

Assessing the Preservation of Institutional Repositories with DRAMBORA: Case Studies from the University of Glasgow

by Perla Innocenti, Giuseppina Vullo

Introduction

The Digital Curation Centre (DCC, <http://www.dcc.ac.uk/>) in the UK and the EU-funded Digital Preservation Europe (DPE, <http://www.digitalpreservationeurope.eu/>), project jointly released the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA, <http://www.repositoryaudit.eu/>) in early 2007, with the goal to provide a practical, evidence-based toolkit for assessing repositories and digital libraries. Subsequent iterative development has led to the refinement of its methodology, and the release of DRAMBORA Interactive, a freely available online tool aimed at streamlining the core risk assessment process. DRAMBORA represents a bottom-up approach that takes risk and risk management as its principle means for determining digital repositories' success and for charting their improvement. The tool's development and ongoing evolution has been informed at all times by practical research. Almost twenty international repositories have been subject to assessment using DRAMBORA, enabling the validation of its primary methodology and offering insights into potential shortcomings and the extent of its applicability in a range of diverse preservation contexts. Furthermore, these exercises have enabled initial research into repository profiling, which attempts to identify commonalities within subsets of the repository community in order to inform and facilitate subsequent repository development and evaluation. There are currently more than 130 repository profiles activated with DRAMBORA interactive, ranging from a variety of institutions and countries, from the United States to Europe, Asia and Australasia. This paper describes the DRAMBORA methodology, focusing on its benefits and developments, and introduces DRAMBORA Interactive. It goes on to describe the audit results of two case studies at the University of Glasgow, chosen to compare the assessment within a long-term perspective of a growing institutional digital repository (Enlighten, the

PERLA INNOCENTI is Preservation Research Fellow at HATII at the University of Glasgow. She is Co-Principal Investigator in the EU-funded projects Sustaining Heritage Access through Multivalent Archiving (SHAMAN) and Digital Library Interoperability, Best Practices and Modelling Foundations (DL.org). She is also involved in repository design, audit research as part of Digital Preservation Europe (DPE) and Digital Curation Center (DCC), coordinating activities and development for the Digital Repository Audit Method Base on Risk Assessment (DRAMBORA) Toolkit.

GIUSEPPINA VULLO is Preservation Research Fellow at HATII at the University of Glasgow. She is currently working in the EU-funded Digital Library Interoperability, Best Practices & Modelling Foundations (DL.org), where she coordinates the Working Group on Quality for Digital Libraries. She has been a DPEX fellow at HATII in 2008, where she worked on digital collections assessment, applying DRAMBORA and InterPARES 3.

University of Glasgow Institutional Repository Service, <http://www.lib.gla.ac.uk/enlighten/>) and a closed digital collection (The Correspondence of James McNeill Whistler online edition, <http://www.whistler.arts.gla.ac.uk/>).

1. DRAMBORA. Description of the methodology

1.1 The landscape of digital repositories assessment criteria

The contemporary domain landscape suggests that information repositories are likely to play a role of considerable importance in the pursuit of digital preservation assurances.

In order to legitimise decentralisation to smaller scale repository environments, it is essential that the community has appropriate mechanisms available to support repository assessment, and determine the competencies of those charged with information stewardship responsibilities. Management, staff, financiers and partners must all be satisfied that their efforts are capable of meeting formal expectations. Similarly, information creators, depositors and consumers naturally hope to obtain similar assurances of the capabilities of the organisations providing maintenance, preservation and dissemination services.

Considerable work has been undertaken to develop preservation audit checklists, intended to represent the objective benchmarks against which repositories' efforts are judged. The two primary examples, both released in 2007, are:

- 1) The Trustworthy Repositories Audit and Certification (TRAC) Criteria and Checklist¹ describes approximately ninety characteristics that repositories that aspire to a certifiable, trustworthy status must demonstrate they have;
- 2) The nestor Catalogue of Criteria for Trusted Digital Repositories² reflects the regional needs of the nestor community. Structured similarly to the TRAC document, this provides examples and perspectives that are more representative of a German operational, legal and economic context.

