Project Acronym: Greening Information Management
Version: V2
Contact: Diane.McDonald@strath.ac.uk
Date: 18/02/10

Project Document Cover Sheet

Project Information			
Project Acronym	Greening Information Management		
Project Title	Assessing the Role of Digital Information Management Practice in Reducing Environmental Impact of ICT Use in Higher Education		
Start Date	11/05/09	**End Date**	07/12/09
Lead Institution	University of Strathclyde		
Project Director	Dr Diane McDonald		
Project Manager & contact details	Dr Diane McDonald Head of Innovation Services & Projects, Learning Services University of Strathclyde 155 George Street, Glasgow G1 1RD Email: Diane.McDonald@strath.ac.uk		
Partner Institutions	None		
Project Web URL	http://www.greeningim.org.uk/		
Programme Name (and number)	JISC Preservation Programme		
Programme Manager	Neil Grindley		

Document Name			
Document Title	Final Report		
Reporting Period			
Author(s) & project role	Diane McDonald, Project Director/Manager; Emma McCulloch, Information Expert, Archie McDonald, Greening Expert		
Date	18/02/10	**Filename**	GIMFinalReport.docx
URL	*http://www.greeningim.org.uk/FinalReport/GIMFinalReport.docx (subject to JISC's approval)*		
Access	☐ Project and JISC internal	☒ General dissemination	

Document History		
Version	**Date**	**Comments**
V1	07/12/09	Original submission
V2	18/02/10	Revised submission as per NG's comments

Greening Information Management Final Report

Diane McDonald, Emma McCulloch, Archie MacDonald

University of Strathclyde

With additional input from Caroline Breslin and Jim Everett
University of Strathclyde

February 2010

JISC

University of Strathclyde Glasgow

Acknowledgements

The Greening Information Management Project was commissioned by JISC to assess the role of digital information management practice in reducing the environmental impact of ICT use in higher education.

The project team would like to thank the following individuals for helping with peer reviewing the various project outputs: Steve Bailey, JISC infoNet; Vicky Mays, University of Hull; Gareth Knight, King's College London; Caroline Breslin and Jim Everett, University of Strathclyde.

Sourcing case studies proved particularly difficult due to the novelty of the approach. The project team would therefore particularly like to thank the following individuals and institutions for agreeing to take part in case studies - Vicky Mays and Chris Awre, University of Hull; Alan Slevin and Alan Dawson, University of Strathclyde; Gareth Knight and Richard Palmer, CeRch, King's College London; Anne Grzybowski, Tom Armitage and EDINA Management Team, University of Edinburgh.

Finally, we would also like to thank the following organisations for participating in activity review interviews:

DCC: Digital Curation Centre

JISC Green ICT Programme and, in particular, the Programme Manager, Rob Bristow

JISC infoNet

JISC Legal

HE Institutions (University of Hull, University of Strathclyde, King's College London, University of Edinburgh,)

MIMAS

SCONUL: Society of College, National and University Libraries

SCURL: Scottish Confederation of University and Research Libraries

SLIC: Scottish Library and Information Council

UCISA: Universities and Colleges Information Systems Association

Executive Summary

As the recent JISC report on 'the 'greening' of ICT in education [1] highlights, the increasing reliance on ICT to underpin the business functions of higher education institutions has a heavy environmental impact, due mainly to the consumption of electricity to run computers and to cool data centres. While work is already under way to investigate how more energy efficient ICT can be introduced, to date there has been much less focus on the potential environmental benefits to be accrued from reducing the demand 'at source' through better data and information management. JISC thus commissioned the University of Strathclyde to undertake a study to gather evidence that establishes the efficacy of using information management options as components of Green ICT strategies within UK Higher Education environments, and to highlight existing practices which have the potential for wider replication.

The resultant Greening Information Management project assessed the role of digital information management practice in reducing the environmental impact of ICT use in higher education through a four phase approach: (i) a focussed review of existing literature and activity relating to Greening Information Management (GIM); (ii) development of a GIM Assessment Framework designed to be used by those responsible for digital resources to either assess the greenness of existing practices or to identify specific information management options that could be used to reduce the overall amount of digital resources stored within a given information environment; (iii) testing of the framework through a practical analysis of collections of digital resources held by four different institutions; and (iv) an analysis of the efficacy of using information management options as components of Green ICT strategies within UK HE environments, based on the findings from steps (i)-(iii).

The literature and activity review found little evidence of existing work regarding the use of information management to further the Green ICT agenda. That said, as information management is concerned, amongst other things, with destroying unused data/information and rationalising storage, the review did support the potential for utilising information management options to reduce disk storage requirements and hence data centre energy consumption. The review suggested that a range of automated and manual techniques when used in combination with Information Lifecycle Management (ILM) and stewardship strategies offer the potential to reduce digital resources. The most relevant techniques (GIM options) are: Enterprise Content Management (ECM); de-duplication; manual pruning/weeding; enterprise-level, centralised data storage repositories; tiered data storage systems; Master Data Management (MDM); and resource limitation.

Reduction in disk usage, however, is not the primary objective of information management and therefore a positive environmental benefit should not be assumed. For example, retention is viewed as a positive practice in many types of information environment. Further, achieving buy-in to change working practices to focus on reduction of digital resources is not an easy task. Organisational practices, business objectives and culture are all likely to affect decisions regarding the retention of digital information. Integrated institutional strategies which advocate good business practice within a context of overall environmental stewardship and responsibility should aid the reduction of stored digital resources. However, considerable change management activities are likely to be required if green information management options are to be implemented effectively.

The GIM Assessment Framework which was developed from the review findings provides a tool that allows information managers to explore and make an informed decision regarding how to use information management options to intelligently decrease the amount of digital resources – data or information – that needs to be stored and managed. The framework consists of 3 stages. *Stage 1: Baselining the current Information Environment* is concerned with capturing the current information environment including lifecycle and stewardship requirements. *Stage 2: Selecting GIM Options* is concerned with assessing which of the range of GIM options offer the potential to reduce the storage requirements of digital resources within the information environment under consideration. Finally, *Stage 3: Assessing GIM Efficacy* is concerned with exploring the business case for the selected GIM option – i.e. whether it will deliver energy savings and positive business benefits in a cost effective manner. Stage 3 can also be applied retrospectively to identify the green costs and benefits associated with an existing information management approach.

The Framework was applied in four case studies covering a range of different information environments – central institutional shared drives (University of Hull), an institutional repository (University of Strathclyde), archive collections held within a research centre (CeRch, King's College London) and a fairly broad institutional perspective incorporating records management policies and experiences of a data centre (University of Edinburgh). As well as helping refine the Framework, this enabled data to be gathered to help assess the real-world potential for implementing more energy efficient data management measures. In particular, the Framework helped identify particular GIM options that could potentially be used to reduced digital storage and hence the carbon footprint within the different institutions and assessed their efficacy.

Analysis of the efficacy of GIM based on the case study findings and the review supports the thesis that implementation of GIM options within an overarching ILM and stewardship strategy could contribute to a positive environmental impact, provided they are incorporated into institutional policy and endorsed by senior management. However, the innovative nature of the approach being investigated – i.e. seeking to reduce the environmental impact of ICT through a reduction in digital resources – meant that the study needed to adopt a qualitative approach which explored possible implementation scenarios only. Further quantitative evidence from real implementations is required. Without such investigation it is difficult to predict whether options will bring significant benefits relating to environmental impact or whether likely benefits may be countered by unexpected costs. For example, the environmental benefits arising from the implementation of such options may result in costs in other areas (e.g. staff costs, training needs). For GIM to be effective, the full extent of costs require to be examined in detail and weighed up and institutional information management policies need to be aligned with the green agenda.

In conclusion, while this innovative approach to information management further positions JISC at the forefront of the Green ICT agenda, further work is required before the sector will be able to capitalise on its potential. If this is to be achieved, it is recommended that:

1) Quantitative action research be undertaken to assess the potential capacity for reducing disk space and energy consumption through the GIM options identified in the project.
2) Evidence-based good practice advice relating to GIM be developed.

3) JISC work with other relevant organisations to arrive at holistic positions which address the Green agenda and ensure effective stewardship of resources.

4) The GIM Assessment Framework be used in combination with the JISC infoNet Impact Calculator to create a practical and useful tool to assist information managers in making greener decisions and formulating business cases be assessed

All project outputs including the literature and activity review, GIM Assessment Framework and case studies are available via the project website at http://www.greeningim.org.uk/.

Table of Contents

1 Background and Overview

1.1 Background

While information and communication technologies (ICT) and digital information or data are increasingly core to the activities of Higher Education (HE), as the recent JISC report on '*The 'greening' of ICT in education'* [1] highlights, the energy usage associated with the data centres and other storage facilities makes a significant contribution to the overall carbon footprint of ICT, and one which is continuing to increase. Given that the government and funding councils are setting challenging carbon reduction targets for the HE and FE (Further Education) sectors, it is imperative to understand how environmental impact might be reduced at source to help meet these targets. The JISC Green ICT Programme [3] is thus examining how to address such issues through minimising the environmental impact of ICT.

Over the last 40 years ICT have increasingly underpinned education, research, knowledge exchange and administrative activities within HE and FE institutions. This ICT usage continually generates and reuses digital resources in many forms. For example, every time an email is sent a new digital message is created and potentially stored by the sender, receiver and email provider. If these digital resources are not properly managed and deleted when no longer required, it will be impossible to efficiently retrieve information as and when required. Further, if there is no strategy for deleting information when it is no longer of use, the volume of digital resources will continue to increase unabated, demanding ever larger data centres and other storage devices, all of which significantly contribute to the environmental impact of ICT. HE institutions are thus awash with digital resources and the size of these assets directly impacts upon data storage requirements.

This suggests that how we manage, use and store digital resources is of significant interest in the quest to reduce the negative environmental impact of ICT. The Green Data Project [4] in the US supports this premise, "assert[ing] that technology greening is inextricably linked to sound data management" and stresses that "Green IT begins with Green Data". However, as '*The 'greening' of ICT in education'* report recognises, not all proposed methods of reducing the environmental impact of ICT do in fact deliver the expected benefits. Of particular concern is whether attempts to reduce the amount of digital resources stored will interfere with good stewardship of the resources and the effectiveness of the education, research, knowledge exchange or administrative business functions which the resources underpin. Thus, JISC commissioned a 6 month study by the University of Strathclyde to assess the efficacy of using data and information management options as components of Green ICT strategies within UK HE environments.

1.2 Aims and objectives of the study

The aim of this study was to provide evidence that establishes the efficacy of using information management options as components of Green ICT strategies within UK Higher Education environments, and to highlight existing practices which have the potential for wider replication. Specific objectives were:

- To investigate and summarise existing literature and activity in the 'green' information management field;
- To identify and articulate the various ways that those responsible for managing information within UK HE institutions could or might use technological means to

intelligently decrease the amount of digital information that it is necessary for them to manage, whilst still demonstrating effective stewardship of all required information;

- To undertake a practical analysis of collections of information or data held by *at least* four different institutions to determine the real-world potential for implementing more energy efficient data management measures, especially information lifecycle management (ILM);
- To produce a report detailing the results of the above analysis and drawing conclusions about the likelihood of these measures to make an effective sector-wide contribution to the Green ICT agenda;
- To provide materials which will encourage individual institutions to take greater action to achieve more energy efficient data management;
- To produce recommendations to JISC and other bodies for further related work in this area.

1.3 The approach adopted

The study was undertaken in 4 phases. Existing research projects such as SusteIT [5], focus on the technological aspects of energy saving practice within institutions. The GIM project was intended to be less hardware focussed, considering instead options of greening information management within various systems employed. The first phase was to conduct a focussed review of existing literature and activity relating to Greening Information Management (GIM) [6]. As GIM is a relatively new concept, the review sought to identify pertinent current practice that might inform future GIM strategies and policies rather than concentrating solely on actual green information management activities, as these are very limited. The second phase of the study involved development of a GIM Assessment Framework [2] designed to be used by those responsible for digital resources to either assess the greenness of existing practices or to identify specific information management options that could be used to reduce the overall amount of digital resources stored within a given information environment. A framework approach was adopted because given the range of information types held by institutions, from email, research data, management information, institutional records, e-learning content and personal file stores etc, and their very different stewardship requirements, neither a universal GIM approach nor a prescriptive assessment approach are possible. The Framework was developed through synthesis of the literature and activity review findings and was tested and further refined through application in real-world contexts and peer review.

This testing was carried out by undertaking a practical analysis of collections of information or data held by four different institutions and constituted phase 3 of the study. The four case studies of a range of different information environments – central institutional shared drives (University of Hull), an institutional repository (University of Strathclyde), archive collections held within a research centre (CeRch, King's College London) and a fairly broad institutional perspective incorporating records management policies and experiences of a data centre (University of Edinburgh). As well as helping refine the Framework, this enabled data to be gathered to help assess the real-world potential for implementing more energy efficient data management measures. The final phase of the study was to use the findings from the literature and activity review and from the real-world case studies to analyse the efficacy of using information management options as components of Green ICT strategies within UK

HE environments. This was achieved by undertaking an impact analysis which examined the feasibility of implementation, direct and indirect costs, benefits and implications of implementing GIM options.

1.4 Scope and contents of the report

This report presents the findings of the Greening Information Management Project [7] to JISC and the HE sector. In doing so, it draws heavily on two related outputs – *Greening Information Management: a focussed literature and activity review* [6] and the *Greening Information Management Assessment Framework* [2]. These reports and the case studies can be found in full at the project website: http://www.greeningim.org.uk/.

