Carbon footprint of the knowledge industry and ways to reduce it

Prof. Gobinda Chowdhury, Director, Centre for Information and Knowledge Management, University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia
Gobinda.Chowdhury@uts.edu.au

Prof. Michael Fraser, Director, Communications Law Centre, University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia
Michael.Fraser@uts.edu.au

World Digital Libraries 4(1): 9–18

Abstract

The current practices of content creation, access, and super distribution followed by the global content industry are neither economically nor environmentally sustainable. Since research studies on the carbon footprint of the content industry are few and far between, adequate data are not available. This paper shows that the current practices in the production of printed content produce a massive carbon footprint. It is argued that these emissions can be significantly reduced by developing a digital content supply network. Features of a digital content network are briefly discussed along with an indication of how this can help reduce carbon footprint of information services. The paper also points out areas of further research in this direction.

Introduction

Since the beginning of the Gutenberg era, the information industry has followed a typical business model, where content is produced as bundled products such as a book with several chapters, a journal issue with a pre-determined number of papers, and so on. A certain number of such bundled products—copies of books, copies of a certain issue of a journal, and so on—are produced by publishers in order to attain economies of scale. Such information products are either purchased by consumers for a set price or are purchased and managed by libraries for use by their clients. This model continued till about 15 years ago, with some minor changes taking place during the past four decades or so due to the arrival of remote online databases, which provide access to individual articles of journals, conference proceedings, and so on, via subscription. Throughout this period, the basic model for content access was either through purchase of bundled content by the consumer or access through libraries, barring some exceptional types of information resources such as free newspapers, institutional reports and publications, pamphlets, and so on. Essentially, consumers were either required to pay for the content or access content through libraries. The latter played a key social role in the democratization of access to knowledge.

Although it is not often highlighted as one of the major players in national economies, the size of the content industry and its contribution to the economy is quite significant. A somewhat dated figure reveals:

- Total **annual revenues of US book publishers amounts to** $26.8 billion (2004 figures) (Eco-Libris 2007);
- **Total annual revenues of European book publishers amounts to** Euro 22.3 billion (2004 figures) (Eco-Libris 2007);
- In the UK, the publishing industry, which is the second largest in Europe, has a turnover of over £18.4 billion, with its 8000 plus companies employing around 164,000 people and contributing to over 8% of GDP (http://www.berr.gov.uk/whatwedo/sectors/publishing/index.html);
- The estimated value-addition of the copyright industry in Australia is approximately 100 billion Australian dollars, which is about 10% of the GDP, while the annual compound growth of the industry is 4.7%, as compared to 3.6% of GDP (Price Waterhouse Coopers 2008); and
- Nielsen Book Scan reports that in 2008, 61 million books were sold retail in Australia, with a total value of Aus$1.21 billion (Nielsen 2009).

In 2007, the value of the global book market was estimated to be $127.5 billion, which is estimated to rise to $160.7 billion in 2012 (up 26 % from 2007). In 2008, the US book publishers' net revenues touched $40.32 billion (Healey 2009). If we consider the entire content industry, the size of the market will be several times larger. Yet, the processes of content creation, distribution, and access follow an extremely traditional business model. In this model, consumers have two choices: either to purchase priced content, which is edited or peer-reviewed and is thus reliable in terms of value; or to access free content, which may or may not be edited or peer-reviewed and hence may or may not be of a certain quality. Nevertheless, in most cases, the consumers have to acquire the entire work, even if they want to use only a small portion of the item. Consumers often wish not only to access granular content, but also to aggregate granular content elements and then super distribute the aggregated content package(s). Taking publishing as an example, consumers must either:

- Buy books and subscribe to journals, and then photocopy or scan the granular elements they actually want for physical aggregation and for physical distribution; or
- Go online from one proprietary site to another, and get access to the digital content

in different ways, on different terms and conditions, and then attempt to repurpose and distribute extracts of content within the limits of the various different digital rights management systems and licences.