Both TRAC and nestor are compelling reference materials, and their usefulness in informing the development and retrospective evaluation of repositories is widely acknowledged. However, neither is sufficient in isolation. By their very nature, check-lists like these adopt a top-down assessment philosophy: both examples seeking to define an objective consensus of the priorities and responsibilities that should exist within any repository environment. By relying solely on nestor or TRAC, one implicitly disregards the great variety that is visible across contemporary digital repository platforms. The question persists, is a one-size-fits-all approach to assessment and certification really useful for those within the curation community? Both TRAC and nestor's criteria have been painstakingly phrased to ensure their flexibility, and facilitate optimal general applicability. But despite such efforts, it appears evident that within the community there is the need for a more tailored assessment solution that takes into account atypical repository qualities, as either a companion piece, or alternative, to the other existing guidelines.

The Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)³ developed by the Digital Curation Centre and Digital Preservation Europe is designed

1 TRAC was developed by a consortium jointly overseen by the US National Archives and Records Administration and the Research Libraries Group (prior to its absorption within OCLC), and is now maintained by the Center for Research Libraries. Available: <<http://www.crl.edu/PDF/trac.pdf>>.

2 The nestor Catalogue developed in Germany by the Network of expertise in Digital long-term preservation. Available: <<http://edoc.hu-berlin.de/series/nestor-materialien/8/PDF/8.pdf>>.

3 DRAMBORA toolkit, DRAMBORA Interactive and more information on ongoing activities are available at <<http://www.repositoryaudit.eu/>>. See also Andrew McHugh – Raivo Ruusalepp – Seamus

to address such shortcomings. Its bottom-up approach enables repositories to relate their benchmarks for success more explicitly to their own aims and contextual environment, enabling an increased granularity of understanding of preservation approaches and challenges. Furthermore, by focusing explicitly on the process of assessment, rather than simply listing desirable repository characteristics, it provides considerably more opportunities for evidence-supported, demonstrable excellence, and consequent repository confidence. A key strength is that DRAMBORA is capable of being used both independently and in association with more objective guidelines.

1.2 DRAMBORA opportunities and outcomes

Digital curation can be characterized as a process of transforming controllable and uncontrollable uncertainties into a framework of manageable risks. The DRAMBORA process focuses on risks, and their classification and evaluation according to individual repositories' activities, assets and contextual constraints. The methodological outcome is a determination of the repository's ability to contain and avoid the risks that threaten its ability to receive, curate and provide access to authentic and contextually, syntactically and semantically understandable digital information.

DRAMBORA acknowledges the heterogeneity that exists within the digital world, refraining from explicitly describing the characteristics that repositories should demonstrate. Instead, parameters for success are aligned with the subjective mandate, objectives and activities of individual repositories. Specific contextual factors and constraints are considered only where they are relevant. This ensures that the results of the audit process are, from the participating repository's perspective, wholly applicable and immediately useful. The process aims to provide repositories with formal understanding of their own mandate and objectives, to provide them with a detailed and manageable breakdown of fundamental challenges, promote communication within the organisation as a whole and facilitate subsequent external audit whether based on TRAC, nestor or any other repository assessment criteria.

1.3 Origins and alignment with international initiatives

In 2006 and early 2007 the Digital Curation Centre (DCC) undertook a series of pilot audits in a diverse range of preservation environments. Various repositories participated, exhibiting a range of different characteristics.⁴ As well as providing the participating organisations with an objective and expert insight into the effectiveness of their operation, and determining the robustness and global applicability of those metrics and criteria already conceived⁵, the audits were aimed at exploring the optimal means for conducting assessment of repositories. The research set out to develop an increased understanding of how evidence can be practically accumulated, assessed, used and

Ross – Hans Hofman, *Digital Repository Audit Method Based on Risk Assessment*, Digital Curation Centre (DCC) and DigitalPreservationEurope (DPE), 2007, <<http://www.repositoryaudit.eu/>>.

⁴ The British Atmospheric Data Centre (BADC); the National Digital Archive of Datasets (NDAD); the National Library of New Zealand's National Digital Heritage Archive (NDHA); the Florida Digital Archive (FDA) at the Florida Centre for Library Automation; and the Beazley Archive (BA) at the University of Oxford were among those that kindly agreed to take part. For a mechanisms to roll out audit and certification services for digital repositories in the United Kingdom, see Seamus Ross – Andrew McHugh, *The role of evidence in establishing trust in repositories*, «D-Lib Magazine», 12 (2006), n. 7/8, <<http://www.dlib.org/dlib/july06/ross/o7ross.html>>.