The report proceeds as follows. In section 2, how information management might be utilised as part of a 'green' ICT agenda is discussed. Next, in section 3, an overview of the Greening Information Management Assessment Framework is presented and its application discussed. Summaries of the four case studies which were developed using the GIM Assessment Framework are then provided in section 4. The efficacy of using information management options as components of Green ICT strategies within UK HE environments is then assessed in section 5. The report ends, in section 6 by summarising the conclusions drawn from the study, outlining the implications of the findings and highlighting a series of recommendations designed to improve the potential to leverage information management as part of the 'green' ICT agenda.

2 Information Management and Green ICT

As the Green Data Project [4] in the US asserts, “storage technology, mainly in the form of large disk arrays, is becoming the biggest consumer of energy in contemporary data centers”. However, at present, as much as 60 percent of the capacity of every disk drive that is used to store digital business information contains stale, duplicated, contraband or otherwise useless data. For example, as a series of interviews with relevant stakeholders in HE [6] highlights, we do not tend to take stock of what it is actually necessary to retain and what might be reasonably disposed of. Thus, as the Green Data Project argues, “collective failure to apply *data discipline* to our business information system storage – to purge junk data and to archive data with little chance of re-reference onto greener archival media – is what drives the acquisition of more and more energy consuming hardware year after year, increasing exponentially the carbon footprint of IT in the process.”[4] Better management of digital resources is required.

Information and data management strategies are already being introduced to ensure that the digital resources maintained by an organisation efficiently and effectively meet its business needs and associated legal and regulatory requirements. If information management is to be employed as part of a Green ICT strategy then it should ultimately result in less computing power through the reduction of storage space, a key factor in progressing the wider Green ICT agenda. While, environmental impact has to date not been a feature of information management strategies, as IBM note, information lifecycle management (ILM) is usually implemented “in a manner that optimizes storage and access at the lowest cost”. [8]

The literature and activity review identified two key and complementary types of approach to information and data management – strategic and holistic lifecycle approaches and practical technology-driven approaches. These two types of approach, their relationship and their applicability to the GIM agenda are briefly summarised below. This is followed by an overview of barriers to and potential enablers of GIM. The section ends with a reflection on the current state of play and the literature and activity review process.

2.1 Digital resource management approaches

The format of the digital resources held by institutions varies greatly, deriving in the main from the particular application in use. Typical formats include: emails, database entries, repository records, application data, static and dynamic web content, audio and video as well as raw research data. These formats vary greatly in terms of the structure of the information they contain, whether it is dynamic or static and the extent of associated metadata. Further, the applications that generate the digital resources also generate further hidden data pertaining to the information or data – e.g. usage logs, metadata etc.

2.1.1 Strategic and holistic lifecycle approaches

Information and data lifecycle management and stewardship

Information lifecycle management (ILM) comprises a set of strategies for successfully managing information throughout its period of existence. There are a number of different approaches: e.g. the JISC infoNet infoKit on managing the Information Lifecycle [9] from creation, through active use, semi-active use and final outcome (disposal or preservation); the Curation Lifecycle Model (CLM) [10] devised by the DCC (Digital Curation Centre) [11];

an activity lifecycle of digital assets [12]; and IBM's approach which is described as "a process for managing information through its lifecycle, from conception until disposal, in a manner that optimizes storage and access at the lowest cost" [8].

The inclusion of disposal considerations and the focus on optimising storage, at least in the IBM model, suggest that applying lifecycle management techniques to digital resources could result in an overall reduction in digital resource storage.

The activity review indicated varying degrees of ILM adoption in the community. Data centres tend to have adopted more formal ILM methods, using more formal tools to assess their digital assets and liabilities[13]. Although not always overtly acknowledged, institutional records management policies are often based on the underlying principles of ILM with records management, having been defined by one institution as "a discipline which utilises an administrative system to direct and control the creation, version control, distribution, filing, retention, storage and disposal of records, in a way that is administratively and legally sound, whilst at the same time serving the operational needs of the University and preserving an adequate historical record." [13]. Relevant ILM assessment tools include: the Data Audit Framework (DAF) [14], which provides organisations with a methodology "to identify, locate, describe and assess how they are managing their research data assets" [15]; and the DRAMBORA (Digital Repository Audit Method Based On Risk Assessment) toolkit "intended to facilitate internal audit by providing repository administrators with a means to assess their capabilities, identify their weaknesses, and recognise their strengths" [16].

Stewardship

Effective stewardship – "data-quality management, data security, auditable compliance with privacy and disclosure guidelines, ILM, and business-continuity planning and disaster recovery" [13] – is key to successful information management. This will be particularly critical in attempts to implement GIM options as it is imperative that the drive to reduce overall digital resource storage does not compromise the business functions of an institution.

Within the HE sector, the RIN (Research Information Network) has produced '*Stewardship of digital research data: a framework of principles and guidelines*' [17]. This focuses on the roles and responsibilities of various stakeholders; development to international standards; ease of retrieval and use of validated sources; efficient and cost-effective use and management; and development of long term valuable research resources. However, stewardship requirements depend on the type of digital resources and business requirements, and so the RIN approach is not directly transferable to, for example, administrative functions within institutions. Marks [18] provides a useful, more general high-level approach to stewardship based on: a holistic approach; senior level-sponsorship; alignment with business processes; extensive change management support and a thorough understanding of information ownership.

The activity review carried out as part of this study indicated that while institutions were highly concerned with stewardship of information, especially relating to legal or statutory requirements, this was often implemented at a departmental or individual business process level and not as part of a holistic institution-wide approach to the stewardship of an institution's information in its entirety.

2.1.2 Technology that supports green information management

The literature and activity review identified a range of technical options that can be applied to further an institution's efficiency with a view to reducing its digital information storage. The options listed below can be tied to individual stages within ILM models, and could be considered within the process of applying an ILM model to an organisation's information.

Enterprise Content Management (ECM) software can constitute a whole system used at the corporate level or a set of different tools that can be used at a corporate and/or at a more devolved level. ECM includes the software, and associated processes. It should help in the implementation of an ILM policy that covers retention of information and associated processes. It would also include the effective use of metadata (assign, use, capture). Many vendors have ECM products including: IBM, Oracle, HP, Symantec, FileNet and Alfresco (open source).

De-duplication is a software solution to minimise the amount of data stored. Duplicate files and file-segments are replaced with pointers to a single original. Providers of de-duplication software include ExaGrid, NEC, IBM, Symantec and various open source products are also available.

Manual pruning/weeding is another method by which de-duplication might be achieved either in conjunction with a retention schedule or on a more ad-hoc basis. Digital records that no longer have value to the organisation can also be removed by manual pruning/weeding. This needs to be underpinned by a retention schedule nd institutional policies which need to be disseminated, training provided and enforced.

Enterprise-level, centralised data storage repositories with network and/or web access can minimise the need for local copies of information/data to be kept. May also incorporate collaborative workspaces, for example, using SharePoint.

Tiered Data storage systems is a technology solution where active, regularly used data is kept on high performance servers, less active data is held on lower performance servers and unused data is stored on removable or near-line storage media. Systems are likely to use products from a number of vendors but may have a software product to integrate all the hardware and software, e.g. Hitachi.

Master Data Management (MDM) software aids the identification of different versions of documents – this can be a time-consuming data reconciliation process. Closely related version control software is used to store and keep track of the different versions of file/documents. Version control has been widely used by the IT industry for many years and is now seen in many other products such as content management systems (CMS) (e.g. SharePoint). MDM software is more recent but is used in many sectors of business with products from many vendors (e.g. IBM, Oracle). JISC infoNet has created a guide that addresses some of the issues associated with using version control [19].

Resource limitation using quota systems to limit the amount of disk space or space on an email server, has been used for many years. This option forces users to organise their digital resources and not keep what they do not need. This has fallen out of favour in recent years, presumably largely because disk space has become much cheaper and portable storage media more pervasive.

The diagram below (Figure 2) indicates how each of the above options relates the ILM model. Options are mapped to various stages of the information lifecycle, to illustrate the key points at which options might be applicable within a given information environment.

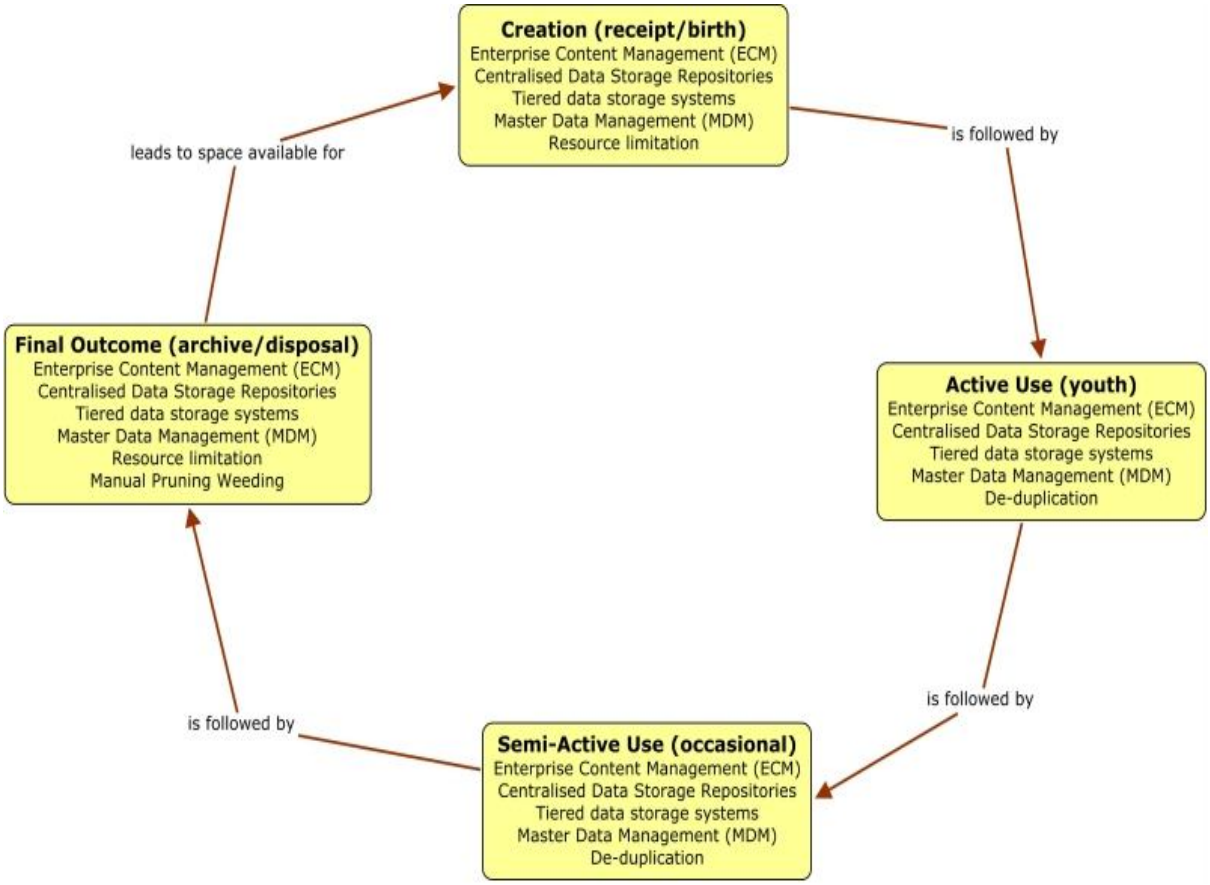


Figure 2 – Relationship between ILM and greening options

2.2 Potential barriers and enablers of green information management

The literature and activity review revealed a number of barriers to the reduction of digital resource storage as well as potential enablers which may help increase the greenness of information management.

2.2.1 Barriers to green information management

Perceived quality of service and data permissions problems lead to people circumventing centralised information resource strategies:

The JISC Work-with-IT *Knowledge Exchange Across Rural Borders* case study [20], which examines technology-driven changes to working practices, is a prime example. In this case study, duplicate copies of knowledge bases have evolved due to local support and access problems. Not only does such duplication have a negative environmental impact, but it also poses issues for version control and data validity.

Perceived conflict between business requirements and green information management is particularly problematic. Current ideas of 'best practice' may need to evolve.

As the management team of a JISC data centre highlighted, replication is seen as a positive practice for effective data management but negative in terms of saving energy. Thus, the Green agenda was not viewed as a strong motivator for data centres.

Technology limitations may also impede reduction in data storage. In particular, current software applications may not be compatible with green information management practice. Such issues should be eradicated as new software releases begin to include full support of document sharing and collaborative working features. However, bespoke software may prove much more problematic.

While the Greening Information Management project at the University of Strathclyde is employing a central shared document management system to store information and facilitate collaboration, lack of a collaborative mode within the referencing software being used means that we have to hold multiple duplicate copies.

Statutory and regulatory requirements such as the FOI (Freedom of Information) Act 2000 [21] and the DPA (Data Protection Act) 1998 [22] have a direct effect on the information stewardship requirements of institutions [23]. Where a holistic approach to sound information management and stewardship is not adopted this can lead to increased data storage and retention commitments.

FOI legislation demands that qualifying information be identified, available, retrieved and disseminated within a given time period (normally 20 working days). This in turn requires that institutions effectively manage their information to ensure compliance. However, the requirement to retain information may lead to multiple forms of storage being adopted in order to attempt to avoid potential archive retrieval problems associated with media deterioration and technology obsolescence [24].

Culture, working practices and lack of trust can limit the greenness and arguably also the effectiveness of information management options. Further, it is not just habit which is giving rise to 'un-green' practice but also a concern regarding the ongoing stewardship by the institution of an individual's digital resources.

