaction, management restructuring, regulatory compliance, fiscal benefits, and return on investment (ROI) can be addressed. In the opinion of Visalakshi et al. (2013: 64), Green IT may be 'simple, plain, common sense'. The positive effects of Green IT are extensively studied in academic literature (Harmon and Auseklis, 2009; Brooks et al, 2012; Lei and Ngai, 2013; Subburaj et al, 2014; Esfahani et al, 2015ab).

The energy consumption of ICTs (as well as the corresponding energy costs) are growing exponentially as a result of the data deluge. From 2000 to 2005 consumption of electricity in data centers doubled, while electricity consumption worldwide grew by (only) 16.7 percent per year (Koomey, 2008). From 2005 to 2010, the consumption of electricity in data centres alone jumped with 56 percent (Koomey, 2011; Cook, 2012). This increase in electricity consumption results in increased carbon dioxide emissions. According to Dubey and Hefley (2011), each PC or laptop in use generates about four tons, each server about eight tons of carbon dioxide every year, although there are many possibilities to lower those emissions (Boccaletti et al, 2008). The use of 'renewable' energy resources (water, wind, solar, geothermal, tidal, wave, and biofuel resources) could affect these emissions positively (WNA, 2015), but in 2014 the use of renewable energy resources for electricity generation is still quite low. The U.S. Energy Information Administration estimates that in 2012 almost 21 percent of the world electricity generation was from the use of renewable energy resources, with a projection for nearly 25 percent in 2040 (EIA, 2012; EIA, 2014). This means that carbon dioxide emissions will be a problem in the foreseeable future. Curbing back data storage could have very positive effects on energy use. In 2008, storage networks were responsible for 15 percent of total ICT energy costs (HP, 2008). This percentage had, in our estimate, doubled in 2011, given the increasing need for data storage as a result of multiplication of data, social media, and fear of not being compliant (Van Bussel, 2012a). Studies have shown that electricity costs can approach 50 percent of the overall energy costs for an organization (Harmon and Auseklis, 2009). In January 2013, an average in-house server in the USA costs $731,94 in electricity (Hammond, 2013).

Summarizing: ICTs have a large energy footprint. The electricity use for ICTs has shown remarkable growth, which resulted in rising costs for electricity consumption. The data deluge (and the use of more and more ICT resources to manage this deluge) threatens [1] to drown all positive effects of Green IT and [2] to raise energy costs exponentially.

2. Research question, objective, and methodology

2.1 Research Question

Market research firm IDC estimated in 2007 that the amount of annually generated data exceeded the storage space globally available (Gantz and Reinsel, 2007). The data deluge is threatening to prevent both DA and Green IT to reach their objectives. To keep data and records accessible over time, to allow for Green IT to reach its environmental effects, and to prevent energy costs from rising unnecessarily, it becomes vital to curb data storage.

The bulk of all preserved data and records is stored on hard disks, consuming more electricity than necessary. Although electricity use of servers can be largely reduced by using them efficiently keeping the powered-up servers utilized and keeping as many as possible powered-down, this does not work in archiving and storage infrastructures. Research has shown that search and data mining activities for access will spread fairly evenly across all servers, making it impossible to keep servers powered-down (Adams et al, 2011; Adams et al, 2012). For curbing data storage, it will be necessary to appraise data and records value (over time), to implement data and records value appraising methods and tools, and to completely and permanently delete data and records that have lost their economic, social, cultural, financial, administrative, fiscal and/or legal value (Robyns, 2014; Niu, 2014).

The research question we want to answer in this paper is if it is possible to methodically reduce the amount of stored data and records in organizations based on the value of information, using the Green Archiving Model we have developed. Reducing the amount of data and records in organizations helps in allowing them to fight the data deluge and to realize the objectives of both DA and Green IT. At the same time, methodically deleting data and records should reduce the consumption of electricity for data storage. As a consequence, the organizational cost for electricity use should be reduced.

2.2   Research Objective

We have tested the viability of our Green Archiving Model in previous research with two exploratory case studies. In Van Bussel and Smit (2014) we stated that this model could be used to increase awareness in organizations for the environmental aspects of data storage and that the objectives of DA and Green IT could be realized using the model. The objective of this paper is to ascertain that our Green Archiving model can be used to methodically reduce the amount of stored data and records based on their value and that it can be used to reduce the costs for consumption of electricity for data storage.

Green Archiving intends to raise awareness of the environmental effects of ICTs (like increased carbon dioxide emissions) and to the effects of the data deluge on the accessibility of data and records. It tries to define solutions for [1] the rising amount of data and records and [2] the constantly rising costs of electricity. Green Archiving integrates Green IT with two leading theories of DA: the theories concerning the Information Value Chain (IVC) and Archival Retention Levels (ARLs). Both theories can be used to reduce the amount of stored data and records based on assigned information values. Operationalizing Green Archiving, organizations curb power consumption, lower needs for storage capacity by permanently deleting data and records based on their value, and develop 'low power' ICTs (Forrest, Kaplan and Kindler, 2008). That way, Green Archiving realizes the objectives of Green IT and curbs data storage, allowing DA to realize its objectives of fast accessibility of past experiences, transparent accountability, data and records-driven decision-making, knowledge-sharing, and reducing costs of transactions (Barosso and Hölzle, 2007; Schwarz and Elffers, 2010; Orgerie et al, 2014; Van Bussel and Smit, 2014; Pine and Mazmanian, 2015). Green Archiving is a relatively new subject and is not extensively studied yet within the context of information and archival sciences.