⁵ Most notably, at that stage, the RLG/NARA Draft Audit Check-list for Certification of Digital Repositories, which would latterly be published as TRAC.

discarded throughout the audit process. A methodology for performing repository audit was quickly established and subjected to considerable subsequent refinement. In March 2007 the process was formalised as the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA), and a first textual version of the toolkit was released.

Important consensus about the breadth of repository characteristics that must be exposed to scrutiny during an assessment process was reached during a meeting of the authors of DRAMBORA, TRAC and nester in early 2007. Adopting a broad view that echoed the work done by RLG/OCLC in their seminal 2002 *Trusted digital repositories – attributes and responsibilities*, ten general principles of repositories were conceived. The ten principles⁶ are varied, encompassing more than simply technological considerations, extending to organisational fitness, legal and regulatory legitimacy, appropriate policy infrastructures, mandate and commitment, and every aspect of object management, including ingest, preservation, documentation and dissemination. For DRAMBORA's purposes, these can be conveniently grouped according to three core criteria classifications, each influenced by contextual factors and exposed to risk, as illustrated in Figure 1.

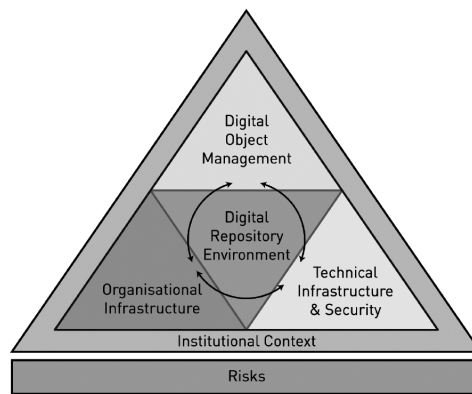


Figure 1: interrelationships within a digital repository environment
© HATII at the University of Glasgow

1.4 Methodology

DRAMBORA's approach is flexible, and responsive to the structural and contextual variety evident within textual and audiovisual repositories: its metric for success is directly linked with repositories' own aims.

Evidence and demonstrable success are at the very forefront of the DRAMBORA process. The first phase of assessment reflects this, a process of information accumulation, aggregation and documentation. The repository's strategic purpose, its action plan, and any contextual factors that influence or limit its ability to meet its objectives must each be made explicit. A hierarchical analysis is undertaken; definition of the repository's mandate is the first step of an increasingly focused scrutiny, requiring detailed descriptions of fundamental repository objectives as well as the activities intended to ensure their successful achievement. The outcome

⁶ *Core requirements for digital archives*, Center for Research Libraries (CRL), 2007, <<http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=92>>.

of this phase is a comprehensive organisational overview, which immediately leads into the latter phase, concerned with the identification of risk.

The issue of risk has been considered from a number of perspectives within the context of digital curation and preservation. For instance, a variety of work has sought to analyze the risks associated with particular file formats, perceiving the risk as something intrinsic to what a digital repository does, based upon the technical challenges associated with maintaining the usability of digital files and storage media.⁷

More recently some authors, such as Ross⁸ and Ross and McHugh,⁹ have described the inherent uncertainty associated with digital preservation.