Akin to the tendency to still print out paper copies of emails which arises from "the cultural or lifestyle habits of workers" [25], anecdotal evidence from the activity review process

illustrates that individuals still retain their own local copies of emails, files and research publications simply through habit or 'the way they work'.

A perceived increased cost associated with GIM is also acting as a barrier.

A general belief was expressed that short-term costs would increase to e.g. provide additional staff to implement GIM and to deliver associated training. It was felt however, that in the longer-term, the implementation of GIM would result in cost savings, thus making any short-term financial investment justifiable.

Finally, **a lack of coherent strategy on green digital resources management** means that, at best, mixed or contradictory messages are being sent and, at worst, there is no thought to the environmental impact of our use of digital resources.

The EDINA Management Team highlighted conflict between greener information management practices and current practice in some areas. For example, for organisations whose remit is to provide information services, replication of key data is viewed as a positive practice for effective data management and service provision but negative in terms of saving energy.

2.2.2 Potential enablers of greener information management

Link with mandatory requirements

At a recent JISC workshop on the Green agenda attended by key stakeholders from government and business, and carbon emissions experts, it was forecast that there may increasingly be a move towards pseudo green taxes and accountability through organisations' annual reports. Such moves are highly likely to focus organisations' attention on their green credentials.

Competitive advantage: In the increasingly competitive market of HE, student opinion may also be highly influential.

A recent survey of students as part of the Creating Academic Learning Futures (CALF) research project indicated that an institution's green credentials were a key differentiator in students' choice of institution [26].

Align the GIM case with the business benefits that such approaches will also deliver.

Eric Olson provides a methodology and a tool set that can be used to assess an enterprise and formulate a green strategy that is underpinned by 3 principles [27]

1. A green strategy fosters a common culture of awareness and action

2. A green strategy facilitates decisions and transformation initiatives that improve the environment

3. Green strategies have attractive value propositions that are cost effective

Organisational leadership: The SusteIT project's recommendation for "Clear organisational commitments, and effective implementation processes such as greater responsibility for energy consumption" [1] was further supported by the literature and activity review. Synthesis of the review findings suggests that this leadership should include:

- Institution-wide holistic green strategies which are incorporated into all institutional activities
- Holistic information management and stewardship strategies
- Strategic leadership and direction

Effective change management: is key to any changes to technology-enhanced working practices [28]. Again, synthesis of the review findings suggests that this leadership should include development of:

- Change skills of individuals
- Personal awareness of green issues and positive contribution that can be made by individuals
- Good practice advice and exemplars relevant to specific contexts

Information literacy: Extending traditional information literacy skills and ICT skills is central to the success of effective information management, typically underpinned by information lifecycle management models. To be able to asses information and its value in the context of an organisation requires traditional information literacy skills. One definition of this, from CILIP (Chartered Institute of Library and Information Professionals), defines it as: "Information literacy is knowing when and why you need information, where to find it, and how to evaluate, use and communicate it in an ethical manner" [29]. There is also little point in investing in new technology-based tools to store information more efficiently if no one can use these tools. There is therefore a requirement to have adequate levels of information and computer literacy and this may result in additional training requirements. The greening case studies and literature review confirmed this, and increased training requirements were cited as a likely additional cost to the implementation of greening options.

2.3 Reflections on information management and the Green ICT agenda

Prior to this study, there was no assessment of the feasibility or efficacy of employing information management options as part of a Green ICT agenda. However, some institutions are undertaking significant information management activities although they do not appear to be considering them from a green perspective. For example, the recent JISC TechWatch report on Enterprise Architecture illustrates how three Universities – Cardiff, Liverpool John Moores and King's College London - are "describing and aligning the functional aspects of an organisation: its people, activities, tools, resources and data/information, so that they work more effectively together to achieve the organisation's business goals." [30]. Other institutions are also actively pursuing an Information Management agenda – for example, at the University of Hull, a new 'Information Management' section has been set up to drive forward better ways of enabling and supporting information management within the institution [31]. However, most of the institutions and sector services contacted felt that they had little to contribute to the activity review undertaken as part of this project as their activities did not consider environmental impact, although the majority expressed a keen interest in the outputs of this study.

That said, as information management is concerned, amongst other things, with destroying unused data/information and rationalising storage, the literature supports the potential for utilising information management options to reduce disk storage requirements and hence also data centre cooling requirements and thus has the potential to reduce the overall carbon emissions that result from ICT usage. Reduction in disk usage, however, is not the primary objective of information management and therefore a positive environmental benefit should not be assumed. Data and information management may be viewed as a combination of managing the lifecycle of an organisation's information via appropriate usage, retention, storage and discard strategies. These strategies should be governed by an appropriate information stewardship policy and can be implemented through technological solutions or working practices. There are various tools which can be used by an organisation to assess different information management requirements. Further, debate exists regarding the most appropriate options for different types of information. A holistic perspective is required, however, if information management is to deliver business and potentially green benefits.

Organisational practices, business objectives and culture are all likely to affect decisions regarding the retention of digital information. Integrated institutional strategies which advocate good business practice within a context of overall environmental stewardship and responsibility should aid the reduction of stored digital resources. However, considerable change management activities are likely to be required if green information management options are to be implemented effectively.

In conclusion, the literature and activity review indicate that there is a reasonable case to be made for employing information management as part of an overall Green ICT strategy. The lack of direct evidence, however, indicates that institutions will need to be provided with guidance on how to assess whether green information management is feasible in particular contexts and whether it is viable in terms of its impact of business and the environment.

Full details of the Greening Information Management project literature and activity review and the methodology used can be found in the accompanying report [6].

3 A Framework to Assess the Potential for Greening Information Management

The GIM Assessment Framework developed by the project team provides a tool that allows information managers to explore and make an informed decision regarding how to use information management options to intelligently decrease the amount of information that needs to be stored and managed. Such a reduction in digital resources stands to reap a number of benefits for institutions, departments and individuals, provided effective stewardship of the information is maintained. These include: environmental benefits, financial savings, more efficient access to digital resources, more consistent and accurate retrieval of information, better compliance with legislation (resulting from efficient retrieval of FOI requests, for example) and an enhanced reputation.

The GIM Assessment Framework focuses on maximising the potential environmental benefits while retaining effective stewardship because as the JISC-funded SusteIT project [1] into Green ICT recognises, the energy usage associated with the data centres and other storage facilities is a significant contribution to the overall carbon footprint of ICT, and one which is continuing to increase. The Framework is based on the candidate greening information management options identified from the literature and activity review [6] undertaken as part of the study, and has been tested and further refined through case studies [32] and peer review.

The Framework is designed to be used by those responsible for digital resources to either assess existing information management practices or in the development of new information environment and practices. Relevant information systems may range from institutional repositories, management information systems through websites and e-Learning and research resources to personal email and file storage. Given the range of information types and their very different stewardship requirements, neither a universal green information management approach nor a prescriptive assessment approach are possible; rather, the GIM Assessment Framework specifies the three stages, captured in Figure 1, p21, that are undertaken as part of an assessment of the potential for greener information management.

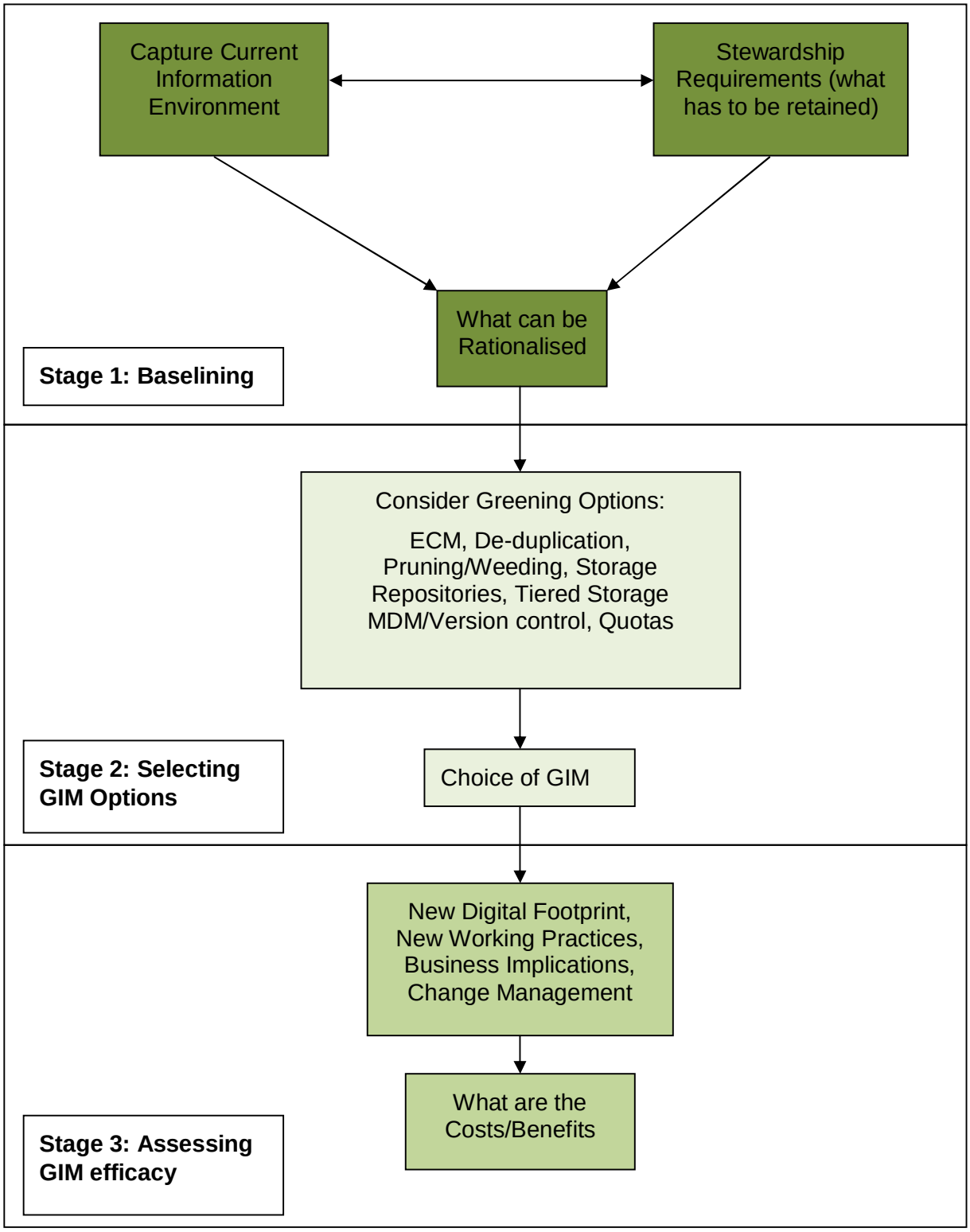


Figure 1 – GIM Assessment Framework Diagram

Box 1, below, summaries the three stages of the GIM Assessment Framework: Baselining the information environment; Selecting GIM options and Assessing their greening information management efficacy.

Box 1: The GIM Assessment Framework

Stage 1: Baselining the current Information Environment is concerned with capturing the current information environment and its stewardship requirements.

Stage 2: Selecting GIM Options is concerned with assessing which of a range of GIM options – techniques and technologies – offer the potential to reduce the storage requirements of digital resources within the information environment under consideration.

Stage 3: Assessing GIM Efficacy is concerned with exploring the business case for the selected GIM option – i.e. whether it will deliver energy savings and positive business benefits in a cost effective manner. Stage 3 can also be applied retrospectively to identify the green costs and benefits associated with an existing information management approach.

3.1 Stage 1: Baselining the information environment

In order to be able to decide whether digital storage requirements can be reduced by applying GIM options – the techniques or technologies summarised in section 2 – it is first necessary to establish a baseline of the information environment under consideration. Thus, stage 1 of the GIM Assessment Framework is concerned with baselining the information environment by capturing the current information environment and assessing stewardship requirements.

There are a number of tools available which can be used to help baseline the information environment under consideration; however no one tool completely captures the current information environment in the detail required. Possible tools and an analysis of their usefulness in differing contexts is presented in section 3 of the accompanying report on the GIM Assessment Framework [2].

3.2 Stage 2: Selecting greening information management options

There are a wide range of existing and emerging information management options which could potentially be employed to reduce the amount of digital resources that are held within institutions. For example, options include Enterprise Content Management (ECM), de-duplication, Master Data Management (MDM) and version control. The use of technologies and storage systems such as data storage repositories, tiered data storage systems and resource limitation is also relevant to the potential for greener information storage. These options all bring differing potential advantages and disadvantages when seeking to reduce digital resources storage and hence carbon footprints. Selecting which potential GIM options are relevant within a given information environment is therefore, as Figure 1, p21 captures, the subject of stage 2 of the GIM Assessment Framework.

In the accompanying report on the GIM Assessment Framework [2], stage 2 is described in section 4. First, the broad selection criteria for assessing GIM options – business implications, working practices implications, technology footprint implications and cost of implementation and support – are discussed. Then a range of information management options that may help to reduce energy consumption whilst still maintaining effective stewardship of information are described, an outline of likely implications for their feasibility based on the GIM selection criteria provided, and their applicability to differing types of digital resource summarised. These implications are indicative only and their relevance needs to be explored in relation to the specific information environment under investigation.

3.3 Stage 3: Assessing greening information management efficacy

The business case for proposed GIM options – whether they will actually reduce energy consumption while still ensuring good stewardship of information and fulfilling an institution's business needs without introducing unreasonable financial costs – requires to be made. As Figure 1, p21 captures, stage 3 of the GIM Assessment Framework is concerned with assessing the efficacy of selected GIM options or indeed the greenness of information management strategies within an institution. Such an assessment of efficacy could equally be applied retrospectively to existing information management strategies.