Till date, the overall process of content access, use, and super distribution, which is very common in academic institutions like universities, remains very cumbersome. Complex and onerous statutory and commercial blanket licensing schemes have been developed to address this market failure. In response, a contested grey economy of copies has been built by one section of the consumer community to form a legitimate market for content that meets consumers' needs.

In this paper, it is argued that the prevailing business model is not environmentally sustainable, and that while other industries and businesses are concerned with reducing their carbon footprint, the processes of knowledge creation, access, and distribution and the content industry that supports these activities are not concerned with properly addressing and mitigating the impact of the industry on matters related to environment and ecology. The lack of detailed studies on the carbon footprint and environmental impact of knowledge-intensive activities hints at the hypocrisy amongst scholars, academics, and researchers who talk about and study the carbon footprint of every sphere other than the knowledge sector. This paper proposes a model of new content service network that will help reduce the carbon footprint of the content industry and of knowledge-intensive activities. However, this initial study has to be supported by more detailed ones on the environmental impact of knowledge creation, distribution, and access.

Environmental impact of the content industry

Studies on the environmental impact of the content industry are few and far between. However, the data gathered from these studies reveal some massive figures.

According to certain estimates, 10.2 kg of CO_2 is generated by each printed book over its lifetime (Ritch 2009). Therefore,

- 235.6 million books sold annually in the UK (Booktrade.info 2010) would generate 2.4 million tonnes of CO_2;
- 1 billion books produced annually in Germany (Space Daily 2009) would generate 10.2 million tones of CO_2; and
- 4.15 billion books produced in the USA (Eco-Libris 2007) would generate 42.3 tonnes of CO_2.

Thus, the annual book production in Germany and USA combined would generate 52.5 million tones of CO_2.

To put these figures into perspective, the CO_2 emission from power plants—the world's most significant contributors to greenhouse gases—in various countries are as follows (Center for Global Development 2007):

- US: 2.8 billion tones;
- China: 2.7 billion tonnes;
- Russia: 661 million tonnes;
- India: 583 million tonnes;
- Japan: 400 million tonnes;
- Germany: 356 million tonnes;
- Australia: 226 million tonnes;
- South Africa: 222 million tonnes;
- UK: 212 million tonnes; and
- South Korea: 185 million tonnes.

(*Note*: The figures above are in US tonnes, where one tonne is equivalent to 2,000 pounds, as opposed to a metric tonne, which is equivalent to 2,204 pounds).

Another way of understanding the environmental impact of the content industry could be to make a comparison with the following figures:

- The average carbon footprint per person in the UK is 10 tonnes (www.climatechangewales.org.uk/public/?id=112) and

- More than 30 million **trees are cut down annually to produce books sold in the US** alone (Eco-Libris 2007)

While the total carbon footprint figure for the entire journal publication industry is not available, a study commissioned by a leading journal publisher Reed Elsevier noted that the total carbon footprint for producing the journal *Fuel* in 2007 was 'just over 40 tonnes of carbon dioxide, and an ecological footprint of almost 21 global hectares' (http://www.reedelsevier.com/corporateresponsibility08/PDFFiles/fuel-footprint-study-exec-sum.pdf). The figure in terms of the total carbon footprint of printed journals will be staggeringly high if the total number of all the print journals in the world is multiplied by this figure, and the transportation and distribution costs of print journals around the globe are also taken into account!

Environmental impact of copying

The actual carbon footprint figures are cause for greater concern if other factors are taken into consideration. Let us take photocopying as an example. Photocopying under statutory licences is a legal activity, although it is well understood that the total number of copies that are made under the fair dealing provision of the copyright law, and those that are copied illegally, will be several times the figures of legally licensed copy pages. Even these figures are alarming. Copyright Agency Limited (CAL), Australia estimated that in 2008–09, over 2 billion pages were copied legally under licences in Australian educational, government, and some business institutions (CAL 2009). Assuming that an average book contains 300 pages (Kozak 2003), this amounts to about 50,000 tonnes of CO_2 emission.