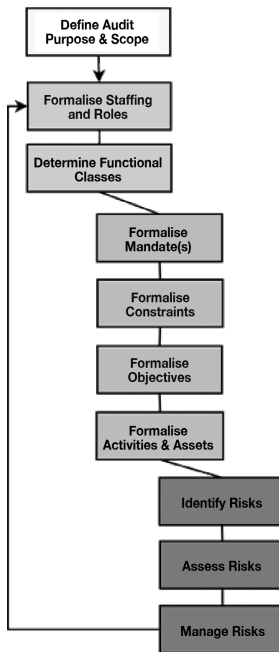


Figure 2: DRAMBORA audit workflow

⁷ Risk Communication Tool, ERPANET, 2003, <<http://www.erpanet.org/guidance/docs/ERPANETRiskTool.pdf>>. Cornell University Library Virtual Remote Control (VRC) tool, Cornell University, <<http://irisresearch.library.cornell.edu/VRC/methods.html>>. JISC, Managing Risk: a Model Business Preservation Strategy for Corporate Digital Assets, 2005, <http://www.jisc.ac.uk/whatwedo/programmes/programme_preservation/programme_404/project/managingrisk.aspx>. Gregory W. Lawrence – William R. Kehoe – Oya Y. Rieger – William H. Walters – Anne R. Kenney, *Risk management of digital information: a file format investigation*, CLIR Report no. 93, 2000, <<http://www.clir.org/pubs/reports/pub93/pub93.pdf>>. Victoria L. Lemieux, *Managing risks for records and information*, ARMA International, 2004. Nancy Y. McGovern – Anne R. Kenney – Richard Entlich – William R. Kehoe – Ellie Buckley, *Virtual remote control: building a preservation risk management toolbox for web resources*, «D-Lib Magazine», 10 (2004), n. 4, <<http://www.dlib.org/dlib/aprilo4/mcgovern/04mcgovern.html>>.

⁸ Seamus Ross, *Uncertainty, risk, trust and digital persistency*, NHPRC Electronic Records Research Fellowships' Symposium Lecture, University of North Carolina at Chapel Hill, 2006.

⁹ Seamus Ross – Andrew McHugh, *The role of evidence in establishing trust in repositories*, «D-Lib Magazine», 12 (2006), n. 7/8 (also published in «Archivi e computer», August 2006, <<http://www.dlib.org/dlib/july06/ross/07ross.html>>).

The risk identification, assessment and management part of the DRAMBORA process is where conclusions are derived from the organisational picture conceived within the first phase. Risk is utilised as a convenient means for comprehending repository success – those repositories most capable of demonstrating the adequacy of their risk management are those that can have, and engender, greater confidence in the adequacy of their efforts. Preservation is after all, at its very heart, a risk management process. The fundamental temporal challenges of preservation are naturally complicated by future uncertainties. Threats relating to any number of social, semantic and technological factors are capable of inhibiting long term access to digital materials.

1.6 DRAMBORA Interactive

In early April 2008, in response to usability issues associated with an entirely paper-based approach, a second version of the toolkit was released as DRAMBORA Interactive, a freely available web based tool (Fig. 3).¹⁰ DRAMBORA interactive leads auditors through the individual stages of the assessment process, recording and displaying responses and providing greater structure to facilitate a more comprehensive coverage. The tool provides robust security provisions, supporting multiple repository contributors, but protecting potentially sensitive information from non-authorised access.

The tool's implicit workflow exactly reflects the core DRAMBORA methodology. In addition, characteristics of each registered repository can be described in detailed terms, with technological, organisational and resource related issues made explicit. This facilitates the intelligent comparison of objectives, challenges and risks with those of peer repositories, again, intended to maximize the assessments' breadth of coverage. The tool is equipped with numerous reporting mechanisms to visualize the repository's status, and support the improvement planning process.

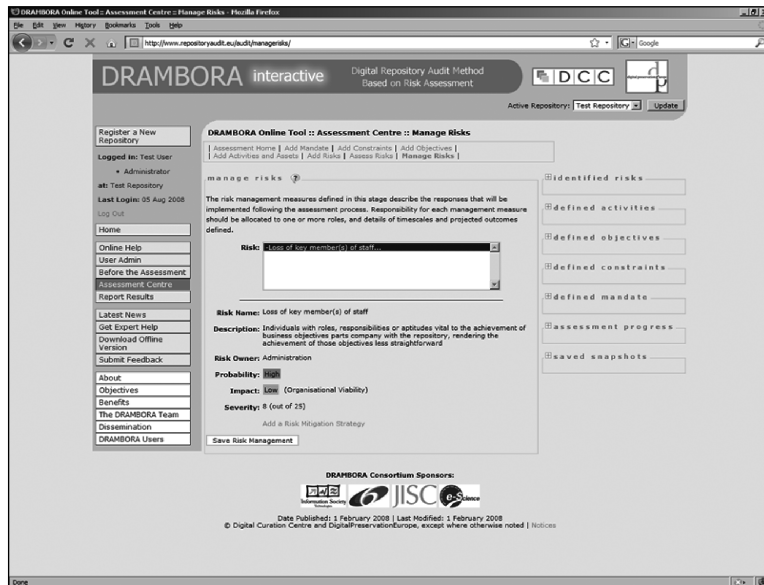


Figure 3: DRAMBORA Interactive interface: Risk management section

¹⁰ <http://www.repositoryaudit.eu/>.