Details of how assessment of the efficacy of GIM options might be achieved is the subject of section 5 of the accompanying report on the GIM Assessment Framework [2]. This involves examination of the criteria which should be considered, including: greening (reduction of the amount of digital storage), business implications, costs and legal issues. Advice is also provided on identifying costs, benefits and issues brought about by introducing GIM options. This includes[1] reference to some key assessment tools such as INSIGHT [34], the Benefits of ICT Investment Landscape Study's (BIILS) Evaluation Framework and Toolkit [35] and the SusteIT ICT Energy and Carbon Footprinting Tool [36]. The different categories that may result in costs or savings, the barriers to success, changes to working practices, business implications and how change is managed are also discussed.

3.4 Using the GIM Assessment Framework

The GIM Assessment Framework is designed to guide information managers in selecting appropriate GIM options for use in their particular information environment. It is extensive in nature, covering a wide range of GIM options, issues and baselining and assessment techniques. While this approach allows information managers to select tools with which they are already familiar and contextualise to their own institution and information environment, this may be a very intensive activity. However, the advantage of the framework approach is that it can be applied exhaustively or with a light touch, using only selected key elements from the various baselining and assessment techniques referenced within the full Framework.

As part of the Greening Information Management project, four case studies were undertaken to explore the real-world potential for implementing GIM. The case studies used a range of guideline questions synthesised from the various baselining and assessment techniques on the GIM Assessment Framework. This easy to use, light touch approach, which is presented in the report on GIM Assessment Framework [2], is reproduced below.

In stage 1, the following questions were issued to participants in order to encourage them to capture the current information environment and to assess stewardship requirements.

[1] The newly released JISC InfoNet Records and Information Management Impact Calculator [33] was not available for consideration when the GIM Assessment Framework was developed.

Stage 1: Baselining the current Information Environment

1. What digital resources do you have within your organisation/service at the moment? (Pick one specific area if your resources are very broad in scope.)
2. What stewardship requirements do you have?
3. Why are you retaining information resources not subject to legal requirements?
4. Have you used any tools like DAF or DRAMBORA to assess your information environment?
5. How are resources currently managed?
    a. Individual/group responsibility?
    b. Audits undertaken?
6. Are you able to identify resources/files that could be rationalised in some way?
7. Does your institution's technical infrastructure influence how your information is stored in any way?
8. Do you follow any specific guides to best practice relating to information management?

From the responses to the stage 1 questions, indications of what resources could be rationalised emerged and were discussed with participants.

In stage 2, a number (approximately 3) of the GIM options were discussed and the following questions were asked to assess the effect, suitability and effectiveness of each of the options within the specific information environment being considered. This was facilitated using the following questions.

Stage 2: Selecting GIM Options

1. For each of the GIM options, would this technique or technology result in benefits to your information service/collection?
2. Are there any additional organisational benefits this GIM option might bring?
3. Do you think this would also result in reduced digital storage?
4. Which GIM options appear inappropriate / infeasible? Why?
5. Will information stewardship requirements still be met if the GIM option is implemented?
6. Will the GIM option enable more effective/efficient compliance (FOI etc)?
7. Will the method or combination of methods be applicable to all the different types if information/data in the organisation?

This led to the identification of 1-3 GIM options which had the potential to reduce the storage of digital resources in the information environment under consideration.

In stage 3, the option most appropriate to the case study's institution/situation was discussed. Questions in the box below were asked to identify possible costs/savings and benefits/issues of the chosen technique.

Stage 3: Assessing GIM Efficacy

1. When implementing the chosen option identified as potentially useful in Phase 2, can you identify any changes to:

    a. Working practices

    b. Business implications

    c. Digital footprint

    d. Change management

    Do you have anything further to add on implementing the chosen technique?

2. Do you foresee costs in any of the following categories related to new investment?

    - Staff costs (including change management, additional/less, role changes)
    - Compliance costs
    - Development costs
    - Training Costs
    - Infrastructure costs
    - System support
    - Software costs (commercial, in-house, open source)

    Costs will be relative – e.g. an extra person at "x" staff grade, cost of a new product, cost of a new server. There could also be potential cost reductions (savings).

3. Can you identify benefits or issues relating to the choice of GIM option? Possible categories: Greening, Financial, Educational/research, Organisational/ administrative, Compliance, External. The benefits may affect different stakeholders - e.g. academics, professional services, admin staff, students, etc. can you comment on this?

4. Questions on key issues in relation to costs/benefits:

    a. What are the barriers to success?

    b. How can you/will you measure/evaluate the level of success of the GIM option and who will be responsible for this?

    c. Who will follow this up and ensure any issues are addressed?

    d. How will you address scenarios where benefits are perceived differently by different stakeholders? e.g. the GIM option will save costs by improving efficiency for the institution but administrative staff will have an increased workload.

Scale of Investment (scale of project) The following key issues relating to determining how to evaluate costs/benefits were used as prompts if necessary.

- Type of Benefit
- Scope of Benefit (how wide and who benefits)
- Complexity (qualitative methods more complex)
- Alignment (aligned with corporate objectives)
- Availability of Comparison Data (e.g. Benchmarking requires)
- Type of Decision (appraisal, monitoring)
- Institutional Culture (management structure)

Feedback from the case studies helped to refine both the guideline questions and the Framework itself.

3.5 Reflections on the GIM Assessment Framework

3.5.1 ILM and the Role of the GIM Assessment Framework

Stage 1 of the GIM Assessment Framework captures the ILM process, stewardship requirements and helps to identify candidate digital resources for rationalisation, disposal or archiving. Stages 2 and 3 are geared toward identifying options that can be used to realise the benefits of ILM and assessing the value of such options.

The GIM Assessment Framework can be used as follows:

- A scoping tool to identify potential GIM options and their potential costs and benefits.
- A guide to finding toolkits and methodologies that can aid assessment of an organisation's information environment, stewardship requirements, and the costs and benefits of greening options.
- An aid to development of a business case for implementing GIM options. This includes scoping and identification of a range of relevant tools.
- As part of a larger toolset to aid the assessment and management of information. Other tools might include DAF [15], DRAMBORA [16], ISMG [37], MANDATE [38], SusteIT ICT Energy and Carbon Footprinting tool [36], JISC infoNet Tools [19], BIILS [35] and INSIGHT [34].
- As a tool to raise awareness of the issues surrounding the GIM agenda and a motivator to encourage green attitudes, strategies and options in information management.

3.5.2 Observations from the case studies

Case study participants were not always familiar with the GIM options analysed in the Framework and sometimes there was confusion in what some of the techniques are, for example, de-duplication versus pruning/weeding. Some viewed de-duplication and pruning/weeding as potentially the same thing. However, in the context of the Framework pruning/weeding is viewed as a primarily manual activity whereas de-duplication predominantly an automatic software-controlled activity.

One comment noted from the case studies was that there is a need for more detail on what the costs and benefits of implementing and supporting specific GIM options would be. One main driver for establishing the detail on costs/benefits is the need to inform the development of a business case for senior management before any implementation of GIM options could take place. To aid this process some case study participants expressed an interest in investigating if the tools developed by the BIILS [35] or INSIGHT [34] studies might help with more detailed cost/benefits analysis.

3.5.3 Peer review

Case study participants were asked to peer review the Assessment Framework. Feedback gathered from this process was positive and did not result in any further refinements. The process confirmed that such a Framework is useful to help establish the efficacy of GIM options in relation to specific information environments. Supplementary tools and techniques would optimise the application of the Framework, since, for example, the information environment benchmarking exercise could be assisted using a tool like DAF and actual costs could be assessed using a tool such as the INSIGHT [34] methodology.

> It [the GIM Assessment Framework] was very useful for gaining a greater understanding of the topic. We're beginning to think of ways to improve our data management practices, so this study will likely be extremely useful in justifying changes to management practices and the purchase of different hardware.
> **The Centre for e-Research (CeRch) – King's College London.**

> The peer review process supported the usefulness of the Assessment Framework, which was found to be a useful tool, particularly for records managers.
> **University of Hull.**

3.5.4 Relation to Information and Records Management Impact Calculator

The JISC infoNet Impact Calculator was not released until after the GIM Framework and case studies were complete and it's implications could not be considered during that time.

Within the Framework, institutions are encouraged to assess or audit their current information environment, which ties in with the benchmarking stage involved in implementing the Impact Calculator. Potential costs and benefits are then identified. The ability to actually quantify likely benefits and to assess whether or not these outweigh the associated costs is an aspect of the Impact Calculator that compliments the Framework. The Impact Calculator can be used to supplement this stage in the Assessment Framework, to provide a better rounded overall picture of potential outcomes resulting from the introduction of an identified information management option. It is felt that the added-value of being able to offer a means to quantify likely benefits will act as a driving force in motivating changes in working practice and therefore an effective driver for the Green agenda.

The JISC infoNet Impact Calculator [39] comprises an "Impact Calculator Guidance Notes" document and two Microsoft Excel spreadsheets, one of which is the empty calculator

template and the other the calculator populated with a worked example. There are close parallels between the impact calculator and the GIM Assessment Framework in that they encourage benchmarking an organisation's current information environment or position and identifying potential costs and benefits associated with a change in business process.

The description on the JISC infoNet website is that: "This tool will include the means by which any organisation considering or currently engaged in an initiative to improve the management of records and information can capture and measure three distinct sets of data: performance information before and after completion of the initiative and the costs of implementing it. A comparison of the three sets of data will enable a more complete picture of the impact of implementing records management solutions to be derived through looking at its results in the immediate, short and medium term" [40].

The Impact Calculator focuses on a single process and is flexible enough to be used for a range of information and records management processes, both digital and non-digital. It provides a method of assessing different aspects of a management process that will affect an institution before and after the process is changed.

The information on the costs and benefits after a change in information management practice - referred to as a "process change" in the Impact Calculator - can be captured annually with the tool for up to 5 years. The tool can be used to estimate costs of a new initiative as part of business case before deciding whether to implement a new management process. This is achieved by entering estimates in the "to be" state in the calculator and later updating, if implemented, with real performance information after the first year of implementation. Actual savings over time can be calculated using the Impact Calculator.

Depending on the scope of the business process undergoing the change, a significant amount of effort may be needed to collect the information required for the calculator. The guidance notes advise consulting the relevant experts within one's institution to gather the information, which will spread the workload and add credibility to the final results [41]. Although the Impact Calculator focuses on a single business process that is to be redesigned/changed, multiple applications of the impact calculator can be used to calculate savings/benefits for a number of related initiatives.

The Impact Calculator includes the costs required to bring about the process change and additional annual ongoing costs. These are not broken down any further than staff and non-staff costs. The guidance notes provide a checklist of major cost headings to aid the identification of areas where costs may be incurred by the changes. The impact calculator could be of great help and compliment the GIM Framework by providing a tool to give predictive values to possible monetary and non-monetary benefits to adopting GIM options. The Impact Calculator is complementary to the INSIGHT Framework [34] which would be of particular help in identifying costs that will form input to the Impact Calculator. It is also complementary to the BIILS toolkit, [35] which has as a primary focus evaluating the methodologies to asses costs and benefits, but could also help in identifying costs and benefits.

4 Case Studies of Real-World Potential for Greening Information Management

Four case studies were conducted in order to produce a practical analysis of collections of information or data held by different institutions to determine the real-world potential for implementing more energy efficient digital resource management options. Participants were first issued with Stage 1 of the Framework, together with corresponding questions. Time was then given for them to report back on this section to provide a picture of the information environment being considered. Following a period of reflection, stages 2 and 3 of the process were then conducted by phone or in person where possible.

The case studies were selected to cover different types of digital resources and different contexts. The University of Hull case study examined the use of shared network drives, the University of Strathclyde case study focussed on Strathprints, the institutional repository, the King's College London case study focussed on archive collections held at the Centre for e-Research (CeRch), and the University Edinburgh case study focussed on its institutional records management policies and practices, with some additional input from EDINA, the JISC-funded data centre based there. These case studies are summarised in subsections 4.1 – 4.4 respectively. The section ends with reflections on the case studies. Full details of the individual greening information management case studies can be found at http://www.greeningim.org.uk/CaseStudies.aspx.

4.1 University of Hull: shared network drive

The managed digital resource featured in this case study was a shared network drive used by the Library, Archives, IT Services and Information Management section at the University of Hull. The resource is managed by Academic Services at Hull and is not currently subject to formal audits or risk assessments.

One outcome Hull hoped to achieve from the case study was a better understanding of the tools available to identify the costs and benefits of implementing new (green) information management options. This would enable the development of a business case for the possible implementation of new techniques. The three GIM options discussed were: manual pruning/weeding, enterprise-level data storage repository and resource limitation (quotas).

Manual pruning/weeding is currently used, although it is deemed desirable to introduce a retention schedule to increase current low take-up of the strategy. Enterprise-level storage repository usage is seen as beneficial but departmental-level staff may perceive a loss of control over their data. Quotas are currently used as a way of managing storage on the shared network drive. Staff circumvent this by downloading files onto mobile devices such as flash drives, which raises security and search and retrieval issues.

The use of an enterprise-level centralised data storage repository was selected as the GIM option considered most appropriate to the information environment explored at Hull.

Keys to success would include:

- developing a business case for an enterprise-level centralised data storage repository;
- providing a range of access levels (including private/restricted areas);

- significant promotion of the storage repository and its benefits throughout the institution.

