# Green Archiving.

*Archiving for a sustainable world*

**Author(s)**
van Bussel, G.J.; Smit, Nikki

**Publication date**
2014

**Document Version**
Author accepted manuscript (AAM)

# Green Archiving
# Archiving for a sustainable world

**Nikki Smit BCH:** Research Group Digital Archiving & Compliance HvA Amsterdam University of Applied Sciences, School of Design and Communication, Amsterdam, nsmitters@gmail.com

**Dr Geert-Jan van Bussel:** Professor of Digital Archiving & Compliance
HvA Amsterdam University of Applied Sciences, School of Design and Communication, Amsterdam, g.j.van.bussel@hva.nl

**Abstract**: *Computers create environmental problems. Their production requires electricity, raw materials, chemical materials and large amounts of water, and supplies (often toxic) waste. They poison dumping sites and pollute groundwater. In addition, the energy consumption in IT is growing exponentially, with and without the use of 'green' energy. Increasing environmental awareness within information science has led to discussions on sustainable development. 'Green Computing' has been introduced: the study and practice of environmentally sustainable computing or IT. It is necessary to pay attention to the value of the information stored. In this paper, we explored the possibilities of combining Green Computing components with two theories of archival science (Archival Retention Levels and Information Value Chain respectively) to curb unnecessary power consumption. Because in 2012 storage networks were responsible for almost 30 % of total IT energy costs, reducing the amount of stored information by the disposal of unneeded information should have a direct effect on IT energy use. Based on a theoretical analysis and qualitative interviews with an expert group, we developed a 'Green Archiving' model, that could be used by organizations to 1] reduce the amount of stored information, and 2] reduce IT power consumption. We used two exploratory case studies to research the viability of this model.*

**Keywords**: Green Archiving, Green Computing, Sustainable IT, CSR, Archival Science.

## Introduction

The development of the World Wide Web together with its possibilities and opportunities has led to changes in information management. The information deluge is putting great pressure on information technology (IT) infrastructures. These infrastructures barely can process the enormous amounts of information (Van Bussel & Henseler, 2013). IT affects the environment in various ways. Its production requires electricity, raw materials, chemical materials and large amounts of water, and supplies (often toxic) waste. Computers and peripherals are changed two or three years after purchase. IT dumping sites are poisoned and groundwater is polluted (Murugesan, 2008). In addition to this, as a result of the information deluge, IT energy consumption is growing exponentially, with and without the use of 'green' energy. From 2000 to 2005 power consumption of data centers doubled, while power consumption worldwide grew by 16.7% per year (Koomey, 2008). From 2005 to 2010, the power consumption of data centers alone jumped with 56 % (Koomey, 2011; Cook, 2012). Storage networks in 2008 were responsible for 15% of total IT energy costs (HP, 2008). This percentage has doubled in 2012, given the increasing need for information storage as a result of the multiplication of 'useful' information, social media, and the fear of not being compliant (Van Bussel, 2012). This increase in energy consumption results in increased greenhouse gas emissions. According to Dubey & Hefley (2011), each PC or laptop in use generates about four tons, each server about eight tons of carbon dioxide every year!

**Purpose and research method**

The influx of information within organizations, the constantly rising costs of energy and the increased greenhouse gas emissons are reason enough to study possibilities to curb power consumption, lower storage capacity and develop 'low power' IT. Studies have shown that IT power costs can approach 50 % of the overall energy costs for an organization (Forrest, Kaplan & Kindler, 2008). Combined with an increasing environmental awareness within information science, discussions have emerged on sustainable development (Schwarz, Elffers 2010). Within the hard- and software industry, Green Computing (or Green IT or ICT Sustainability) has been introduced: the study and practice of environmentally sustainable IT, to save costs (mainly) and for corporate social responsibility (CSR) (Harmon & Auseklis, 2009). Although Green Computing is an important development, to curb storage in an information deluge it is necessary to pay attention to information value (over time) and, thus, to completely destroy information that has lost its value. In this paper, we want to develop a model of Green Archiving, which can be used by organizations to 1] reduce IT power consumption, and 2] reduce the amount of stored information. The research method we used was a combination of desk research, qualitative research (interviews) and two small exploratory case studies. We researched scientific literature with an IT, information management and archival science perspective. The findings of this desk research were used and tested during interviews with an expert group. This group consisted of independent IT, information management and archival science experts known for their expertise and innovative ideas. The information acquired through these interviews was important for this study, for Green Archiving is a new subject and is not extensively studied within the context of information science.