1.7 DRAMBORA at the University of Glasgow

Evaluation of digital libraries is a rapidly developing research area. There is currently no unique methodology encompassing all digital library's evaluation aspects. There are mainly two reasons for this: digital libraries are complex systems and require interdisciplinary approaches for their investigation; digital libraries young history means that a temporal analysis of their evolution is quite challenging.

The DRAMBORA methodology was applied within the DPEX (Digital Preservation Europe Exchange Programme),¹¹ at the Humanities Advanced Technology and Information Institute (HATII)¹² at the University of Glasgow in May-June 2008. The toolkit was chosen to analyse how repositories and collections respond to a digital preservation assessment that involves all their organisational and functional aspects. The evaluation has been conducted to investigate the differences between a growing evolving repository and a completed online collection, including their shared characteristics and challenges. In a digital environment, in fact, the boundaries between silos, archives and repositories are more blurred than in the analogue environment. Moreover, digital preservation involves any kind of digital archive – even those that are hidden behind the simplest resource such as a small digital collection – because any collection has a repository behind it. These issues have been held in common in the following two case studies:

1. *Enlighten, the University of Glasgow Institutional Repository Service*¹³

This project was chosen in order to evaluate an active digital repository that offers born-digital research outputs, bibliographical references and full texts. The digital preservation issues are evident as there's no analogue form of these records. Enlighten is entirely managed by Glasgow University Library staff and it has been promoted during its first years because of its aims, related to the Open Access movement.

2. *The Correspondence of James McNeill Whistler, online edition, University of Glasgow 2003-2007*¹⁴.

This project was chosen to test the assessment of the digital repository behind a digital archive. The electronic edition has been considered as the primary historical source, constituted of original structures and digitised materials. The scope of the audit was to understand whether and how the historical and the scientific value of this source is going to be guaranteed in the long-term. The online edition includes the letters owned by the Glasgow University Library as well as other ones owned by

11 <<http://www.digitalpreservationeurope.eu/exchange/>>. Giuseppina Vullo's activities within DPEX at HATII also focused on the InterPARES 3 assessment methodology application and on the comparison between the two methodologies.

12 <<http://www.hatii.arts.gla.ac.uk/>>.

13 Enlighten, <<http://www.lib.gla.ac.uk/enlighten/>>, is the University of Glasgow's institutional repository service and includes published material by members of the University of Glasgow, including peer-reviewed journal articles, published conference papers, books and book chapters.

14 Based on records of over 13,000 letters, the online centenary edition <<http://www.whistler.arts.gla.ac.uk/>> makes available a total of some 10,000 letters covering the period 1855-1903. The edition includes all the letters written by Whistler and all the letters written to him; letters mentioning Whistler are also included on a selective basis.

other institutions (Library of Congress, New York Public Library, Freer Gallery of Art) and private persons. Materials currently available are dated 1855-1903.

What follows in the next sections is an overview of the audit results of both institutions.

2. Enlighten

2.1 The organisation

The University of Glasgow is one of the UK's largest universities with almost 16,000 undergraduate, 4000 postgraduate students and over 5000 staff. Its mission is to undertake leading-edge, internationally-competitive research while offering a challenging student-centred learning environment, sustaining and adding value to Scottish culture and society, to the natural environment and to the national economy. Glasgow University Library holdings include 2 million books and journals, covering a large range of subjects.

Launched in April 2006, Enlighten (Fig. 4) is based on the work done by DAEDALUS project,¹⁵ funded by JISC from 2002 to 2005, whose goal was to explore an institutional repository model using different software (EPrints, DSpace and PKP Harvester) for different content, such as published and peer reviewed papers, pre-prints, grey literature and theses.

Since the publication of the Glasgow University Statement on Open Access,¹⁶ changes in policy by both funders and publishers have significantly increased the free availability of full text material. Many publishers now allow the full text final versions of journal articles to be deposited in repositories and many funding bodies require a full text copy of research to be made freely available. The University believes the time is now right to adopt a similar policy requiring authors to deposit the full text of their publications in Enlighten when they are able to do so.

The usage of EPrints has dramatically increased in the last years: within the context of about 1500 fulltext, the usage statistics have shown the downloads increased from 150.000 in April 2006 to over 946.000 two years later.