Identified cost issues:

- short-term costs are likely to increase in relation to training, infrastructure, system support and software;
- in the long term, this investment will bring savings in terms of staff costs and compliance costs;
- increased functionality of the new system is likely to justify the initial cost of making the change;
- it remains to be seen whether these longer term benefits outweigh the initial costs incurred.

4.2 University of Strathclyde: Strathprints (Institutional Repository)

The managed digital resource featured in this case study was an institutional repository (Strathprints), an open access repository that provides access to all of the University of Strathclyde's research outputs and other material produced by University staff.

Three GIM options were deemed relevant to Strathprints. These are ECM (Enterprise Content Management), de-duplication and version control.

One benefit of ECM is that it would facilitate better integration of email correspondence and repository content. Wider organisational benefits resulting from the integration of repository content and associated email files include centralisation of repository related material at institutional level, increased transparency and decreased need for interoperability across related systems. It is thought likely that the use of an ECM system would result in a reduction in digital storage capacity, through the improved ability to identify information being held.

Duplication of records and/or items across systems is seen as positive, where a lack of interoperability is evident between such systems, since multiple records increase the visibility of research outputs and make preservation of the publication more likely.

One difficulty in establishing a policy relating to version control is that it is not always clear what constitutes a 'version', or exactly what constitutes a preprint, postprint, author final draft, and so on.

De-duplication was selected as the GIM option considered most appropriate for Strathprints. This technique may relate to duplication of items or de-duplication of content within items. The former is already being undertaken in Strathprints and the latter is deemed to be a feasible option for implementation. The introduction of further de-duplication techniques is likely to introduce added responsibility for proxies (departmental staff responsible for depositing outputs of all staff within their department). Highly ranked departments (in RAE terms) may be keen to deposit their work in Strathprints to create improved visibility and the converse may also be true. That is, lower ranking departments may be reluctant to deposit their work since visibility of outputs is unlikely to be considered desirable.

Likely benefits:

- reduction in staff time (and hence financial savings);

- increased compliance;
- increased visibility of individual researchers;
- reduced carbon footprint, efficient use of disk space.

Identified cost issues:

- competing costs and benefits for different parties within the institution;
- increased training requirements for proxies.

4.3 King's College London (KCL): The Centre for e-Research (CeRch)

The managed digital resource featured in this case study was the archive collections held at the Centre for e-Research (CeRch) which is located in Information Services and Systems at King's College London and incorporates the former Arts and Humanities Data Service (AHDS) Executive and the Methods Network. The main categories of resources include: legacy AHDS collections, JISC-funded projects and JISC collections (stored in a "dark" archive).

Formal assessment of CeRch's information environment has been undertaken using the DRAMBORA toolkit to assess the broad risks associated with data storage and management across all digital collections and projects. This has informed the development of management practices and procedures. The Preservation Exemplars at King's (PEKin) project [42] is currently using a combination of DAF and DRAMBORA to assess digital assets within the college. It is currently uncertain whether resources/files might be rationalised in some way. Although there is no legal requirement to maintain the AHDS data, there may be implications for the organisational reputation if it were rationalised.

Three GIM options were deemed relevant to CeRch. These are tiered storage, the use of a storage repository and de-duplication.

CeRch already implement a tiered storage model to a degree. The Centre plans to make greater use of this type of tiered approach to collection management. It is probable that the full range of information held could be handled in this way. There is a policy in place to store files uncompressed. The increased ability to prioritise the use of high-performance hardware is likely to result in overall energy savings at an institutional level.

Intra-file de-duplication is a potential means of reducing the disk storage used by CeRch. Within the AHDS collections, a small amount of duplication occurs, primarily in web site content. The applicability of de-duplication techniques to all types of information held by CeRch is dependent on the policies imposed by, and the contracts held with, information providers and funders. For example, JISC Collections and JISC projects could not be subjected to any de-duplication techniques introduced. The technique would result in storage, energy and financial savings. Considering duplication of entire files, and specifically within its role as the AHDS Executive, CeRch adopted a data management strategy that complied with the OAIS Reference Model, storing multiple manifestations of the same object on disk. A technique that could limit this type of duplication (e.g. a 'migration on demand' service, as proposed in the CEDARS project) would likely reduce storage requirements, but at the expense of increasing processing requirements to produce derivatives on-the-fly.

CeRch has plans to implement Fedora storage repository software to house a number of its collections. Fedora will be used to manage all the different manifestations of an object. Various automated systems will be introduced to handle metadata creation, format conversion, and so on. The process of introducing Fedora would bring implications for managing associated processes. Considerable changes would be required in terms of staff training, infrastructures operated and so on. CeRch are introducing it on a gradual basis, to try to identify all the areas of change that are likely to require management.

Centralised data storage repository was selected as the GIM option considered most appropriate.

Likely benefits:

- reduce the storage capacity used through the ability to create relationships between related digital objects;
- increase effective stewardship due to the introduction of automated processes, previously undertaken on a manual basis;
- enables more efficient compliance with legislation;
- users of the service are likely to develop a perspective of trust;
- common infrastructure could potentially be applied to many different projects, thereby optimising funding and streamlining working practices and processes.

Identified costs:

- training requirements would increase in the short term but may reduce overall in the longer term;
- overheads are likely to be reduced (through automation);
- systems support requirements will change, although it is unclear whether this will constitute an overall cost or benefit;
- more development staff would be required in the short term.

Potential barriers to the introduction of Fedora include:

- lack of interoperability with other existing and established services;
- lack of general acceptance of open source products;
- uncertainty of maintenance costs and requirements of in-house systems when compared to commercial products;
- lack of a business case for green computing lack of business justification.

4.4 University of Edinburgh: Records Management and EDINA

The University of Edinburgh case study took a different approach to the rest, in that it did not focus specifically on a managed digital resource. Rather, it considered the benefits and drawbacks associated with existing records management policies within the institution.

Of the list of seven GIM options presented in the Assessment Framework, it was found that the Records Management Section currently implements two – pruning/weeding and version control – and also promotes the active retention and disposal of records in line with retention schedules established for a range of specific information types.

Pruning/weeding and version control were deemed appropriate options that could be applied across an institution. Likely benefits of pruning/weeding include:

- improved search time
- rationalised results sets
- reduced IT backup and restoration times
- improved compliance for e.g. Freedom of Information and Data Protection requests

Likely benefits of version control include:

- increased efficiency in working practices in the creation and development of electronic documents
- improved ability to assess files for retention or disposal

Identified cost issues:

- overall changes in the price/performance profile of modern storage devices is such that energy saving is unlikely to be significant unless the degree of redundancy was excessive

4.5 Reflections on the case studies

4.5.1 Greening information management options

In the first three case studies, participants provided insight into the greening GIM options they considered most feasible to introduce to their current information environment. The University of Edinburgh case study differed in perspective as it considered options that are already implemented within the institution and for which associated policies are in place; namely, retention and disposal, manual pruning/weeding and version control.

Options selected by case study participants varied, although three options were each chosen by two institutions. Of the remaining four options presented each was chosen once. This indicates that each of the seven options included in the Assessment Framework is considered to have the potential to green information management in specific information environments. Edinburgh discussed retention and disposal, which is manual pruning/weeding in line with a retention schedule rather than on an ad-hoc basis.

Across three of the case study environments - Hull, Strathclyde and CeRch - both the de-duplication and centralised storage repository options were selected twice. Manual pruning/weeding was chosen by Hull and Edinburgh and version control was chosen by Strathclyde and Edinburgh. All other options chosen (3 in total by each participant) were unique. A summary of these findings is presented in Table 1 below.

Institution	Greening Information Management Option selected		
University of Hull: Academic Services (Institutional Shared Drives)	Manual pruning/weeding	Centralised storage repository	Resource limitations (quotas)
University of Strathclyde: Strathprints (Institutional Repository)	De-duplication	Version control	ECM
King's College, London: Centre for e-Research (CeRch) (Archive Collections)	De-duplication	Centralised storage repository	Tiered storage
University of Edinburgh: Records Management Section	Manual pruning/weeding	Version control	Retention and disposal (pruning/weeding)

Table 1: Summary of GIM options selected in case studies

Results suggest that manual pruning/weeding, de-duplication and the use of a centralised storage repository are options that are potentially applicable to a range of different types of information being held. The most well-defined and heavily structured collection of information included in the case studies is probably that of Strathprints, Strathclyde's institutional repository. The very nature of this resource means that the centralised storage repository was not a feasible option to consider (since it already constitutes just that). Similarly, quotas were not deemed to be a potentially applicable option here since the institutional repository is required to make available all research output. A restriction on file size or server size would therefore be inappropriate. In the context of a shared network drive, however, as at the University of Hull, quotas were deemed to be of great potential value as a means of reducing the overall amount of data stored. It was highlighted however, that staff try to find ways to work around the quotas imposed, such as storing files using portable media, rather than addressing the issue at source.

Tiered storage was considered a viable potential option by CeRch, who provide information collections with different access permissions. It was felt that in this context, where some material can be held offline, it would constitute good environmental practice to move files that were less well used, or further on in their information lifecycle, to a less energy intensive platform. It is unlikely that such an option would be considered useful in the repository context, since this tiered storage approach can effectively be created within the EPrints software itself e.g. by making files accessible to repository administrators only, or to all users.

4.5.2 Costs, benefits and impacts

All the case studies identified that there were both costs and benefits associated with GIM options. However, due to time and resource constraints it was not possible to ascertain whether the benefits (greening or otherwise) would outweigh the costs, both in the short and long term. More in depth study to examine in detail the potential costs and benefits is needed to build the business case that senior management would require before

implementing any new greening information management option. However, the case studies in effect scoped the costs and benefits that need to be considered in detail, which is in itself a useful first step of the business case development process.

The Hull, Strathclyde, KCL and Edinburgh case studies also identified a high likelihood of staff changes (types, numbers, working procedures) would be required to implement most GIM options.

Should a formal master document management or version control procedure be introduced to Strathprints; workflow, associated working practices would be devolved to proxies (i.e. departmental staff responsible for depositing outputs of all staff within their department). Ad hoc checks to remove different versions of a publication currently lies with library staff; should a more formal and robust method be introduced, it would be best handled by the proxies responsible for uploading material to the repository.

Alan Slevin, Strathprints, University of Strathclyde

It is thought likely that the introduction of further de-duplication techniques will result in a long-term reduction in costs as a result of decreased staff time required to manage related tasks.

Alan Slevin, Strathprints, University of Strathclyde

The University of Hull predicts a reduction in staff time, as a result of introducing greening information options, due to the improved efficiency in locating documents for all purposes.

Vicky Mays, University of Hull.

A common infrastructure, such as Fedora, could potentially be applied to many different projects, thereby optimising funding and streamlining working practices and processes. The increased ability to automate procedures as a result would have a direct impact on staff time.

Gareth Knight, CeRch – King's College London.

Some participants felt that the Green agenda was not a strong motivator and although institutions may have instituted green polices this has not always translated into changes in information management practice. One participant noted a perceived conflict between

keeping multiple copies of records to aid preservation and a more greening approach which advocates reducing overall disk usage.

Some confusion over the validity of participation was evident. The GIM Assessment Framework is intended to assess the efficacy of GIM options within specific information environments, yet some parties approached were more comfortable directing the project to a representative of the broader institutional viewpoint, as they did not feel they would have the level of responsibility required to recommend implementation of specific options. It was generally found, however, that those representing the wider institutional perspective were unable to commit the time required for the case study process to be undertaken.

4.5.3 Baselining information environments

Two of the institutions have used formal ILM related tools (e.g. DAF) to assess their information assets and this has informed the development of information management practices and procedures.

As part of the PEKin project [42] at KCL, DAF has been used to inform a new data assessment methodology to help ascertain e.g. where data is stored and who is responsible for it. This has been combined with the risk assessment element of DRAMBORA, to provide a robust Assessment Framework. Based on requirements and risks, decisions can be made on when resources should be moved to an alternative storage platform, and what nature this platform should take.

The cross-service repositories and preservation group at the University of Edinburgh has been working on auditing EDINA's business data using DAF, using the Inventory of Data Assets.

4.5.4 Information Lifecycle Management

Case study participants' knowledge of ILM theory varied. The key case study participant at the University of Hull is a records manager, so ILM is considered a key concept to effective information management within that institution in relation to both paper and digital records.

"I did see many overlaps/parallels between ILM theory and the case study process."

Vicky Mays, University of Hull.

KCL's case study contact described the importance of ILM to the institution's PEKin project:

> It [PEKin] investigates the information lifecycle of research and administrative data. Although these data types differ in terms of the length of time for which they must be stored and the activities that they must support, management of the data lifecycle requires the application of similar processes. A common infrastructure to manage the lifecycle of such resources is being developed.
>
> **Gareth Knight, CeRch – King's College London.**

The remaining case study participants did not allude to the relevance of ILM theory, although this is not to say if they are familiar with it. Due to likely variation of existing knowledge regarding ILM and the current project's desire for participants to focus on specific GIM options in the time they had available for contribution, it was decided that explanations of ILM theory would not form part of the case studies per se; rather, the theory would inform the case study process and questions asked therein (as also mentioned in **Error! Reference source not found.**).

4.5.5 The case study process

Finally, while the case studies were successful in providing a practical analysis of collections of information or data held by different institutions to determine the real-world potential for implementing more energy efficient digital resource management options, this process was also successfully used to test and refine the overall GIM Assessment Framework and to produce the practical 'toolkit' presented in subsection 3.4. However, securing case studies to participate in the Greening Information Management project proved difficult and areas such as JISC data centres and email repositories were not included. This may be attributed, in part at least, to time constraints and to two specific concerns. Firstly, as also encountered in the activity review, a perceived lack of activity in the specific area of green information management meant that prospective participants felt they had little to offer. Secondly, when looking at non-central information environments, some information managers felt that the ability to reduce digital resources depended on institutional strategies and/or centrally provided technologies and therefore they were not in a position to comment.