2.3   Research Methodology

This paper is based on the research project that was first reported on in Van Bussel and Smit (2014). We are using here the results mentioned in that paper, with additional results from one of the exploratory case studies used. We have added also results from an extensive research in existing case studies on DA, Green IT and data deduplication. Our exploratory research was a combination of desk research, qualitative interviews with information technology and information management experts, a focus group and two exploratory case studies. We researched scientific literature with an ICT, information management and archival science perspective. We collected literature with a key word search in Google Scholar, Microsoft Academic Search and in the Digital Library of the University of Amsterdam (indexes on IT, information science / management, archival science / management). The key words used in this search were: 'Green Computing', 'Green IT', 'IT power use', 'IT power costs', 'information value', 'archival appraisal', 'archival disposal' and 'environmental awareness'. The findings of this desk research were used, discussed and criticized in: [1] individual, semi-structured interviews with ten ICT, information management and archival science experts (three scientists, two consultants, three CTO's, and two storage industry specialists); [2] a focus group, consisted of six (other) experts (two Green Computing consultants, two information managers and two storage managers). We used the information acquired through desk research, interviews and focus group to develop a provisional Green Archiving model. In Van Bussel and Smit (2014), this model was than tested for validity in two small exploratory case studies.

3.   Exploring the stage: Composing the Green Archiving Model

3.1   Green IT Components

Analyzing literature, published case studies, interviews and focus group discussion, we discern six components of Green IT research. The first is product longevity (Visalakshi et al, 2013; Agarwal and Nath, 2011). As Walker (1995: 21) already stated, 'product's longevity is influenced by the durability of its component parts; its capacity to be repaired, maintained and upgraded; and its aesthetic qualities'. Product longevity helps in ensuring an intelligent utilization of resources in manufacturing processes, which account for 70 percent of the natural resources used during a computers lifecycle (Mingay, 2007). The second component of Green IT is software and deployment optimization (Ahamad and Ravikanth, 2012; Badbe, 2014; Choudhary, 2014), that influences the amount of computer resources required for any given computing function. It is a way for saving energy that includes algorithmic efficiency, resource allocation, virtualization, and terminal servers. Virtualization is a very popular method of optimization in archiving and storage environments. With it, many virtual versions of

devices or resources (servers, storage devices, networks or operating systems) could be made, using only one actual device or resource, making it much more energy efficient (Ren et al, 2015). The third component is power management (Schlomann et al, 2015; Kashyap et al, 2015; Visalashki et al, 2013). Even with a turned off monitor, a computer will consume as much energy as a powered but idle computer. Almost one-third of the energy consumption of an organization's PC population is wasted as a result of PC's that are unused but still turned on. With power management, it is possible for organizations to annually save up to $60 per computer (Gunaratne et al, 2005). This translates also in a reduction in the pollution emissions from reduced electrical generation. Materials recycling (Shalini and Prasanthi, 2015; Visalashki et al, 2013; Kwon et al, 2006), the fourth component, refers to recycling or reuse of computers or electronic waste, including finding other uses for ICTs, or having them dismantled, allowing for the safe extraction of materials for reuse in other products. Recycling ICTs can keep harmful materials such as lead, mercury, and chromium out of landfills (Murugesan, 2008). Telecommuting, often referred to as teleworking, occurs when paid workers work away from their normal place of work, usually from home. It is not clear if the reduction of a company's energy consumption matches the rise in energy consumption at home in order to work remotely, but reducing the amount of cars moving employees back and forth will produce a carbon dioxide emission reduction (Asgari and Jin, 2015; Srivastava et al, 2015; Thompson, 2009). The last, sixth component, low power IT (or energy-efficient computing) (Hopper and Rice, 2008; Düben et al, 2014; Ahmad and Ranka, 2012; Lee et al, 2013) has been designed to use less electronic power. The increasing electricity costs are forcing hardware developers to rethink their technologies. A variety of approaches have been proposed to trade the accuracy of the hardware fabric in return for savings in resources used such as energy, delay, area and/or yield and, therefore, lead to reduced costs. The components software and deployment optimization, telecommuting, and low power IT are important for reducing power consumption for data storage.

3.2   Archival retention levels and information value chain

Information and Archival science are interdisciplinary fields concerned with the analysis, collection, classification, storage, retrieval, dissemination, appraisal, disposal and preservation of data. They use methods and techniques to appraise and select organizational data for long-term (or indefinite) preservation or to permanently delete appraised data (Shepherd and Yeo, 2003; Xie, 2013; Smallwood, 2013). Appraisal is the process of establishing the 'value' of data and records, qualifying that value, and determining its duration. The primary objective of appraisal is to identify the data and records to be continuously preserved for an unlimited period of time (Duranti, 1994). Appraisal establishes the value of organizational data over time, be it economic, social, cultural, financial, administrative, fiscal and/or legal value (Cook, 2013). Many appraisal approaches are based on content evaluation of records (very impractical in an electronic age of information overload), causing a decontextualization. Shepherd and Yeo (2003: 151) stated that appraisal 'should be based on analysis of organizational purposes and the systems that support them'. The focus of appraisal is moved from the data and records to the organizational contexts that created them. In the appraisal process, 'judgements of value' are made to decide what to keep and what to destroy. Penn (2014) re-interprets appraisal through the philosophical frameworks of axiology and demonstrates that the concept of value has a wider resonance than has been previously considered. Appraisal always results in retention schedules, which assure that all data and records are retained and disposed according to their quantified value: the time (in years) that data and records should be retained, according to considerations of organizational risks and assigned economic, social, cultural, financial, administrative, fiscal and/or legal value. Minimizing risks (especially those of litigation) also means systematic disposal immediately after the expiration of the assigned retention period (Shepherd and Yeo, 2003). Two theories of archival science offer tools for appraising data and records: the theories of Archival Retention Levels (ARLs) (Den Teuling, 2001) and Information Value Chain (IVC) (Van Bussel, 2012ab).