In accepting to undertake the DRAMBORA assessment, Glasgow University Library provided a large amount of technical documentation. Preparatory online and telephone communications with the staff anticipated the audit, and showed a constant interest in taking part in the project. Before the audit, a DRAMBORA 2.0 interactive online profile was created. The assessment took place in two working days. The staff was well organized and efficient. They offered a comfortable space to conduct the audit, provided extra information and actively discussed the assessment findings and suggestions. Since the second day, the staff of Enlighten recognized that the DRAMBORA audit process was yielding numerous benefits, and providing insights that would undoubtedly prompt further investigation and probable response from stakeholders.

Following the identification of risks, a significant part of the time spent on site during the audit was committed to risk assessment; for each risk, repository staff discussed the severity of the threat and provided probability scores.

15 <<http://www.lib.gla.ac.uk/daedalus/>>.

16 <<http://www.lib.gla.ac.uk/enlighten/statement.html>>.

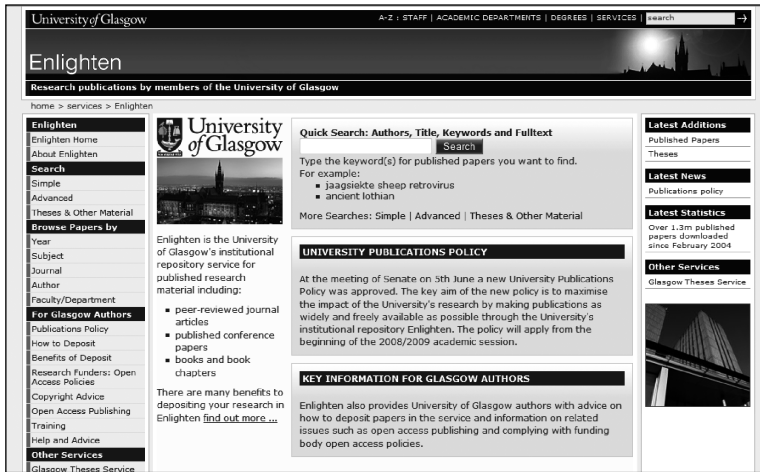


Figure 4: Enlighten homepage

2.2 Collections in Enlighten

Enlighten is the University of Glasgow service, encompassing three digital repositories for different types of material:

- *EPrints*¹⁷ is the repository for published and peer-reviewed papers, i.e. peer reviewed journal articles, books, book chapters, conference proceedings, lecture notes and thesis;
- *DSpace*¹⁸ is the repository for digital content including working papers, technical reports, theses and pre-prints by members of the University of Glasgow for other research material;
- *Theses*¹⁹ is the repository for a collection of full text higher degree theses successfully defended at the University of Glasgow.

The public source for the mandate of the repository is available on the DAEDALUS project website,²⁰ but is not explicit in Enlighten (whose slogan is “Freeing Research at the University of Glasgow”).

2.3 Preservation planning and activities

Preservation planning responsibilities are currently not explicitly described within any job description. Decisions on preferred ingest file format (PDF) and the storage solutions have been made collectively as a result of the DAEDALUS project. The Enlighten team is involved in SHERPA DP²¹ and PRESERV.²² Digital preservation needs of the repository are under investigation and some strategy options will be developed depending on the resources. Adequate back-up and recovery measures are in place to ensure that the repositories do not suffer significant failures.

17 <<http://EPrints.gla.ac.uk>>.

18 <<https://dspace.gla.ac.uk>>.

19 <<http://theses.gla.ac.uk>>.

20 <<http://www.lib.gla.ac.uk/daedalus/overview.html>>.

21 <<http://ahds.ac.uk/about/projects/sherpa-dp/index.html>>.

22 <<http://preserv.eprints.org/>>.

2.4 Technical infrastructure

Within Enlighten, the service development side of DAEDALUS was initially going to use only two pieces of repository software, *EPrints* and *ETD-db* for these. *DSpace* was added as a third choice after the project started. During the course of the project only *EPrints* and *DSpace* were left, because it was considered feasible to support two pieces of software rather than three. *EPrints* and *DSpace* provided complementary features which were suited to the different content policies being adopted for published papers and for *preEPrints* and working papers. Both *DSpace* and *EPrints* were installed on a single server. Since the launch of Enlighten, the repositories have been moved to a new server, now running on Linux.²³

2.5 Audit scope and process

Enlighten has been audited by Perla Innocenti and Giuseppina Vullo from HATII on 2nd and 3rd June 2008.