**Theoretical foundations**

The theoretical foundations of Green Archiving are to be found in scientific literature on Green Computing (Brooks, Wang & Sarker, 2012) and archival science (Shepherd & Yeo, 2003). Murugesan (2008, p. 25-26) defines Green Computing as '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'. Analyzing scientific literature, we discern six components of Green Computing research: product longevity (Agarwal & Nath, 2011), software and deployment optimization (such as virtualization) (Ahamad & Ravikanth, 2012), power management (Murugesan, 2008), materials recycling (Murugesan, 2008), telecommuting (Thompson, 2009), and low power IT (Ahmad & Ranka, 2012). Those components were confirmed in the expert group interviews. Archival science uses methods and techniques to appraise and select organizational records for long-term (or indefinite) preservation (Shepherd & Yeo, 2003). Important for this study are two theories of archival science: the theories of Archival Retention Levels (ARLs) (Den Teuling, 2001) and the Information Value Chain (IVC) (Van Bussel, 2012). Designating ARLs in organizations is, as Smit (2012) states, information alignment: it defines detailed functional responsibilities for the retention and storage of information. Only on the designated ARL information needs to be retained. Identical information (duplicates) retained on other organizational levels may be disposed of. In a digital environment, duplicates are (according to Paul & Baron (2007)) stored in different forms and places and in various business processes, not being the designated ARL. In hospitals, in 2009 an average organization's duplicate rate was typically between 5-10% (McClellan, 2009). It is (non-scientifically) estimated that in 2013 in most organizations 30% of all files are duplicates (Proofpoint, 2013). Using ARL checklists can seriously reduce the amount of information stored. According to Van Bussel (2012), the IVC ensures that the informational and evidential 'value' of information is utilized in business

processes. 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. This chain is instrumental for the performance of business processes and the provision of trusted information (Van Bussel, 2012). In this paper, we concentrate on the appraisal, selection and disposal processes, in which 1] the value of information is appraised, 2] that information that has lost its value over time is selected and 3] that 'unneeded' information is completely disposed of (Shepherd & Yeo, 2001). Retention schedules are used to operate those processes. Almost 80% of all information in an organization can be destroyed over time, as was postulated in the expert group. It was emphasized that the effects on energy use could be substantial.

**Model of 'Green Archiving'**

Based on desk research and interviews with the expert group (in which many theoretical findings were elaborated and confirmed), we were able to develop a Green Archiving model. We think this model can be used to 1] reduce power consumption (and environmental waste), and 2] reduce the amount of stored information. This model of Green Archiving combines Green Computing with the archival theories of ARLs and IVC (figure 1).