According to the HP carbon footprint calculator (HP Eco solutions 2009) a large office photocopier (Canon imageRUNNER 3300) consumes 2,736 kWh electricity (assuming that it prints 100,000 pages per year), and emits 3,087 kg of CO_2. As per this calculation, the CO_2 emission for photocopying 2 billion pages will be (at the rate of 3,087 kg per 100,000 pages) 617,740 tonnes.

This figure may vary depending on the type of photocopier and its frequency of use, special facilities like air conditioners, maintenance and repair, and so on**.** It should be noted that there is an environmental cost involved even when a photocopier is idle. According to one calculation (RISO 2009), the cost of printing 100,000 pages in a month would be $107.88 per year, but the cost of electricity use in idle and sleep mode would be $382.50. The study also states that 'a mid-volume (20–44 copies per minute) copier in a low-volume office can use 70% more energy per page than an efficient low-volume (under 20 copies per minute) copier.'

However, even using the rough estimate, the environmental cost of photocopying 2 billion pages per year under the CAL license yields a massive 617,770 tonnes of CO_2. It must be noted that this figure of 2 billion photocopies include only those pages that are copied under the license—the number of pages copied for super distribution as permitted within the CAL licences. This does not include the pages that are copied by people for their personal use within the fair dealing provisions and other exceptions in the copyright law. In reality, the number of copies made in this manner will be at least equal to—if not more than—copies made under the CAL licenses. So, even if we make a conservative estimate, the CO_2 emissions for photocopying could be 1.2 million tonnes or more.

To put this into perspective, in Australia, the per capita CO_2 emission in 2007 was 18.8 tonnes (http://www.iea.org/co2highlights/CO2highlights.pdf). Taking into account Australia's population of 22 million, and considering the figure of 1.2 million tonnes of CO_2 emission from photocopying alone, it may be noted that the per capita CO_2 emission from photocopying alone was 0.055 tonnes, or 55 kg per person.

Now, let us look at this scenario from a different perspective. CO_2 emissions from legal licensed photocopying in Australia are equivalent to the annual CO_2 emissions from four power plants. This observation is based on figures from Carbon Monitoring for Action (CARMA) (http://carma.org/region/detail/18) data, which state that the 766 power plants in Australia together generate 224 million tonnes of CO_2 per year. As mentioned earlier, this calculation does not include CO_2 emissions from photocopying done in the wider corporate sector, nor does it include copying of copyright-free content, copying under exceptions or illegal copying. Taking photocopying of all kinds of knowledge-related activities in all the sectors in Australia into account, it may be noted that the total carbon footprint from photocopying content (books, journals, reports, and so on) is at least equal to that produced by 6–8 power plants of average capacity.

How can emissions be reduced?

Now, the question is, can this massive emission be reduced, and can we create a more environment-friendly system for knowledge creation, distribution, access, and reuse? In order to answer these two questions, one needs to identify the major contributors to CO_2 emissions in the prevailing supply chain for knowledge creation, distribution, access, and reuse. The two biggest contributing factors in producing CO_2 emissions in these fields are:

- The number of printed copies of content: books, journals, newspapers, conference volumes, and so on. As noted earlier in this paper, the number of books published in Germany and the USA contribute to 10.2 million and 42.3 million tonnes of CO_2, respectively
- Photocopying of content: both legally and illegally, which makes a significant additional contribution to the carbon footprint

How can we reduce these two activities and yet improve the process of knowledge creation and knowledge use that are the two cornerstones of a knowledge society? In the rest of this paper, we propose the architecture of a novel content supply chain that will help improve the economic and environmental sustainability of the content industry, and will thus ensure a more environment-friendly knowledge creation and knowledge access infrastructure.