The first theory concerns itself with designating ARLs in organizations to store and retain data that is unique, authentic, relevant and contextual. The ARL theory is part of appraisal methods that consider organizational contexts and purposes more important than the specific content of data and records. ARLs define detailed functional (organizational) responsibilities for the retention, storage and archiving of unique, authentic, relevant, and contextual data and records (Smit, 2012). Data value is appraised according to the organizational level that is responsible for the collection, analysis, processing, and storage of that specific data. This organizational level is the designated ARL. At the ARL the data are retained as long as the retention schedule permits. This schedule makes the economic, social, cultural, financial, administrative, fiscal and/or legal value of the data and records (retained at every ARL) explicit and defines its archival value: a time (in years) after which this information should be irreparably destroyed. Identical data retained at other functional levels within the organization and without a new business objective (duplicates) can be immediately deleted, permanently and

irreparably. When using ARL's, it will be necessary (to prepare for litigation procedures) to capture data about the organizational level these duplicates are being kept and the persons who have accessed them (Van Bussel and Henseler, 2013). In digital environments, these duplicates can be stored in different forms and places and in various business processes, not being the designated ARL (Paul and Baron, 2007). The effect of using ARL's on the organizational need for storage capacity will be substantial, as published case studies indicate. In hospitals, for instance, an average organization's duplicate rate in 2009 was typically between 5-10 percent (McClellan, 2009). Deduplication lets an organization keep 20 times more data in a given amount of storage (Geer, 2008). The practical experiments of Mandagare et al (2008) show that between different deduplication techniques the space savings amount for almost 30 percent, which was confirmed in Dutch (2008) and Proofpoint (2013). Using ARL checklists can reduce the amount of data stored up to a minimum of 30 percent, which has direct effects on costs and needed storage capacity. The organizational use of ARLs can be seen as contextual data deduplication.

The IVC theory defines the utilization of the informational and evidential value of data and records in business processes to improve trusted information management and the performance of business processes (Van Bussel 2012ab). The IVC includes all processes of information management and manages data generation, data identification, data capture, data storage, data processing, data distribution, data structuring, data publication, data (re-) use, data appraisal, data selection, data disposal, data retention, data security, data auditing and data preservation. DA uses this chain to reach its purposes: to reduce the costs of transactions, to enlarge the speed of access to organizational experiences, to help in decision-making, to share knowledge, and to realize accountability. For the purpose of this paper, only the processes of data appraisal, data selection and data disposal are important. In the data appraisal process the short- and long-term (or indefinite) value of data and records is defined in order to retain and preserve them for later (re-)use. As stated above, this data appraisal defines the archival value and results in a retention schedule. In the data selection process, data and records are collected and set aside according to the agreed upon retention schedule. In the data disposal process, the set aside data and records are completely, permanently and irreparably deleted (Shepherd and Yeo, 2003). Organizational retention schedules are used to operate those processes. Almost 75 percent of all data and records in an organization can be permanently deleted over time (Archieflandsverordening, 2007). The value of Big Data, for instance, degrades rapidly over the short term. Retaining that data for a long time, hoping it may become valuable or needed some day, is unnecessarily costly and indefensibly risky (Gascon, 2013). In a retention schedule, such a Big Data data set will be appraised to be destroyed after its last use. Normally that would mean up to one year after the moment that specific Big Data data set was last used.

3.3 Green Archiving Model

Combining the components of Green IT with the data reducing components of DA, allowed us to develop a Green Archiving model. As our earlier research showed (Van Bussel and Smit, 2014), that model can be used: [1] to increase awareness in organizations for the environmental effects of the use of ICTs, [2] to reduce the amount of stored data and records, [3] to reduce power consumption for data storage and, ultimately, [4] to reduce greenhouse gas emissions and E-waste in realizing all components of Green IT. This paper concentrates on the aspects [2] and [3] of the model. Aspect [1] was added to the objectives of the Green Archiving Model when conducting interviews with Digital Archiving and Green IT specialists and exploring the research question within a focus group. In their professional experience, the specialists encountered an extremely low organizational awareness of the environmental effects of ICTs and a lack of knowledge about the electricity use needed for storage and the associated organizational costs. Aspect [4] of the model will be part of further research; there are many case studies confirming the fact that Green IT reduces carbon dioxide emissions and E-waste (Murugesan, 2008; Schmidt, 2006). In future case studies we want to research the total effects of the application of the Green Archiving Model. The model of Green Archiving we developed is shown in figure 1.