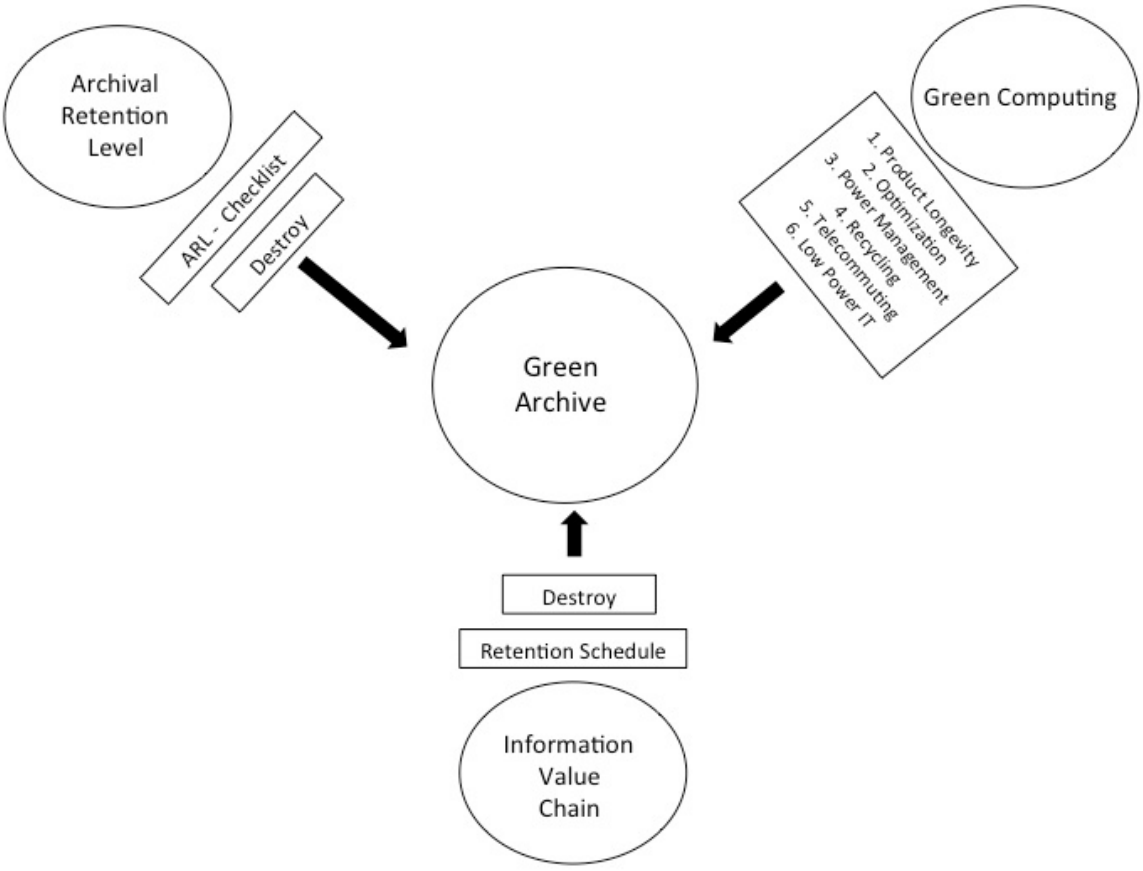


Figure 1. Green Archiving Model: combination of Green Computing, ARL and IVC

**Two small exploratory case studies**

We used this model in two case studies. The purpose of these case studies was to ascertain 1] that the model was viable, and 2] that it could be used in organizations to curb power use. We organized our first case study in the Nederlands Muziek Instituut (Dutch Music Institute), a small organization that operates as national heritage centre for musicians and composers. In using the model, we ascertained that environmental awareness was low. The concept of Green

Computing was unknown, as were the environmental effects of IT. That may be a result of the fact that the Institute used the IT infrastructure of the Koninklijke Bibliotheek National Library of the Netherlands (The Hague) for its storage network. The Institute never discussed energy use and power costs for their two terabytes (TB) of data storage capacity with their hosting partner. Within the Institute, ARL Checklists were not in use, but it was acknowledged (after being instructed about their purpose) that its use would reduce the amount of duplicates within office automation. Because the capacity of office automation storage was not known, it was not possible to quantify this effect. In addition, the Institute didn't use retention schedules for its digital collection and its business processes. In reality, even unneeded information was not disposed of. We concluded the Green Archiving model was a viable model to check environmental awareness, but that this case study didn't give us any information about the possibilities of the model to be used to curb power use. The second case study was in a small international import export corporation, working with subsidiaries in Europe, Asia and South America. In this organization the Green Archiving model was enthusiastically received. Green Computing was well known within the IT department, but not all six components were implemented. The company emphasized Optimization and Power Management. The results were comparable to those described by Dubey & Hefley (2011). The organization didn't use ARL Schedules, but experimented with them in its corporate headquarters. The IT department estimated that almost 35 % of their IT storage capacity of 8 TB was used for duplicate files. It acknowledged that the use of ARL schedules would have a significant effect on the IT storage capacity. The company intends to organize a pilot to study the effects of ARL schedules globally. Retention Schedules were used only for the information 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. The company wants to use the Green Archiving model in its pilot to the effects of Green Archiving on the company's storage capacity, energy costs and Green Computing efforts. We concluded that in this case study the Green Archiving model seemed to be a viable model for organizational use in curbing power use.

**Conclusions and future work**

Both case studies were exploratory; we can only state here that it seems that our hypothesis that the Green Archiving model is a viable model to study possibilities to reduce the amount of stored information and to curb power consumption in organizations, is correct. Much more research is needed to scientifically prove that Green Archiving is really influencing power consumption and the amount of stored information. In future research, we want to organize extensive case studies to research the organizational effects of Green Archiving and the viability of our model. Ultimately, we want to study the possibility of developing a system which will automatically appraise, select and destroy unneeded information. That system will reduce storage capacity and curb power consumption used for information storage, while simultaneously ensuring that organizations have a documented understanding of the past. This is very important for eGovernment, for delivering evidence, and for improving organizational transparency and accountability.

**References**

Agarwal, S. & 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* (pp. 688-693). Washington: IEEE.

Ahamad, S.F., & Ravikanth, P.V., Green Computing Future of Liveliness. *International Journal of Computational Engineering Research (IJCER)*, Vol. 2, Number 4, July-August. (pp. 41-45).

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

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

Cook, G. (2012). *How clean is your cloud ? Catalysing an energy revolution*. (p. 11) Amsterdam: Greenpeace International.

Den Teuling, A. (2001). Selectie in de praktijk. In: F.C.J. Ketelaar, A.J.M. den Teuling. J. van Wijngaarden (Eds.), *Archiefbeheer in de praktijk*. (pp. 4020/1-55). Alphen aan den Rijn: Samson.

Dubey, S., & 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*). (pp. 1-8). Washington: IEEE.

Forrest, W., Kaplan, J.M. & Kindler, N. (2008). Data Centers: How to cut Carbon Emissions and Costs. *The McKinsey Quarterly*, Vol. 8, Number 14. (pp. 4-14).

Harmon, R.R. & 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)*. (pp. 1707-1717). Washington: IEEE.

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

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

Koomey, J.G. (2011). *Growth in data center electricity use 2005-2010*. (pp. iii, 5). Oakland (Ca.): Analytics Press.

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

Murugesan, S. (2008). Harnessing Green IT: Principles and Practices. *IEEE IT Professional*, January– February 2008 (pp. 24-33).

Paul, G. & Baron, J. (2007). Information Inflation. Can the legal system adapt?, *Richmond Journal of Law & Technology*. Vol 13, Number 3. (pp. 1-41).

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

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

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

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

Thompson, J.T. (2009). Three approaches to green computing on campus. *EDUCAUSE Quarterly Magazine*, 32 (3). Online. http://www.educause.edu/ero/article/three-approaches-greencomputing-campus.

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

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