5 The Efficacy of Greening Information Management

Whether greening information management techniques can make a sector-wide contribution to the Green ICT agenda was assessed by undertaking an impact analysis which examined the direct and indirect costs, benefits and implications of implementing green information management techniques. The impact analysis focussed on the following questions:

1. *Can information management techniques be used to reduce digital storage without compromising stewardship?*
2. *What are the direct and indirect costs associated with implementing greening information management techniques?*
3. *What are the resulting environmental and other benefits associated with the reduction in digital storage? How do these compare with costs? Are there negative impacts?*
4. *How feasible is it to implement green information management techniques at present? In the next few years?*

Drawing on the findings from the literature and activities review and the real-world case studies as evidence, each of these questions is considered below. This is followed by reflections on the overall efficacy and the analysis process.

5.1 Can information management options be used to reduce digital storage without compromising stewardship?

As the case study summaries of section 4 illustrate, the evidence from the case studies suggests that in principle information management techniques could be used to reduce digital storage without compromising stewardship. However, this is an 'in principle' assessment only.

A case study from an industry magazine [43] of Burt's Bees' details how the company used a number of techniques including hardware upgrading, server virtualisation and de-duplication to substantially reduce their energy consumption in their data centre. The organisation was fully committed to the Green agenda and very motivated to explore ways to introduce greener ICT whilst still maintaining stewardship over its information.

If green information management is to be realised in practice then the barriers discussed in sections 2 and 4 need to be overcome. In particular, steps will need to be put in place to overcome the extant culture of retention of digital resources. Critical reflection of stewardship requirements, information management practices and working practices will be required. Sector leadership by JISC and strong institutional leadership will be critical if this is to be effective.

5.2 How feasible is it to implement green information management techniques?

The feasibility of implementing greener information management depends on there being scope for improvement and the ability to overcome the barriers identified in the preceding sections. It can thus be considered in terms of the following 5 factors. Firstly, there needs to be an excess of digital resources within the information environment that can be removed without loss of stewardship. This could be either due to duplication or inactive material. Secondly, there needs to be an appropriate GIM technique be it a technological or procedural means of reducing the digital resource storage. Thirdly, there needs to be sound business rationale for implementing GIM. Fourthly, staff and students need to be provided with appropriate support to GIM. Finally, without an embedded culture of environmental responsibility it is likely that any GIM strategy will be circumvented in practice.

Table 2 p40 provides an analysis of the factors affecting feasibility of implementing GIM by summarising the feasibility criteria for implementing GIM (column 1), what they depend on (column 2), how they might be achieved (column 3) and who the key stakeholders are in achieving feasibility (column 4). The feasibility of GIM in a given context will depend on whether the appropriate requirements indentified in column 4 can be met. As the analysis in Table 2 indicates, both institutions and JISC have a role to play in the feasibility of GIM. Further, not all requirements are currently in place. For example, as discussed earlier, the current prevailing attitude that good practice entails the retention and preservation of digital resources is problematic for GIM. It is therefore not practical to provide a definitive answer regarding the feasibility of GIM. Rather, Table 2 provides a checklist of actions that individual institutions and JISC need to undertake if GIM is to be feasible.

Feasibility criteria	Depends on	Requirements	Action required of
Potential to decrease digital resource storage	Specific information environment: - Current practice - Stewardship requirements Prevailing attitude to retention	Information stewardship requirements; Change to retention attitudes	Institutional information manager; JISC
Means of reducing digital resource storage	Given IE context; Range of GIM options;	Indentify range of GIM options which can facilitate a reduction in digital storage in the given IE context	Information managers
Sound business case for GIM	Direct and indirect costs, benefits and impact	Tools to develop sound business case Identification of costs, benefits and impacts	JISC Business and information manager
Support for implementation of GIM	Develop holistic information lifecycle management appropriate to context	Leadership Training/Information literacy skills	Senior management All those storing data
Embedded culture of environmental responsibility	Institutional and sector leadership	Embedding of environmental responsibility across all strategies and policies	Senior management, JISC

Table 2: Analysis of factors affecting feasibility of GIM

5.3 What are the direct and indirect costs associated with implementing greening information management options?

Direct costs relate to resources or activities which have a direct benefit on the activity or project under consideration or are necessary for the aims and objectives of the activity or project to be achieved. As Table 2, p40 identifies the activities required to make GIM feasible, it also can be used as a starting point for identification of direct costs, suggesting 5 categories of costs. Firstly, there will be direct costs associated with development of a GIM information strategy and policy. These are likely to relate to staff time or buying in expertise in information management. Secondly, there will be direct costs associated with the chosen GIM technique. These include: hardware or software costs directly associated with the chosen GIM technique – e.g. the introduction of a tiered storage solution or de-duplication technology. These may also require new technology or information management support structures. Thirdly, there will costs associated with developing the business case – again primarily staff time. Fourthly, there will be costs associated with change management and support. These will include dissemination activities and staff training. Finally, there will be costs associated with development and implementation of institutional strategies and policies on the environment. These will again primarily involve staff time and as they are required for general environmental responsibility, they will only partially be attributable to GIM.

Indirect costs by their very nature are much more difficult to assess. In general, these will need to be identified on a case by case basis. Potential indirect costs include those associated knock-on effects to working practices and business processes. For example, introduction of information lifecycle management and say pruning techniques may mean that information can be retrieved more effectively, thus reducing time on task. Thus the cost could be in fact a net saving. Alternatively, it could lead to extra time being spent on manual archiving or pruning of information and hence a net cost.

As UCISA argues, a full business case for ICT related activities should also include consideration of the cost that will be incurred from not implementing a particular option [44]. Such costs are in general not considered within the HE sector and are therefore not well researched. However, in the case of greening information management such costs might reasonably include those associated with: inefficiencies in information retrieval due to duplication; invalid data due to retention of out of date material; loss of students or problems in staff retention due to perceived lack of environmental awareness; and potentially penalties for not achieving government or funding council targets on energy savings. While not directly transferable, the outputs of the UCISA study into the ‘Cost of Downtime’ which is currently being undertaken should provide a general method for identifying and estimating costs associated with not implementing greening information management.

Costing GIM not only involves identification of where the costs arise but also being able to measure the costs. As the reports of the JISC-funded INSIGHT Cost-Benefit [34] and Benefits of Investment in ICT Landscape Study (BIILS) [35] projects show, identification of costs is also limited by the types of costing systems currently employed within the HE sector which makes it difficult to readily identify the full costs associated with services[2]. However,

[2] This is beginning to change in research intensive institutions where the requirement to use full economic costing based on TRAC for research proposals is driving change.

marginal costing approaches which focus on the change in costs are in widespread use and should be suited to identification of the costs associated implementing GIM. Measuring the costs may also be further complicated by the fact that GIM may be implemented in conjunction with wider information management solutions such as introducing an enterprise architecture solution. Deciding which costs are attributable to the GIM element will be more problematic and a simple percentage of costs may need to be used.

5.4 What are the resulting environmental and other benefits associated with the reduction in digital storage? How do these compare with costs? Are there negative impacts?

Until there are well-established real-world implementations of GIM it is difficult to identify all the potential benefits. However, drawing on the literature and activity review and case studies, three broad categories of potential benefits can be identified – environmental benefits, business process benefits and reputational benefits.

5.4.1 Environmental benefits and impacts

As discussed in sections 1 and 3, the theory behind GIM is that reduction in storage of digital resources equates to a reduction of energy consumption related to provision of data storage. This theoretically affords environmental benefits as the energy savings reduces carbon emissions.

Unfortunately, removal of a single file does not necessarily lead to energy savings; rather energy savings arise when the overall reduction in the amount of digital resources means that a disk array or module can be powered down. As Table 3 below illustrates, the number of digital resources which need to be removed to achieve this depends on the type of resource.

File Type	Typical size	Number per Gb	Number per 500 Gb drive
Administrative	100 Kb	10000	5 million
Image file	1 Mb	1000	500000
Scientific data	10 Mb	100	50000

Table 3: Typical sizes associated with different digital resources

The corollary to this is that a reduction in digital resources could, in the interim at least, lead to increased energy inefficiencies as disks with little utilised capacity will generate a higher energy cost per stored file.

When it is established that there will be a reduction in disks, the SusteIT energy and carbon footprinting tool [36] can be used to calculate the energy save. This calculation is based on three variables:

1. Numbers of devices
2. Power consumption data for individual types of device (watts, either measured or taken from manufacturer's data)
3. Usage data of devices in a FHE setting (hours/year).

Using the calculations from the SusteIT energy and carbon footprinting tool an estimate of the annual power use for a single 500 Gb disk server based on the following assumptions is provided in Table 4 below:

- High performance disk servers that have disk drives that is 500 Gb in size
- That the servers are rated 300 watts.
- The server is in active/standby for 8,064 hours (336*24), this assumes the server is off a total of 4 weeks a year.
- In areas where servers are normally located (server room) there will be overheads for storage equipment, cooling and power supply.

Server Rooms	Number of 1U servers	Watts per server	Overhead	Total kWh/year
Disk servers	1	300		2,628
Storage & other server equipment			0.150	394
Server/equipment sub-total				3,022
Server cooling overhead			0.425	
Server power supply overhead			0.075	
Server cooling & power supply overhead			0.500	1,511
Server sub-total				**4,533**

Table 4: Estimate of the annual power usage for a 500 Gb disk server

In order to make the saving of 4533 kilowatt hours per year (kWh/year) predicted in Table 4 a reduction in the storage of files is needed which equates to approx. 5 million admin files or half a million image files or fifty thousand scientific data files – see Table 3, p42. Normal one would assume a heterogeneous mix of files on a server and these examples are only for illustration only.

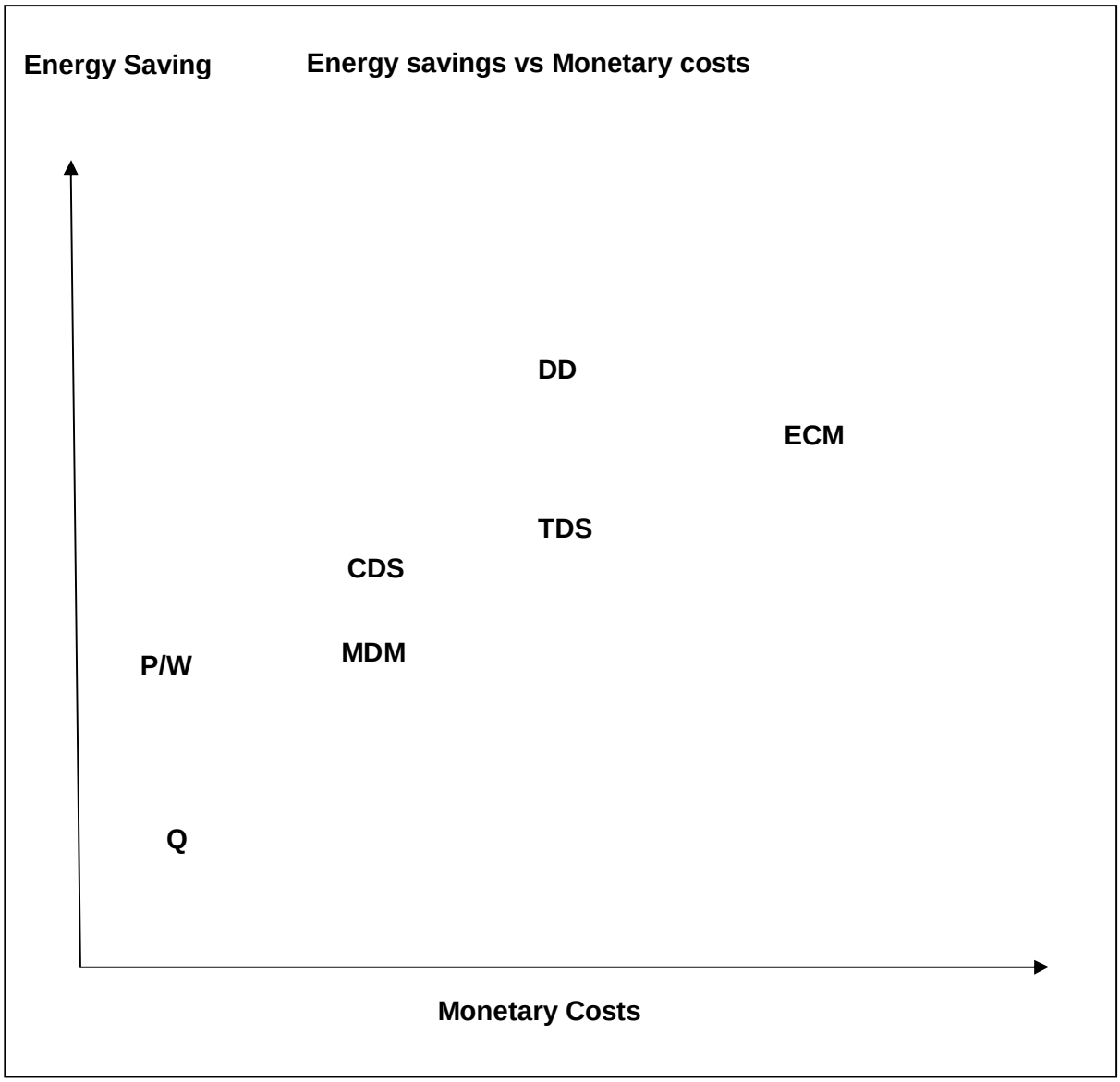
Using the SusteIT energy and carbon footprinting tool's estimates for energy costs (12p per kWh) and energy – CO_2 emission conversion factor (0.537 per kg CO2/kWh) an estimate of annual cost saving and reduction in carbon emissions brought about by taking the server out of use can be calculated:

- Cost saving = 0.12 x 4533 = 544 (£/year)
- CO2 reduction = 0.537 x 4533 = 2434 (kg CO2/year)

Equating an x% reduction in disk space to y% energy savings is too simplistic. The energy consumption arises from what James and Hopkinson [1] term embedded energy used during the manufacturing process, consumption energy which is spent during the lifetime of the equipment and disposal energy spent during decommissioning. As the various GIM

techniques may involve introduction of new hardware or software and the decommissioning of old ICT, all three types of energy consumption need to be considered.