Archival Retention Level

Green IT

ARL-Checklists

Destroy

1. Product Longeveity
2. Deployment / Optimization
3. Power Management
4. Recycling
5. Telecommuting
6. Low Power IT

Green Archive

Destroy

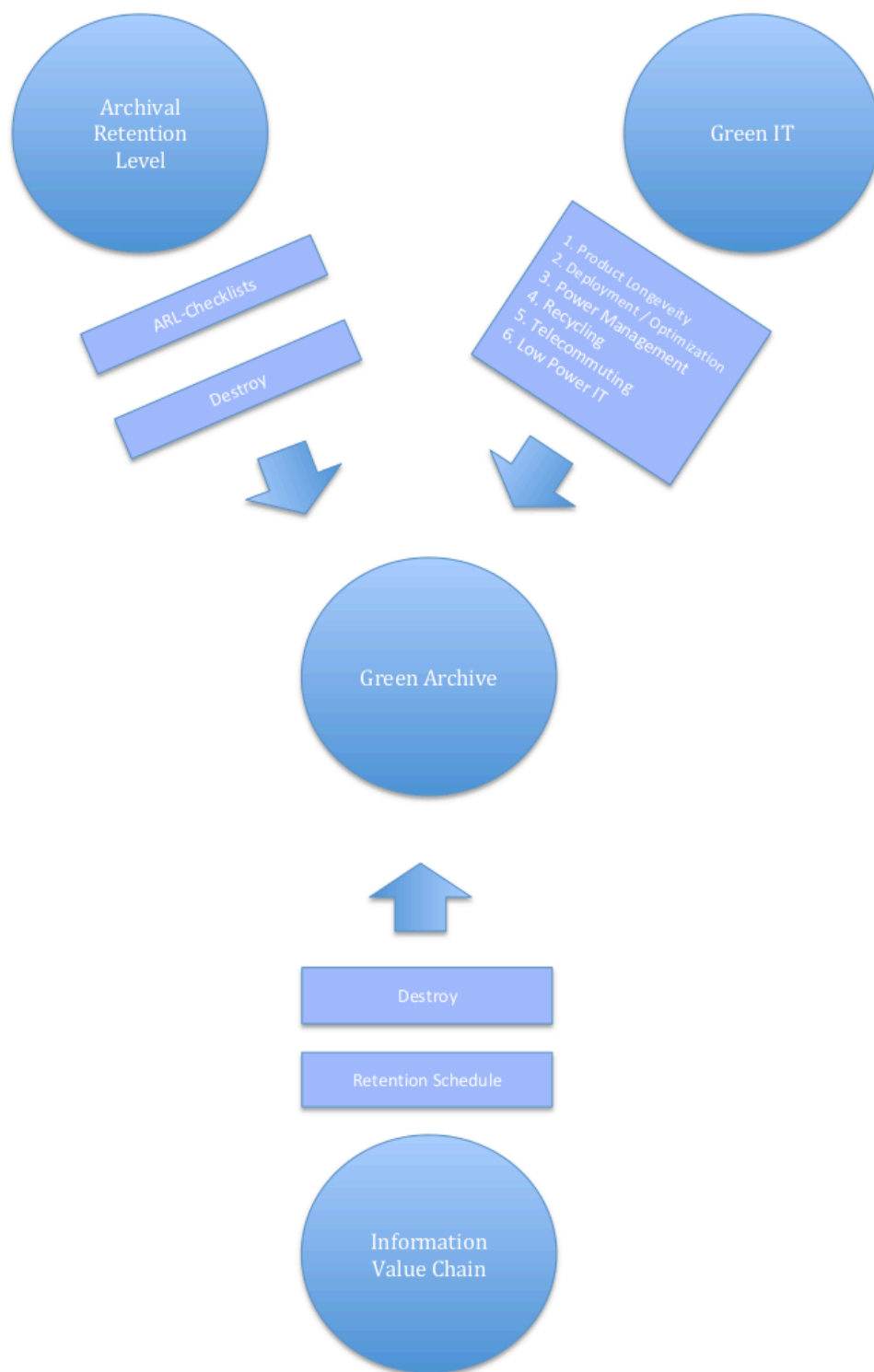Retention Schedule

Information Value Chain

Figure 1: Green archiving model: combination of Green IT, ARL and IVC

4. Testing the stage: experimenting with the Model

4.1 Purpose

The Green Archiving Model has four objectives. In Van Bussel and Smit (2014), we ascertained in a case study in the Dutch Music Institute (Nederlands Muziek Instituut) that the Green Archiving Model was viable and that it could be used in organizations to increase awareness of the environmental effects of ICTs (objective 1). In that paper, we also presented some provisional results in reducing the amount of data (objective 2), and in curbing power use for data storage (objective 3), using a case study in a Dutch international trade organization. We are presenting more definite results of using the Green Archiving Model here, especially concentrating on objectives 2 and 3, following additional research in this trade organization and already published case studies. The fourth objective (reducing carbon dioxide emissions and reducing E-waste) will be addressed in future research.