Design of the new content service network (CSN)

Figure 1 shows the basic architecture of a new content service network, which will enable users to choose digital information from a myriad of information channels and sources, with sufficient levels of granularity, required for their use; and produce new information packages for future reference and distribution to a designated audience through a variety of media, including institutional or personal computers, handheld devices, on-demand print, and so on, within the framework of a new and easy-to-implement business model. The network model, once fully built and implemented, will provide all these and many more facilities, thereby creating a demand-driven, real-time content service model that will benefit content creators/producers, as well as consumers. Overall, the new content service network will play a key role in enabling people to make smart use of information.

The proposed CSN will have five major characteristics that will make it unique and among the currently available models in the content market place:

- It will not replace the current print-based or Internet-based model of content supply. Rather, it will be an additional channel, which will facilitate granular content discovery, access, aggregation, and super distribution.
- It will not build a repository of centrally held content resources. Instead, a single window is proposed for seamlessly accessing content directly from the databases where the content

is already held, such as the author or publisher databases.
- Neither does the system suggest that every piece of content used should be paid for, nor does it recommend that every piece of content be free. It builds an infrastructure that will allow the content creator/producer to decide whether or not the content should be free or paid-for, and the prices can be adjusted by them throughout the lifetime of each piece of content. It will be a content supermarket, where the consumers will be able to pick and choose the content—fee-based, as well as free, side by side—according to their preference.
- It will be a social infrastructure designed to create a level playing field for all types of content creators and producers to make their content available. Content creators and producers would register their content to be accessed through the CSN with a trusted third party, which will be a not-for-profit organization. The system will be built on a set of open standards and protocols, and its operation will be transparent.
- The CSN will include key players such as libraries and various other stakeholders, who add value to the content supply chain. It will create an environment where librarians and other stakeholders (such as market intelligence agencies) can add value in the supply chain, and play a key role in helping consumers get access to quality information through direct and transparent online transactions.

The proposed CSN will be significantly different from the currently available systems and approaches in many ways. It will:
- Create mechanisms for access to—and aggregation of—granular content, including copyright, out-of-copyright, and Creative Commons content currently available from heterogeneous information services, through a single window and a single process, in an on-demand and real-time mode.
- Allow the consumers to lawfully access, aggregate, store, and super distribute content.
- Have an underlying transparent business model that will facilitate all the legal rights management and commercial aspects of the transactions, and thus will hide from the consumers all the complexities of content access, rights management, and the supply chain.
- Create a win-win situation for both consumers and creators/publishers by generating a transparent mechanism for payment and/or attribution for access and super distribution of content.

It is proposed that the content service network will be developed and tested in a B2B transaction model. Once the technology—especially the relevant digital rights management tools and standards—is developed, it can be opened up for B2C transactions. It should be emphasized that the proposed network does not in any way suggest that consumers should pay for all kinds of content. Neither does it suggest that the payment has to be made separately for each transaction. On the contrary, it proposes a framework where the consumers will be free to choose different types of content—free as well as fee-based—in granular form, as required. Consumers will be able to add value, share, and create new knowledge as envisaged in digital library frameworks like the DELOS digital library reference model (DELOS 2006). Further research will help develop the precise systems for pricing and payment for content use and super distribution. Payments may be affected on an e-commerce transactional or subscription model. The other novelty of the network will be that it will provide a one-stop shop for all kinds of activities related to content discovery, access, aggregation, value-addition, use and super distribution, and so on. All these activities can be performed within a transparent legal and business model. The consumer will be able to