The diagram below shows the relative position of the greening options in relation to energy savings and monetary costs.



The greening options in the quadrant diagram above are:

ECM – Enterprise Content Management

DD – De-duplication

P/W – Manual Pruning/Weeding

CDSR – Centralised Data Storage Repository

TDS – Tiered Data Storage

MDM – Master Data Management

Q – Quotas

Care must also be taken not to oversimplify the energy consumption calculation. For example, while a reduction in storage requirements might mean less disks are required, it does not necessarily mean that there will be a reduction in information retrieval requests. Indeed, improved information management could afford the ability to undertake much more detailed business analysis activities which are computationally intensive.

5.4.2 Business process benefits and impacts

As indicated in the discussions of costs in subsection 5.3, the introduction of GIM may afford the business benefits associated with good practice in information management. As the JISC InfoNet infoKit on Information and Records Management [45] identifies, these include:

- Easier retrieval
- Reduction in purchase and maintenance costs
- More efficient access to documents for all uses including legislative compliance (e.g. FOI)
- Requires management buy-in to greening agenda.
- Emphasis on value of information.
- Helps organisations to think of information strategically

The JISC InfoNet Impact Calculator can help by focusing on the change of process (process redesign) that implementing a GIM technique will require and aid quantifying both monetary and non-monetary benefits [41]. It complements the GIM Framework by allowing both predictive and actual benefits to be identified and measured for a new GIM initiative.

5.4.3 Reputational benefits

As the issue of global warming increasingly comes to the fore, it seems reasonable to project that institutions with strong environmental policies which permeate all activities will reap reputational benefits, leading indirectly to institutional benefits. For example, institutions may be able to access improved funding based on reductions in carbon footprints, attract and retain staff and students. Further discussion of benefits can be found in the Suste-IT report [1].

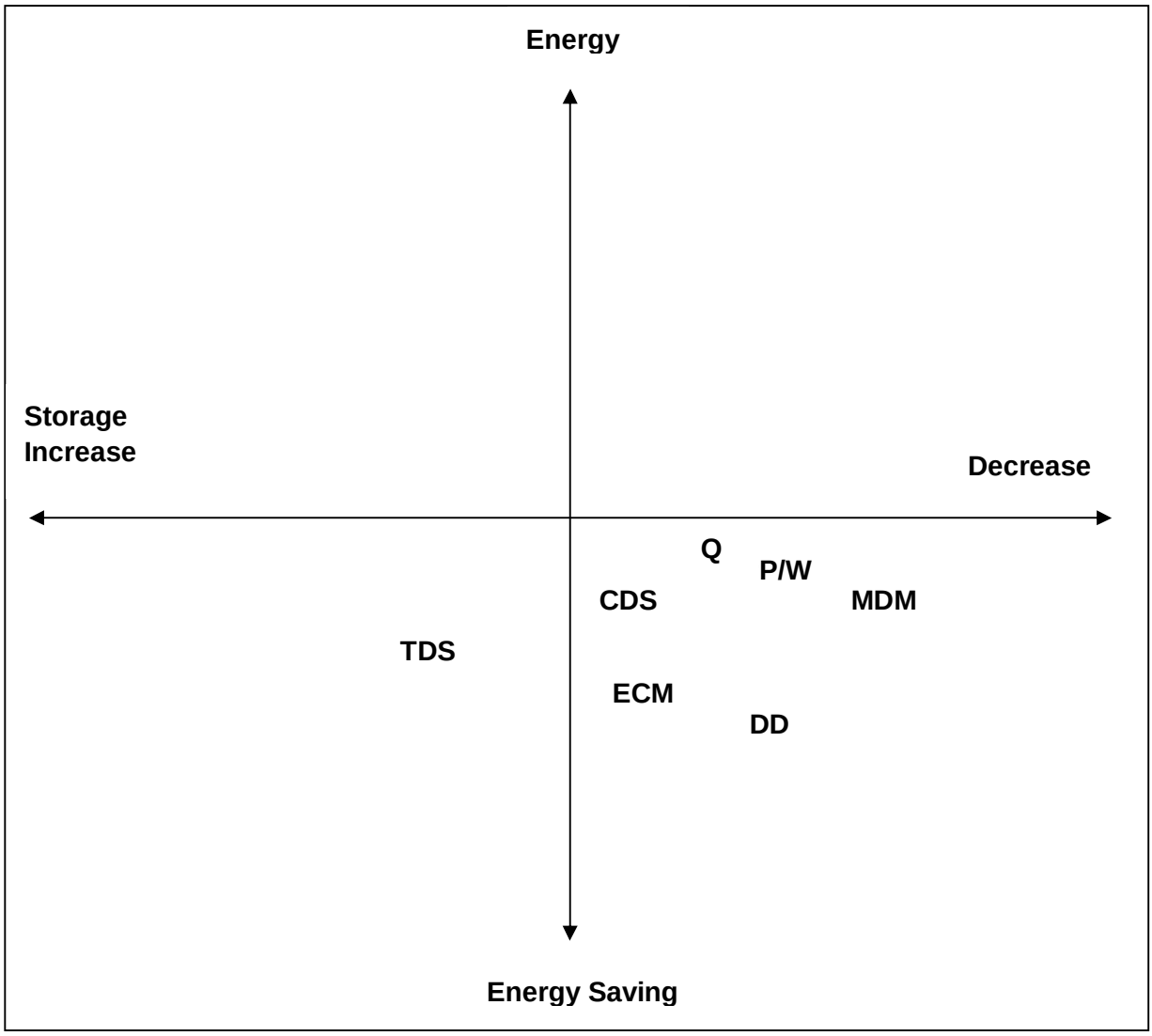
5.5 Reflections on the efficacy of greening information management

The analysis undertaken in the case studies supports the hypothesis that information management can be used to reduce the storage of digital resources and where the reduction is sufficient, it could contribute to a reduction in the environmental impact of HE. However, as discussed, this efficacy assessment is based on the opinions of the case study participants and extrapolations rather than a quantitative study of real-world information management in action. The range of case studies and types of digital resources which were considered does provide a degree of reassurance regarding the applicability of the efficacy analysis; however additional quantitative study is recommended.

Further, before any recommendation could be made regarding the overall role of GIM options within an institutional or sector green strategy, how the environmental saving which these afford compare with (a) that expected by the government, funding councils and society and (b) those afforded by other Green ICT approaches relating to hardware etc needs to be

considered. For example it is not yet known whether the introduction of tiered storage would be more energy efficient that an existing infrastructure.

The quadrant diagram overleaf shows the relative position of the greening options in relation to the two factors: Energy (Savings to costs) and Digital Storage (Increase or Decrease).



The greening techniques in the quadrant diagram above are:

ECM – Enterprise Content Management

DD – De-duplication

P/W – Manual Pruning/Weeding

CDSR – Centralised Data Storage Repository

TDS – Tiered Data Storage

MDM – Master Data Management

Q – Quotas

6 Conclusions & Recommendations

Having analysed the efficacy of greening information management in the previous section, the conclusion of the study and its implications for JISC, institutions, information managers and staff and students in general are briefly considered. This is followed by four recommendations for future work.

6.1 Conclusions

6.1.1 The current state of green information management

Limited activity is being undertaken to help make institutional information management processes greener. Although some energy saving practices are undoubtedly underway in relation to ILM and the retention and disposal of information, such practice is rarely viewed as being motivated by the Green agenda. Rather, it is promoted as a means of streamlining processes or optimising search behaviour; that is, much of the information management activity being conducted is indirect, in that it is not undertaken with a view to reducing energy consumption.

Of the limited activity that is taking place, the majority is being conducted in a piecemeal fashion, with little central coordination or holistic vision. This makes it difficult to quantify success and to measure actual reductions in disk space and energy consumption.

6.1.2 Key barriers and enablers to greener information management

Barriers to achieving greener information management identified were:

- Perceived inadequacy of institution-wide information management strategies and associated infrastructure, resulting in people using non-established working practices
- Perceived conflict between business requirements (stewardship etc) and green information management
- Technology limitations
- Statutory and regulatory requirements
- Culture, working practices and lack of trust
- Perceived increased cost
- Lack of a coherent strategy regarding the greener management of digital resources

Enablers of achieving greener information management were found to be:

- Linking the environmental agenda to existing mandatory requirements
- Competitive advantage
- Aligning the GIM case with the business benefits and embedding policies therein
- Organisational leadership
- Effective change management
- Information literacy

6.1.3 The efficacy of information management as a Green ICT tool

Information management options have enormous potential to green the amount of, and way in which, information is stored within an organisation. Providing stewardship and legislation is not compromised, the study supports the thesis that implementation of GIM options within an overarching information lifecycle management strategy should contribute to a positive environmental impact, provided they are incorporated into institutional policy and endorsed by senior management.

The efficacy of options identified in the GIM Assessment Framework requires to be more thoroughly assessed on a longer-term basis. The implementation of options should be investigated, considering costs and benefits across the short, medium and long term. Without such investigation it is difficult to predict whether options will bring significant benefits relating to environmental impact or whether likely benefits may also bring unexpected costs. It should also be borne in mind that environmental benefits arising from the implementation of such options may result in costs in other areas (e.g. staff costs, training needs), which require to be examined in detail and weighed up, or may prove to be misaligned with institutional policy.

A recent discussion held on the JISC Repositories mailing list [46] indicates the difficulties faced in attempting to estimate the quantity of data generated within specific areas of academia and the storage capacity required to accommodate this data. This makes it difficult to quantify associated costs, although there is undoubtedly a clear environmental impact in terms of the embedded energy relating to hardware and media, and consumption energy relating to hardware and data cooling.

Further work is required to ascertain the overall effect of implementation, both for individual and for combinations of options, before quantification of benefits and costs can be achieved. Such work will confirm (or otherwise) the overall efficacy of using such options in the HE sector and will also help to promote the Green agenda and motivate the adoption of best practice in the area.

6.1.4 Potential GIM strategies

Each of the seven GIM options included in the Assessment Framework was considered, by case study participants, to have the potential to green information management in specific information environments. Results suggest that manual pruning/weeding, de-duplication and the use of a centralised storage repository are options that are potentially applicable to a range of different types of information being held. The implementation of quotas, ECM and tiered storage were considered applicable to more specific information environments.

6.1.5 Timeliness of the study

The innovative nature of the approach being investigated – i.e. seeking to reduce the environmental impact of ICT through a reduction in digital resources – meant that the study has struggled to find relevant literature, examples of good practice in this area and case studies. While, as discussed above, this means that it is impossible to draw a definitive conclusion regarding efficacy, there is sufficient evidence to support the premise. While this continues to position JISC at the forefront of the Green ICT agenda, further work is required in the dissemination of the approach as a method for greening ICT, in gathering further longer-term evidence to develop the business cases and in developing good practice advice and use case scenarios.

6.2 Implications

6.2.1 For JISC

Two major implications for JISC can be identified from this study. Firstly, JISC will need to assimilate the findings of this study into its overall Green ICT programme of work. Secondly, if GIM is to be a feasible component of institutional and sector greening strategies then the impact on other existing strategies and programmes needs to be considered to ensure consistent and coherent messages are sent out. In particular, aspects of the current focus on the Digital Preservation and Asset Management Programme [47] are counter to the core thesis of green information management. The general premise of preservation, whereby it is recommended that multiple copies of digital material be kept (e.g. LOCKSS [48] for e-journals) is at direct odds to the Green agenda. Two of the benefits identified in 'A Continuing Access and Digital Preservation Strategy for the Joint Information Systems Committee' [49] are:

- decreasing cost of digital storage: the message here is that because hardware is becoming low-cost, more can be purchased in order to store more materials; the fact that this practice will result in increased energy consumption is overlooked;
- replication: digital materials can be replicated cheaply so multiple preservation copies are distributed in different geographical locations; again, increased energy consumption resulting from serving multiple copies of materials is overlooked.

Further, the ArchivePress [50], a new JISC-funded project being undertaken by ULCC and the British Library is considering the potential for WordPress [51] to "capture and archive blog content dynamically, *as it happens*". Such an automated approach is likely to result in much material being stored that may not contain any content that would be considered active, or semi-active, in ILM terms. Dynamic storage of 'real-time' information does not dovetail with advocated options from a greening information management perspective, such as adopting a retention schedule, assessing material for deletion/archiving or undertaking manual pruning/weeding, for example.

It follows that recommendations and guidance for good practice in the areas of records management, preservation, curation and so on, as issued from JISC, require to acknowledge the Green agenda, with a view to advocating effective institutional practice, but that also optimise energy efficiency.