4.2 Exploratory research

We organized three different exploratory case studies in a small international trade corporation in Maastricht (The Netherlands), working with subsidiaries in Europe, Asia and South America. These three case studies were: [1] a scan of the ICT infrastructure of the corporation using the model (November-December 2013), [2] a pilot study by the corporation's IT department on the effects of ARL Schedules (March-May 2014), and [3] a pilot study by de corporation's Chief Information Officer on the effects of a Records Schedule on the globally stored data and records (January-March 2015). In [1] the Green Archiving model was enthousiastically received. Green Computing was well known within the IT department, but only the components Software Deployment and Optimization (virtualization of storage servers) and Power Management were implemented. The results of this implementation of these components were comparable to those described by Dubey and Hefley (2011). The IT department admitted that it should be more aware of other Green Computing components. CSR was extremely important for the corporation and implementing other components of Green Computing would be a significant contribution to CSR. The IT department planned to look into the possibilities of Telecommuting and Product Longevity. When we did the exploratory scan, the organization didn't use ARL checklists, but (for this case study) agreed to experiment with them in its corporate headquarters. After a scan of the headquarter's file systems, the IT department estimated that almost 35 % of their IT storage capacity of 18 TB was used for duplicate files. It acknowledged that the use of ARL checklists would have a significant effect on the IT storage capacity. Retention Schedules were used only for the data stored in their document and records management applications, but the IT department acknowledged that both applications were not yet generally in use. Rigorous use of those schedules would certainly have an effect on the IT storage capacity, but the IT department could not quantify those effects yet. In [2] the IT department organized a pilot study on the effects of ARL Schedules in their global information management environment. It analysed all file systems and database management systems, storage area networks ans mail systems in the global operations of the corporation. In this pilot study, the corporation realized the functional responsibilities in the organization structures of their subsidiaries were not clearly defined. It developed a new information management structure based on the IVC theory and integrated it in their business process models. The ARL analysis confronted the IT department with ICTs that were implemented within their subsidiaries without their knowledge. Global data storage capacity was 45 TB (including the headquarter's 18 TB storage capacity). The conclusion of this pilot study was that the use of ARL checklists would diminish global data storage capacity with 30 percent. 37 percent of the company's data storage capacity was used for duplicate files. The IT department calculated that such a reduction of the amount of data would result in less electricity use for data storage and would diminish electricity costs with 25 percent. These results correspond with the results of data deduplication mentioned earlier in this paper. In case study [3} the office of the Chief Information Officer analysed the retention schedules in use in their records and document management applications. They were limited to data and records captured within the document and records management software that was not globally implemented yet. The Chief Information Officer decided to use the European Document Retention Guide 2013 (De Brauw, 2013) as a way to analyse the possibilities of further reducing the data and records stored. The Chief Information Officer concluded that almost 10 percent (3.5 TB) of all data and records stored in February 2015 could be disposed of immediately, because they had no 'value' anymore for the organization as all possible retention periods had passed. Based on the Guide only 5 percent of all data would have to be retained for longer than twenty years. The IT department calculated that using ARLs and Retention Schedules for all data and records in the organization would curb data storage with almost 45 percent and would reduce electricity costs with 35 percent. The

conclusion of the Chief Information officer was that methodical use of the value of information (as expressed within the Retention Schedule used) would improve storage efficiency, reduce the amount of stored data and records, diminish litigation risks, and reduce electricity costs. The Green Archiving model seems to be a viable model for organizational use.

5. Conclusions and future work

In this paper, we studied the possibility to methodically reduce the amount of data and records stored in an organization based on their 'value' and using the Green Archiving Model. The case study showed that the model can be used to reduce [1] the amount of data (45 percent, using ARLs and Retention Schedules) and [2] the electricity consumption for data storage (resulting in a (calculated) cost reduction of 35 percent). These case studies indicate that the Green Archiving Model is a viable model to methodically reduce the amount of stored data and records and to curb electricity use for storage in organizations. That way, it facilitates DA and Green IT to reach their objectives. These case studies only provide us with provisional results. They need to be confirmed in further research. We are planning extensive case studies to research the environmental effects of Green Archiving and the scientific viability of our model. The ultimate goal of this research project is the development of a low power ICT that will automatically appraise, select and preserve or permanently delete data and records. Such an ICT will automatically reduce storage capacity and curb electricity consumption used for data storage. At the same time, data disposal will reduce overload caused by storing the same data in different formats, it will lower costs and it reduces the potential for liability.

References

Adams, I., Miller, E.L., and Rosenthal, D.S. (2011). *Using Storage Class Memory for Archives with DAWN, a Durable Array of Wimpy Nodes*. Technical Report UCSC-SSRC-11-07, Santa Cruz: University of California.

Adams, I.F., Storer, M.W., and Miller, E.L. (2012). 'Analysis of workload behavior in scientific and historical long-term data repositories', *ACM Transactions on Storage (TOS)*, vol. 8, no. 2, art. 6: 1-27.

Agarwal, S. and Nath, A. (2011) 'Green Computing - a new Horizon of Energy Efficiency and Electronic waste minimization: a Global Perspective', *CSNT '11 Proceedings of the 2011 International Conference on Communication Systems and Network Technologies,* Washington: IEEE, pp. 688-693.

Ahamad, S.F., and Ravikanth, P.V. (2012) 'Green Computing Future of Liveliness', *International Journal of Computational Engineering Research (IJCER),* vol. 2, no. 4, pp. 41-45.

Ahmad, I., and Ranka, S. (2012) *Handbook of energy-aware and Green Computing,* London: Chapman & Hall/ CRC,

Archieflandsverordening 2007, MvT, National Archives of the Netherlands Antilles.