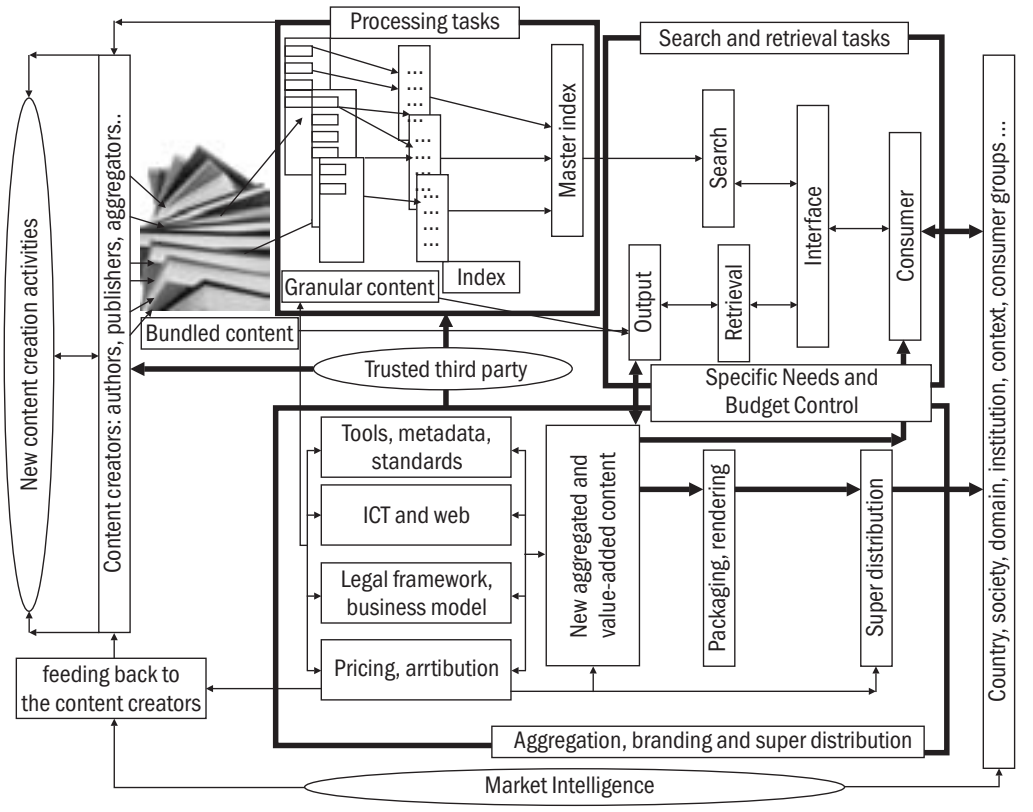

Figure 1 Outline of the content service network

monitor and adjust the overall cost for every transaction involving their selection of content.

CSN: a conceptual view

The new knowledge infrastructure will create an environment where bundled content (produced by publishers) will co-exist with granular content. It will be up to the consumers to decide which content to use under specific circumstances. For example, some consumers may wish to read an entire book, a complete journal article, a complete issue of a journal or even a complete volume of a journal; while others might wish to use specific sections of books, journal articles, and so on for a specific purpose. Consumers should be able to have on-demand access to the content marketplace and pay for what they use, rather than shell out a pre-determined fixed charge for each whole document, even if only a small portion of it is actually used in a given situation.

The new infrastructure will not only facilitate creation of new content packages by aggregating whole or granular content from a variety of channels and sources, but it will also allow consumers to add their own value in a number of ways. First, the aggregated content package will be created using a deep understanding of the context and consumers—using the knowledge of context. For example, in an academic environment, a lecturer accessing and aggregating content online for super distribution to students will apply the structure and strategy/policy of the university, faculty, and

department; the nature of the specific course; nature and characteristics of the students; and so on. Secondly, the users (in this case, the academicians, who are accessing the content for aggregation and super distribution) can add value by incorporating their own content in the form of notes, comments, interpretations, and so on. Once the content has been aggregated, the academy can then add its own branding, form, style, and so on. Therefore, the new product will not be just be aggregated content as such (as happens in the current environment of printed course packs or electronic course reserves), but will be designed and rendered relevant content packages aggregated from the widest range of resources, and targetted for a specific context and environment of downstream users.