The audit interviews were held with staff members involved in the repository management:

- Susan Ashworth, Subject Librarian, Research & Learning Support Servicesearch & Learning Support Services;
- Morag Greig, Subject librarian, Advocacy Manager for Enlighten, E-Theses co-ordinator;
- William J. Nixon, Deputy Head of Library Information Systems.

At the end of the audit, a risk registry was drafted upon the audit results and discussed further with the staff of Enlighten.

2.6 Audit findings

2.6.1 Organisational commitment

The general aim of Enlighten is to make University of Glasgow research materials freely available to the general public. As for the DAEDALUS project, its objectives include:

- establishment and population of a range of OAI-compliant Digital Collections at the University of Glasgow using a range of different OAI-compliant pieces of software;
- role of catalyst for cultural change and ongoing discussions about the “crisis in Scholarly Communication” within the University of Glasgow and the wider community;
- dissemination of experiences and findings to the wider community.²⁴

The depositors are the university staff and the students. From 2009 there will be a unique database to search into the three repositories and it will be integrated to the online catalogue; all “formal” publications in full text where copyright allows will be collected.

2.6.2 Organisational fitness

Most current materials are purchased from the grant allocated by the University Management Group to the Library and subsequently distributed to faculties. As the JISC funding drew to a close, the project staff also spent some time evaluating policy decisions that had been taken during the course of the project. A re-assessment of policies relating to deposit mechanisms and to the question of permitting metadata only records in the repository was carried out. A key step in the transition from project to service was securing the necessary funds to appoint staff within the library to carry out repository work as an established part of their jobs. Two library assistants, one

²³ <http://EPrints.gla.ac.uk/3718/01/Enlighten_oclc_article_.pdf>.

²⁴ *Ibid.*

with supervisory duties, were appointed. These posts are based within the library's bibliographic services department. Currently the staff dealing with Enlighten is composed of six people working part time. Library staff is regularly involved in training sessions and meet regularly, keeping periodical contacts with the computer department.

2.6.3 Legal and regulatory legitimacy

Glasgow University, as a signatory of the Scottish Open Access Declaration, strongly encourages authors at Glasgow University to deposit copies of their published work into the University's Institutional Repository.²⁵ Authors clearly indicated that copyright was a major issue of concern, and that they were dependant on project staff checking and clearing copyright for them. The project staff therefore spent a good deal of time contacting publishers to seek permission for papers to be added to the repository. In some cases this effort provided a useful dialogue. The key difficulty was that there were no existing guidelines or legal precedent, and no expert within the repository staff to turn to for advice. This experience allowed the Enlighten staff to develop an expertise in interpreting copyright agreements and dealing with publishers. It is clear that further advocacy is required to alleviate concerns related to copyright, as the lack of knowledge about the issue is preventing some authors from depositing in a repository.²⁶

2.6.4 Policy framework

The open access policy was encapsulated in the form of a statement issued by the University's Vice-Principal for Research, which was circulated by e-mail to all staff. Additional publicity was secured in the form of an article in the University's internal newsletter. The adoption of this policy was a significant achievement, and much of this can be attributed to the ongoing efforts of senior library staff over the past few years to inform and educate University management about open access and to persuade them of the merits of an institutional repository. The constant repetition of the OA message over a prolonged period of time has had an important role within the success of the service.²⁷ Enlighten repository staff also produced a number of internal operational and management policies. The vast bulk of the workflow is still mediated by staff and there is no automatic bulk ingest, although in the past there have been some bulk ingest of metadata.

2.6.5 Acquisition and ingest

The initial focus for Enlighten is the Glasgow EPrints Service, published and peer-reviewed articles, conference papers and books. Authors have two deposit options: self-deposit and mediated deposit. Self-deposit requires authors to provide bibliographic details for their publication along with any relevant information relating to copyright. They are then required to upload the full text of their publication. Repository staff check the metadata and complete a number of additional fields before the record is made publicly available. If there are any problems relating to copyright or the version of the publication that has been added, the author is contacted. The second option currently offered is the mediated deposit. Authors are asked to send an e-mail providing the basic bibliographic details for their publication along with a suitable version of the full text. Repository staff then creates a record on behalf of the author. Before depositing the thesis the students must solve the copyright issues,

25 <<http://www.lib.gla.ac.uk/enlighten/statement.html>>.