Armitage, J. and Roberts, J. (2002) *Living with Cyberspace. Technology and society in the 21st century,* London, New York: Continuum.

Asgari, H., and Jin, X. (2015) 'Toward Comprehensive Telecommuting Analysis Framework: Setting the Conceptual Outline', *Transportation Research Board 94th Annual Meeting*, No. 15-0282.

Badbe, V.C. (2014) 'Green Computing Approaches', *International Journal of Computer Technology and Applications*, vol. 5, no. 6, pp. 3042-3049.

Barroso, L.A. and Hölzle, U. (2007) 'The case for energy-proportional computing', *IEEE Computer,* vol. 40, no. 12, pp. 33–37.

Bearman, D. (2006) 'Moments of risk. Identifying threats to electronic records', *Archivaria*, vol 62, Fall, pp. 15-46.

Boccaletti, G., Löffler, M., and Oppenheim, J. M. (2008) 'How IT can cut carbon emissions', *McKinsey Quarterly* vol. 37, October, pp. 37-41.

Brooks, S., Wang, X., and Sarker, S. (2012) 'Unpacking Green IS: A Review of the Existing Literature and Directions for Further Research', Vom Brocke, J., Seidel, S. and Recker, J. (eds.), *Green Business Process Management. Towards the Sustainable Enterprise,* Heidelberg: Springer, pp. 15-37.

Choudhary, S (2014) 'A survey on green computing techniques', *International Journal on Computer Science and Information Technology*, vol.5, no. 5, pp. 6248-6252.

Cook, G. (2012) *How clean is your cloud ? Catalysing an energy revolution,* Amsterdam: Greenpeace International..

Cook, T. (2013) 'Evidence, memory, identity, and community: four shifting archival paradigms', *Archival Science,* vol. 13, no. 2-3, pp. 95-120.

De Brauw Blackstone Westbroek (2013) *European Document Retention Guide 2013,* Amsterdam: Iron Mountain..

Den Teuling, A. (2001) 'Selectie in de praktijk', Ketelaar, F.C.J., Den Teuling, A.J.M. and Van Wijngaarden, J. (eds.), *Archiefbeheer in de praktijk,* Alphen aan den Rijn: Samsom, pp. 4020/1-55.

Donnellan, B., Sheridan, C. and Curry, E. (2011) 'A Capability Maturity Framework for Sustainable Information and Communication Technology', *IEEE IT Professional,* vol. 13, no. 1, pp. 33–40.

Düben, P.D., Joven, J., Lingamneni, A., McNamara, H., De Micheli, G., Palem, K.V. and Palmer, T.N. (2014) 'On the use of inexact, pruned hardware in atmospheric modelling', *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 372, pp. 1-16, DOI: 10.1098/rsta.2013.0276.

Dubey, S., and Hefley, W.E. (2011) 'Greening ITIL: Expanding the ITIL lifecycle for Green IT', *Proceedings of Portland International Conference on Management of Engineering and Technology (PICMET 2011),* Washington: IEEE, pp. 1-8.

Duranti, L. (1994) 'The concept of appraisal and archival theory', *The American Archivist*, vol. 57 (Spring), pp. 328-344.

US Energy Information Administration (EIA) (2012) 'International Energy Statistics'. Online.
Available: http://www.eia.gov/cfapps/ipdbproject/IEDIndex3.cfm?tid=6&pid=29&aid=12 (26 May 2015)

US Energy Information Administration (EIA) (2014) 'International Energy Outlook 2013'. Online.
Available: http://www.eia.gov/forecasts/archive/ieo13/electricity.cfm (26 May 2016)

US Environmental Protection Agency (EPA) (2009) *Electronic Waste Management in the United States through 2009*, Washington D.C.: EPA.

Esfahani, M.D., A.A. Rahman and N.H. Zakaria (2015a) 'The Status Quo and the prospect of Green IT and Green IS. A systematic literature review', *Journal of Soft Computing and Decision Support Systems*, vol. 2, no. 1, pp. 18-34.

Esfahani, M.D., A.A. Rahman and N.H. Zakaria (2015b) 'GreenIT/IS Adoption as Corporate Ecological Responsiveness. An Academic Literature Review', *Journal of Soft Computing and Decision Support Systems*, vol. 2, no. 1, pp. 35-43.

Forrest, W., Kaplan, J.M. and Kindler, N. (2008) 'Data Centers: How to cut Carbon Emissions and Costs', *McKinsey on Business Technology*, vol. 8, no. 14, pp. 4-14.

Gantz, J.F. and Reinsel. D. (eds.) (2007) *The expanding digital universe. A forecast of worldwide information growth through 2010,* Framingham (Ma.): IDC.

Gantz, J. and Reinsel, D. (2012) *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East,* Framingham (Ma.): IDC.

Golden, B. (2010) 'Cloud computing: How big is big data? IDC's answer', *CIO*, May 7.

Geer, D. (2008) 'Reducing the Storage Burden via Data Deduplication', *Computer*, vol.41, no. 12, pp. 15-17.

Gunaratne, C., Christensen, K. and Norsman, B. (2005) 'Managing energy consumption costs in desktop PCs and LAN switches with proxying, split TCP connections, and scaling of link speed'. *International Journal of Network Management*, vol. 15, pp. 297-310.