The CSN will have the following characteristics:

- Digital content will remain with the content creator/provider, who will register their content and its granular segments with the CSN.
- Content in granular form will be tagged and indexed using appropriate tools, technologies, and metadata, in addition to its existing metadata. This calls for specific research to find out which currently available tools, technologies, and metadata can be used unchanged or with some modifications/ improvements to suit the specific needs of the new knowledge infrastructure.
- The CSN will have access to granular content, while the actual degree of granularity (chapters, sections or paragraphs, tables, figures, and so on) will be determined by market demand.
- Users will be able to search, locate, and access granular content. They will also be able to pick and mix specific granular content to suit their specific needs and the needs of their target audience (for example, a group of university students) with the help of appropriate tools, interface, and support systems. Again, specific research and evaluation studies will show, which of the currently available tools, technologies, and metadata can be used unaltered or with some modifications/ improvements to suit the specific needs of the new knowledge infrastructure.
- Users will be able to prepare aggregated content packages in real time, as per their specific needs.
- The aggregated content packages can be made available for interactive use with the help of various Web 2.0 technologies and tools.
- The aggregated content can be packaged/ rendered.
- The aggregated content can be saved for future use and/or distributed to a specified target audience.
- The rights for super distribution to the target audience will be managed with the content transaction.

All the above activities can be performed within an easy-to-use and transparent legal and business framework. Thus, the new CSN will establish many-to-many connections among the content creators and consumers. In other words, a given consumer will have access through one interface to a large pool of granular content produced by a variety of creators and publishers—large and small, local and overseas, copyrighted and copyright free. It will be a virtual marketplace, where content creators/providers and consumers will make transactions in real time to access or use content.

In the new CSN, content creators/providers/ publishers will have the ability to set the price, if any. However, it need not be fixed throughout the life cycle of the content. Instead, the price can be negotiable and adjustable depending on the frequency and volume of access, as well as on super distribution. Different prices may be set for different categories of customers or customers in different locations (even in different countries). There can be a differential pricing system in place that will benefit both the consumers and creators/providers of digital content.

It is anticipated that on a global scale this will lead to a multicurrency payment system as an element of the e-business model. All

the transactions will take place through a single window in real time, thereby relieving the consumers from dealing separately and differently with multiple content producers/providers and complex different copyright practices and laws.

Environmental benefits

One of the most significant indirect social benefits of the CSN will be the significant reductions in carbon footprint. Although further studies are necessary to gather data on the carbon footprint of the content industry in each country—and comparative data can only be produced once the CSN is fully developed and implemented—it can be argued that the new model will be able to reduce the carbon footprint figures significantly because:

- it will enable consumers to move increasingly towards use of information in digital form because of its convenience in access, downstream use, interactive use, novel sharing, and peer review facilities provided by the Web technology. In some cases, content will need to be printed, but it will be print-on-demand only, as opposed to the current book and journal industries' practice of printing, overprinting, sale or return, and pulping;
- there will be significant reductions in the physical movement of content for distribution in hard copies because content will only be printed as and when required. Consequently, there will occur reductions in return and pulping of unsold copies;
- less space will be required to store printed materials in libraries, bookstores, warehouses, and so on, where quite often special measures have to be taken for conservation of printed materials; and
- carbon footprint will be significantly reduced because within the new virtual supply chain model there will be less requirement for travel by people associated with the content industry.

Some critics argue that reading a book online produces more—or at least as much—carbon footprint as a printed book. Their studies calculate CO_2 emissions for the Internet, the computers involved, and the entire network infrastructure. However, this is not justifiable because every government, especially in the developed world, is investing huge resources towards building a broadband network. The network and its underlying infrastructure will operate whether or not it is used for online content transaction and use. In this regard, it might be mentioned that a comparison of CO_2 emissions from printed and digital books has been provided by Chowdhury (2010).