6.2.2 For institutions

Institutions require to formulate integrated policy, advocated from a senior level. Such policy should be holistic, well embedded within the institution and complementary to existing policy. Appropriate changes to working practices will require to be made in line with such policy and additional training and potential staffing changes accommodated, both in the short- and longer term.

6.2.3 For Information Managers

Information managers will need to re-examine the rationale for retention of information, moving away from a prevailing attitude that information should be retained where possible, to one which seeks to actively prune information where stewardship permits. A critical analysis of current stewardship requirements will also be required as anecdotal evidence

illustrates that a 'belts and braces' attitude to stewardship prevails in some institutions, rather than one based on a sound legal or business basis.

Implementation of these changes requires a fundamental rethinking of information management objectives. To achieve the required transformation information managers will need to embrace a change in mindset, embedding the balancing of green and business requirements in their work.

The information manager's work may also become more challenging as they may be perceived as the enforcer of unpopular changes. Support and training will be required. The JISC-funded Work-with-IT and Embed-IT projects [52] provide useful advice and change management guidance for implementing technology-enhanced changes to working practices.

6.2.4 For staff & students

The introduction of greening information management techniques may significantly impact working practices. As with the Green agenda in general, if greening information management is to be successful it will require individuals to take ownership of the solution and embed it in their practice

6.3 Recommendations

6.3.1 Quantitative action research be undertaken to assess the potential capacity for reducing disk space and energy consumption through the GIM options identified in the project.

The Greening Information Management project has identified a number of options that can be implemented with a view to reducing disk space. Based on the assumption that a reduction in disk space leads to a reduction in power, the efficacy of employing these options has been examined in relation to a range of different types of information environment. There is no quantifiable evidence, however, to say that this is indeed the case in practice.

To illustrate this with certainty a longer term study is required, within which an institution's current practice would be benchmarked in quantitative terms. That is, counts of files, file sizes, server sizes, energy consumption and so on would be established. Individual GIM options would then be introduced over a set trial period. A second set of measures would then be taken to ascertain the extent of reduced disk space (or otherwise) and hence energy saving achieved.

6.3.2 Evidence-based good practice advice relating to GIM be developed.

A mismatch currently exists between information management and the Green agenda. The study has highlighted that HE institutions typically publish policy documents on recommended practice for information management, whether it be at institutional,

departmental or individual level. Such policy often exists in relation to a number of different types of information, detailing how such resources should be handled, retained, disposed of and so on. It follows that awareness of the value of information management in the traditional sense, exists. Awareness also exists for the Green agenda, particularly in relation to the use of ICT, the use of paper and the efficiency of buildings in terms of heating and lighting. It appears, however that the benefits of effective information management is not viewed as a motivator to progress the Green agenda. There appears to be a gap between the well-established understanding of information management per se, and its potential importance for furthering the efficiency of energy usage within the HE and FE sectors.

Evidence-based good practice advice relating to green information management would greatly help organisations to choose appropriate information management solutions.

6.3.3 JISC needs to work with other organisations and the community in general to arrive at holistic positions which address the Green agenda and ensure effective stewardship of resources.

JISC is actively promoting a Green agenda through recent calls and ICTs. At the same time however, other initiatives are promoting the preservation of blogs and wikis and web archiving programmes as instigated by the British Library [55] and JISC. Such initiatives mean that, potentially, duplicate resources are being stored in multiple places.

The potentially conflicting messages from such initiatives are problematic. If information management is to be used to reduce the environmental impact of ICT, JISC needs to work with other organisations and the community in general to arrive at holistic positions which address the Green agenda and ensure effective stewardship of resources.

6.3.4 The GIM Assessment Framework and the JISC infoNet Impact Calculator be used to create a practical and useful tool to assist information managers in making greener decisions and formulating business cases

The JISC impact calculator could be of great help and compliment the GIM Framework by providing a tool to give predictive values (and ongoing costs) to possible monetary and non-monetary benefits to adopting greening information management techniques. The most effective way to identify costs and benefits of implementing GIM techniques is likely to be by using a combination of: the GIM Framework, the JISC impact calculator [39] and the INSIGHT framework [34].

The outcome(s) of the third phase of the LIFE project [56], running from August 2009 for one year, also has potential value in complementing the GIM Assessment Framework. Closely associated with the stages of the ILM model, LIFE is developing a predictive costing tool to help inform the planning and management processes involved in preserving digital content.

The investigation of further syntheses between the GIM Assessment Framework and other related JISC outputs is highly recommended to provide a means of informing organisations of potential costs and benefits of improved information management, more specifically from an environmental viewpoint.

7 References

1. James, P. and L. Hopkinson, *Sustainable ICT in Further and Higher Education*. 2009.
2. McDonald, D., A. MacDonald, and E. McCulloch, *Greening Information Management Assessment Framework*. 2009, University of Strathclyde.
3. *Greening ICT programme*. [Last Accessed 06/11/09]; Available from: http://www.jisc.ac.uk/whatwedo/programmes/greeningict.aspx.
4. The Data Management Institute. *The Green Data Project*. 2007 [Last Accessed 02/12/09]; Available from: http://www.greendataproject.org/.
5. Hopkinson, L., James, P., *Energy and Carbon Impacts of ICT - User Guide for the SusteIT Footprinting Tool*. 2009. p. p1-17.
6. McDonald, D., et al., *Greening Information Management: a focussed literature and activity review*. 2009, University of Strathclyde. p. 1-42.
7. University of Strathclyde. *Greening Information Management Study*. 2009 [Last Accessed 30/10/09]; Available from: http://www.greeningim.org.uk.
8. IBM. *Information lifecycle management solutions*. [Last Accessed 01/09/09]; Available from: http://www-03.ibm.com/systems/storage/solutions/ilm/index.html.
9. JISC InfoNet (2007) *Managing the Information Lifecycle Management infoKit*. [Last Accessed 01/09/09]; Available from: http://www.jiscinfonet.ac.uk/infokits/information-lifecycle.
10. DCC, *DCC Curation Lifecycle Model.*
11. DCC. *Digital Curation Centre*. 2007 [Last Accessed 02/12/09]; Available from: http://www.dcc.ac.uk/.
12. Moon, M., *Activity lifecycle of digital assets. Journal of Digital Asset Management*, 2007. **3**(3): p. 112-115.
13. University of Edinburgh. *Records Management*. [Last Accessed 01/09/09]; Available from: http://www.recordsmanagement.ed.ac.uk/InfoStaff/RMstaff/records_management_for_staff.htm.
14. DCC. *Data Audit Framework (DAF)*. [Last Accessed 01/09/09]; Available from: http://www.dcc.ac.uk/tools/daf/.
15. DCC *Data Audit Framework (DAF)*. [Last Accessed 01/09/09]; Available from: http://www.dcc.ac.uk/tools/daf/.
16. DCC and DPE. *Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)* 2009 [Last Accessed 01/09/09]; Available from: http://www.dcc.ac.uk/tools/drambora/.
17. RIN (Research Information Network), *Stewardship of digital research data:a framework of principles and guidelines*. 2008, Research Information Network. p. 1-16.
18. Marks, S., *The importance of info stewardship.* Network World, 2006. **23**(23): p. 29.
19. JISC InfoNet *Managing Information to Make Life Easier [A practical Guide for Administrators]*. 1-29 [Last Accessed 01/09/09]; Available from: http://www.jiscinfonet.ac.uk/records-management/guide-for-administrators.
20. Work-with-IT. *National Rural Knowledge Exchange*. JISC Study into the Evolution of Working Practices: Case Studies 2008 [Last Accessed 01/09/09]; Available from: http://ewds.strath.ac.uk/work-with-it/CaseStudies/KnowledgeExchangeacrossruralborders.aspx.
21. UK Government, *Freedom of Information Act 2000: 2000 Chapter 36*. 2000, The Stationery Office Limited.

22. UK Government, *Data Protection Act 1998*. 1998, The Stationery Office Limited.
23. Weedon, R., *Stakeholder interview with JISC Legal (Strathclyde University)*. 2009.
24. Art St. George and the 2007 EDUCAUSE Evolving Technologies Committee, *Imagining Tomorrow's Future Today.* EDUCAUSE Review, 2007. **42**(Number 6).
25. IBM, *"Green" Information Management: The Role of Enterprise Content Management in reducing reliance on paper.* 2008.
26. Romenska, S., *The Learning technologies of the future: technologies that learn?*, in *ALT-C 2009*. 2009: Manchester.
27. Olson, E.G., *Creating an enterprise-level "green" strategy.* Journal of business strategy, 2008. **29**(2): p. 22-30.
28. McDonald, D., D. Cullen, and A. Comrie, *Technology-enhanced working: Advice and Guidance for Staff Development and Change Managers*, in *JISC Study into the Evolution of Working Practices*. 2008, University of Strathclyde.
29. CILIP. *CILIP - Policy and Advocacy - Information literacy: definition*. 2003 [Last Accessed; Available from: http://www.cilip.org.uk/get-involved/advocacy/learning/information-literacy/Pages/definition.aspx.
30. Backhouse, G., *Doing Enterprise Architecture*. 2009, JISC Techwatch.
31. Awre, C., *WEBSITE-INFO-MGT Digest - 28 Aug 2009 to 1 Sep 2009 (#2009-70)*, D. McDonald, Editor. 2009.
32. University of Strathclyde. *Greening Information Management Case Studies*. 2009 [Last Accessed 30/10/09]; Available from: http://www.greeningim.org.uk/CaseStudies.aspx.
33. *Impact Calculator*. Records and Information Management 2009 [Last Accessed 06/11/09]; Available from: http://www.jiscinfonet.ac.uk/records-management/measuring-impact/impact-calculator/index_html.
34. Nicol, D., M. Coen, and C. Breslin, *INSIGHT : A Framework for Evaluating the Costs and Benefits of ICT Investments in Teaching and Learning*. 2004, JISC/University of Strathclyde.
35. University of Strathclyde, *BIILS: Benefits of ICT Investment - Landscape Study: An Evaluation Framework and Toolkit*. 2008.
36. SusteIT, *ICT Energy and Carbon Footprinting tool*. 2009.
37. Coen, M., Cullen, D., Breslin, C., Kelly, U., *A Framework for Information Systems Management and Governance (ISMG) - Self Assesment Toolkit*. 2007, University of Strathclyde. p. 1-36.
38. CDLR. *Mandate - Managing Digital Assets In Tertiary Education Toolkit*. 2006 [Last Accessed 02/12/09]; Available from: http://mandate.cdlr.strath.ac.uk/.
39. infoNet, J. *JISC infoNet - Impact Calculator*. 2009 [Last Accessed 02/12/09]; Available from: http://www.jiscinfonet.ac.uk/records-management/measuring-impact/impact-calculator.
40. JISC InfoNet. *A common framework for measuring the impact of records management*. 2009 [Last Accessed 01/09/09]; Available from: http://www.jiscinfonet.ac.uk/records-management/measuring-impact/.
41. JISC InfoNet, *Impact Calculator Guidance Notes*. 2009.
42. *PEKin: Preservation Exemplar at King's*. 2009 [Last Accessed 27/11/09]; Available from: http://www.kcl.ac.uk/iss/cerch/projects/portfolio/pekin.html.
43. Hein, T., *BEING GREEN Comes Naturally.* Baseline, 2008. **91**: p. p40-42.
44. UCISA (2009) *Cost of Downtime*. [Last Accessed 09/11/11]; Available from: http://www.ucisa.ac.uk/~/media/groups/exec/docs/Cost_of_IT_failure_%20ITT_final%20pdf.ashx.

45. JISC InfoNet (2007) *Records Management infoKit*. [Last Accessed 01/09/09]; Available from: http://www.jiscinfonet.ac.uk/infokits/records-management.
46. JISCmail. *JISC-Repositories Archives*. 2009 [Last Accessed 30/09/09]; Available from: https://www.jiscmail.ac.uk/cgi-bin/webadmin?A1=ind0909&L=JISC-REPOSITORIES.
47. JISC. *Digital Preservation and Asset Management Programme*. [Last Accessed 02/12/09]; Available from: http://www.jisc.ac.uk/whatwedo/programmes/preservation.aspx.
48. DCC. *UK LOCKSS Alliance*. 2006 [Last Accessed 02/12/09]; Available from: http://www.dcc.ac.uk/lockss/.
49. Beagrie, N. (2002) *The Continuing Access and Digital Preservation Strategy for the Joint Information Systems Committee (JISC) 2002-5*. [Last Accessed 02/12/09]; Available from: http://www.jisc.ac.uk/index.cfm?name=pres_continuing.
50. Davis, R.M. *ArchivePress*. 2009 [Last Accessed 02/12/09]; Available from: http://jiscpowr.jiscinvolve.org/category/challenges/.
51. *WordPress*. [Last Accessed 02/12/09]; Available from: http://wordpress.org/.
52. University of Strathclyde. *Work-with-IT: JISC Study into the Evolution of Working Practices*. 2009 [Last Accessed 30/10/09]; Available from: http://ewds.strath.ac.uk/work-with-it.
53. *British Library Digital Preservation Strategy*. [Last Accessed 06/11/09]; Available from: http://www.bl.uk/aboutus/stratpolprog/ccare/introduction/digital/digpresstrat.pdf.
54. Kelly, B. *JISC - PoWR: Preservation of Web Resources*. 2009 [Last Accessed 06/11/09]; Available from: <http://jiscpowr.jiscinvolve.org/>.
55. *Web Archiving Programme*. [Last Accessed 06/11/09]; Available from: http://www.bl.uk/aboutus/stratpolprog/digi/webarch/index.html.
56. *LIFE*^{3}. 2009 [Last Accessed 06/11/09]; Available from: http://www.life.ac.uk/3/.