Hammond, T. (2013) 'Toolkit: Calculate datacenter server power usage'. Online.
Available: http://www.zdnet.com/toolkit-calculate-data-center-server-power-usage-7000013699/ (26 May 2015)

Harmon, R.R. and Auseklis, N. (2009) 'Sustainable IT Services: Assessing the Impact of Green Computing Practices', *Proceedings of Portland International Conference on Management of Engineering and Technology (PICMET 2009),* Washington: IEEE, pp. 1707-1717.

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.

Hodge, G.M. (2000) 'Best practices for Digital Archiving. An Information Life Cycle Approach', *D-Lib Magazine*, vol. 6, no. 1. Online. Available: http://www.dlib.org/dlib/january00/01hodge.html (26 May 2015).

Hopper, A. and Rice, A. (2008) 'Computing for the future of the planet', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 366, pp. 3685-3697, DOI: 10.1098/rsta.2008.0124.

HP (2008) 'Groen computergebruik: energie besparen en helpen de aarde te beschermen'. [Online].
Available: http://m.hp.com/be/nl/hp-news/details.do?id=48932&articletype=news_release (26 May 2015).

Gascon, D. (2013) 'Are you a Data Hoarder? Really, are you?', *Compliance, Governance & Oversight Council, CGOc.* Online. Available: https://www.cgoc.com/blog/are-you-data-hoarder-really-are-you (25 May 2015).

Jacob, J.S. and P. K.G (2012) 'EDPAC: event-driven power aware pervasive computing for effective power utilization in Green Computing', *International Journal of Advanced Information Technology*, vol. 2, no. 1, pp. 55-64.

Jacobs, A. (2009) 'The pathologies of Big Data', *Communications of the ACM,* vol 52, no. 8, pp. 36-44.

Ren, J., Li, L., Chen, H., Wang, S., Xu, S., Sun, G., Wang, J. and Liu, S. (2015) 'On the deployment of information-centric network: Programmability and virtualization'. *International Conference on Computing, Networking and Communications (ICNC), Proceedings,* Garden Grove (CA.): IEEE, pp. 690-694.

Kashyap, P., Naik, A., and Patel, P. (2015) 'Green Computing', *International Journal of Research*, vol. 2, no. 4, pp. 940-944.

Koomey, J.G. (2008) 'Worldwide electricity used in data centers', *Environmental Research Letters,* vol. 3, no. 034008 (July-September), doi:10.1088/1748-9326/3/3/034008).

Koomey, J.G. (2011) *Growth in data center electricity use 2005-2010,* Oakland (Ca.): Analytics Press

Kwon, Y.C., Sang, W.L. and Moon, S. (2006) 'Personal computer privacy: Analysis for Korean PC users', *Advances in Information and Computer Security*, Berlin-Heidelberg: Springer, pp. 76-87.

Ladou, J., and Lovegrove, S. (2008) 'Export of electronics equipment waste', *International journal of occupational and environmental health*, vol. 14, no. 1, pp. 1-10.

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.

Lei, C.F. and Ngai, E.W.T. (2013) 'Green IT Adoption: An Academic Review of Literature'. *PACIS 2013 Proceedings.* Paper 95.

Lee, S.M., Park, S-H. and Trimi, S. (2013) 'Greening with IT: practices of leading countries and strategies of followers', *Management Decision*, vol. 51, no. 3, pp. 629-642.

Lyman, P. and Varian, H.R. (2003) *How much information?, ,* Berkely: School of Information Management and Systems, University of California.

Mandagere, N., Zhou, P., Smith, M.A. and Uttamchandani, S. (2008) 'Demystifying data deduplication', *Proceedings of the ACM/IFIP/USENIX Middleware'08 Conference Companion*, Leuven (Belgium), pp. 12-17.

Manyika, J. (ed.) (2011) *Big Data. The next frontier for innovation, competition and productivity,* McKinsey Global Institute.

McClellan, M.A. (2009) 'Duplicate Medical Records: A Survey of Twin Cities Healthcare Organizations', *AMIA Annual Symposium Proceedings,* Bethesda (Mi.): AMIA, pp. 421-425.

Mingay, S. (2007) *Green IT: The New Industry Shock Wave*, Gartner.

Murugesan, S. (2008) 'Harnessing Green IT: Principles and Practices', *IEEE IT Professional*, vol. 10, no. 1, pp. 24-33.

Niu, J. (2014). 'Appraisal and Selection for Digital Curation', *International Journal of Digital Curation*, vol. 9, no. 2, pp. 65-82.

Orgerie, A.C., De Assuncao, M.D. and Lefevre, L. (2014) 'A Survey on Techniques for Improving the Energy Efficiency of Large Scale Distributed Systems', *ACM Computing Surveys*, vol. 46, no. 4, pp. 1-35.

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.

Penn, ESM (2014) *Exploring archival value: an axiological approach*. Doctoral thesis, UCL (University College London). Online. Available: http://discovery.ucl.ac.uk/1455310/ (26 May 2015)

Pine, K. and Mazmanian, M. (2015) 'Emerging Insights on Building Infrastructure for Data-Driven Transparency and Accountability of Organizations', *iConference 2015 Proceedings,* Newport Beach (Ca.)*,* pp. 1-13

Proofpoint (2013) 'Archive and Governance ROI Tool'. Online. Available: http://www.proofpoint.com/products/archive-governance/roi-tool/index.php (26 May 2015).