Summary and conclusion

The current model of content creation and distribution in print form is certainly damaging for our environment. Therefore, alternative and environment-friendly systems for content creation, access, and use should be developed to save the environment. The new content service network proposed in this paper will reduce carbon footprint by eliminating the level of CO_2 that is currently generated during the production and distribution of printed content. The proposed digital content service network will create a demand-driven content utility service system for institutional users that will allow them to search, find, access, aggregate, store, use, and super distribute granular content from a variety of free and fee-based content producers and services. It will also create a level playing field, thereby ensuring that small content creators and publishers have the same opportunity as others in making their products available through the virtual content marketplace. Thus, it is envisaged that the CSN will create a social infrastructure for content creation and use. The ultimate power, with regard to content creation and use, will rest with the general public (content creators and users), as opposed to a specific industry or business in the content supply chain. Although, further

research is necessary to develop standards and protocols for digital content creation and access through the network, the new model will support better content transaction and uses, thereby facilitating knowledge-intensive activities in a much more environment-friendly way.

References

Booktrade.info. 2010. **UK total consumer market trends for 2009**. Available online at: http://www.booktrade.info/index.php/showarticle/25285. (Last accessed on 27 September 2010).

Center for Global Development. 2007. Carbon dioxide emissions from power plants rated worldwide. In *Science Daily*. Available online at: http://www.sciencedaily.com/releases/2007/11/071114163448.htm. (Last accessed on 23 December 2009).

Chowdhury G. 2010. **Carbon footprint of the Knowledge sector: what's the future?** *Journal of Documentation* 66(6): 934–946

CAL (Copyright Agency Limited). 2009. **Annual report**. Copyright Agency Limited

DELOS. 2006. A Reference Model for Digital Library Management Systems. Available online at: http://www.delos.info/index.php?option=com_content&task=view&id=345&Itemid=#reference_model. (Last accessed on 23 December 2009).

Eco-Libris. 2007. **Book Publishing Industry**. Available online at: http://www.ecolibris.net/bookpublish.asp. (Last accessed 23 December 2009).

Space Daily. 2009. **German book industry faces green change**. Available online at: http://www.spacedaily.com/reports/German_book_industry_faces_green_change_999.html. (Last accessed on 27 September 2010).

Healey M. 2009. Book Industry TRENDS 2009 Indicates Publishers' Net Revenue Up 1.0% in 2008 to Reach $40.32 Billion. Book Industry Study Group. Available online at: http://www.bisg.org/news-5-363-book-industry-trends-2009-indicates-publishers-net-revenue-up-10-in-2008-to-reach-4032-billion.php. (Last accessed 23 December 2009).

HP Eco solutions. 2009. **HP Carbon Footprint Calculator.** Available online at: http://www.hp.com/large/ipg/ecological-printing-solutions/carbon-footprint-calc.html. (Last accessed on 23 December 2009).

Kozak G. 2003. Printed Scholarly Books and E-book Reading Devices: A Comparative Life Cycle Assessment of Two Book Options. Centre for Sustainable Systems, University of Michigan. Available online at: http://css.snre.umich.edu/css_doc/CSS03-04.pdf. (Last accessed on 23 December 2009). [Report No. CSS03-04].

Nielsen. 2009. **Nielsen BookScan.** Nielsen. Available online at: www.nielsenbookscan.com.au/uploads/fs_AustralianPanel_Sept09.pdf.

Price Waterhouse Coopers. 2008. **Making the tangible intangible: the economic contribution of Australia's copyright industries.** Price Waterhouse Coopers. [Prepared for the Australian Copyright Council.]

RISO. 2009. Power consumption: the hidden costs of copiers and printers. Available online at: https://us.riso.com/pls/portal/docs/PAGE/RISO07/CUSTOMERSUPPORT/INFORMATIONRESOURCES/WHITEPAPERS/POWER_CONSUMPTION_WHITE_PAPER.PDF. (Last accessed on 23 December 2009).

Ritch E. 2009. **The environmental impact of Amazon's Kindle: executive brief**. San Francisco. Cleantech Group.