Robinson, B.H. (2009) 'E-waste: an assessment of global production and environmental impacts', *Science of the total environment*, vol. 408, no. 2, pp. 183-191.

Robyns, M. C. (2014). *Using Functional Analysis in Archival Appraisal: A Practical and Effective Alternative to Traditional Appraisal Methodologies*. Plymouth: Rowman & Littlefield.

Schmidt, C. W. (2006) 'Unfair trade e-waste in Africa', *Environmental Health Perspectives*, vol. 114. no. 4, A232.

Shalini, K., and Prasanthi, K.N. (2013) 'Green Computing', *Journal of Telematics and Informatics*, vol. , no. 1, pp. 1-13.

Schwarz, M., and Elffers, J. (2010) *Sustainism is the new modernism. A cultural manifesto for the sustainism era*, New York: Distributed Art Publishers.

Segaran, T. and Hammerbacher, J. (2009) *Beautiful Data: The Stories Behind Elegant Data Solutions,* Sebastopol (Ca.): O'Reilly Media.

Shepherd, E. and Yeo, G. (2003) *Managing records. A handbook of principles and practice*, London: Facet Publishing.

Schlomann, B., Eichhammer, W., and Stobbe, L. (2015) 'Energy saving potential of information and communication technology', *International Journal of Decision Support Systems*, vol. 1, no. 2, pp. 152-163.

Smallwood, R.F. (2013) *Managing electronic records: Methods, best practices, and technologies,* Hoboken (NJ): John Wiley & Sons, Hoboken (NJ), chapter 7.

Smit, N. (2012) *Green Archiving. Digitale archivering en het milieu,* Amsterdam: Reinwardt Academy (BCM Thesis).

Dutch, M. (2008) *Understanding data deduplication ratios,* SNIA Data Management Forum.

Srivastava, K., Sethumadhavan, A., Raghupathy, H., Agarwal, S., and Rawat, S. R. (2015) 'To Study the Indian Perspective on the Concept of Work from Home', *Indian Journal of Science and Technology*, vol. 8, no. S4, pp. 212-220.

Subburaj, S., Kulkarni, S. and Jia, L. (2014) 'Green IT: sustainability by aligning business requirements with IT resource utilisation', *International Journal of Communication Networks and Distributed Systems*, vol. 12, no. 1, pp. 30-46.

Thompson, J.T. (2009) 'Three approaches to green computing on campus', *EDUCAUSE Quarterly Magazine*, vol. 32, no. 3. [Online]. Available: http://www.educause.edu/ero/article/three-approaches-green-computing-campus (26 May 2015).

Tomlinson, B. (2010) *Greening through IT. Information Technology for environmental sustainability,* Cambridge (Ma.): MIT University Press.

UNEP (2006) *Call for Global Action on E-waste*. United Nations Environment Programme, s.l.

Van Bussel, G.J. (2012a) *Archiving should be just like an Apple*TM *en acht andere, nuttige (?) stellingen*, Amsterdam: Amsterdam University Press.

Van Bussel, G.J. (2012b) 'Reconstructing the Past for Organizational Accountability', *The Electronic Journal of Information Systems Evaluation,* vol. 15, no. 1, pp. 127-137.

Van Bussel, G.J. and Henseler, H. (2013) 'Digital Archiving and e-Discovery: Delivering evidence in an age of overload', John, B., Nkhoma, M. and Leung, N. (eds.), *Proceedings of the 4*th *international conference on Information Systems Management and Evaluation (Ho Chi Minh City, May 2013)*, Reading (GB): ACPI, pp. 281-288.

Van Bussel, G.J. and Smit, N. (2014) 'Building a Green Archiving Model: Archival Retention Levels, Information Value Chain and Green Computing', Devos, J. and De Haes, S. (eds.), *Proceedings of the 8th European Conference on IS Management and Evaluation. ECIME 2014. University of Ghent, 11-12 september 2014,* Reading (GB): ACPI, pp. 271-277.

Visalakshi, P., Paul, S. and Mandal, M. (2013) 'Green Computing', *International Journal of Modern Engineering Research (IJMER). Proceedings of the National Conference on Architecture, Software systems and Green computing (NCASG), Paiyanoor (India), May 3, 2013,* pp. 63-69.

Walker, S. (1995) 'The environment, product aesthetics and surface', *Design Issues*, vol.11, no. 3, pp. 15-27.

Widmer, R., Oswald-Krapf, H., Sinha-Khetriwal, D., Schnellmann, M. and Boni, H. (2005) 'Global perspectives on e-waste', *Environmental Impact Assessment Review,* vol. 25, no. 5, pp. 436-458.

World Neclear Association (WNA) (2015) 'Renewable energy and electricity'. Online. Available: http://www.world-nuclear.org/info/Energy-and-Environment/Renewable-Energy-and-Electricity/ (26 May 2015).

Xie, S.L. (2013) 'Preserving Digital Records: InterPARES findings and developments', Lemieux, V. (ed.), *Financial Analysis and Risk Management. Data Governance, Analytics and Life Cycle Management*, Berlin-Heidelberg: Springer, pp. 187-206.

Geert-Jan van Bussel, Nikki Smit and John van de Pas

Geert-Jan van Bussel, Nikki Smit and John van de Pas