26 <http://EPrints.gla.ac.uk/3718/01/Enlighten_oclc_article_.pdf>.

27 *Ibid.*

deciding if to restrict or not access to the thesis; completing a Thesis Access Declaration form.²⁸ This must be signed by a supervisor showing the agreement with the decision of making the thesis publicly available and converting the final version of the thesis to a single PDF file. Repository staff checks the information provided and add further details.

2.6.6 Maintenance of digital objects

The server is stored in the Computer Service department and maintenance is provided by the same department and include filestore for *EPrints*, *Theses* and *DSpace* as well as the operating system. Glasgow University Library has recently purchased a Sun 4150. No virus checking is performed in the ingest phase, but so far there have been no problems. Backups are done using Legato system for the Library catalogue, and tapes for *EPrints*.

2.6.7 Metadata management

Two library assistants are currently working on Enlighten metadata management. Effort varies according to the level enrichment of ingested document and to eventual copyright issues. Personal names are checked and modified in standardized formats, although an authority file is not yet in use.

2.6.8 Access and dissemination

Access to Enlighten is freely available. Usage of the publications in the repository has increased dramatically. In April 2006, there had been 130,000 PDF downloads of papers in Enlighten; two years later there have been over 946,000 downloads. These figures clearly show the significant increase in visibility achieved by making scholarly work openly available. Since the start of the 2007/2008 session, postgraduate students have been required to provide an electronic copy of their thesis in addition to one printed copy. Theses are made available in the Glasgow Theses Service, part of Enlighten. There are more than 100 theses available within the Theses Service, thus making available, on a worldwide scale, the postgraduate research being carried out at Glasgow. Only a few theses have been embargoed under the University's agreed policy.

2.6.9 Preservation planning and action

Currently, Enlighten does not have a specific preservation policy. The declared first aim of Enlighten has been to provide access and not to preserve digital documents. The Enlighten team is involved in SHERPA DP and PRESERV; digital preservation needs are considered, and some strategy options will be developed depending on the resources. The disaster planning is that one of the University of Glasgow.

2.6.10 Technical infrastructure and security

Since the launch of Enlighten the repositories have been moved to a new server, now running Linux. The server is stored in the Computer Service department and maintenance is provided by the same department. *E-Prints* is OAI-PMH compliant and features both the harvesting service and the data provider aspects of the protocol.

3 The correspondence of James McNeill Whistler online edition

3.1 The organisation

The Centre for Whistler Studies was established in 1992 for research on the artist James McNeill Whistler (1834-1903). It led to the complete edition of Whistler's correspondence. The Centre

²⁸ <http://theses.gla.ac.uk/GU_thesis_access_declaration.pdf>.

was closed in 2006 after publication of the correspondence. The project was developed by staff working at the Centre for Whistler Studies, part of the Faculty of Arts at the University of Glasgow. It was a constituent member of the University's Institute for Art History and collaborated with other University departments with research interests in Whistler. These include the Hunterian Art Gallery, which houses 1,000 works of art by Whistler, the Department of History of Art, the French Department and the Centre for American Studies. The Whistler correspondence website is maintained as an important archival research resource, and the Faculty wishes to continue to support the importance and significance of the Whistler holdings. They are a core element of the Hunterian's collections, and constitute part of the unique resources that the University must advertise – which it already does as part of the Faculty's Learning and Teaching strategy. While it is a completed project, maintenance and upgrading are ongoing, as are international relationships, particularly those with American institutions which fostered the Centre's projects in the past.

3.2 Records in the James McNeill Whistler Online Edition

Based on records of over 13,000 letters, the on-line centenary edition makes available a total of some 10,000 letters covering the period 1855-1903. The edition includes all the letters written by Whistler and all the letters written to him; letters mentioning Whistler are also included on a selective basis.

The Correspondence of James McNeill Whistler Project²⁹ is a public web database (Fig. 5) which houses information pertaining to all aspects of correspondence of the artist James McNeill Whistler (1834-1903). It includes related tables describing people, works of art, subjects, institutions and places that relate to the correspondence of the artist, providing a rich semantic context. A separate bibliographical database provides additional referencing information. Linked to the database are transcriptions of the correspondence and biographical data. The project also includes a set of low-resolution images of works of art used on the website to accompany data about works referenced in the correspondence.

The screenshot shows the homepage of the 'The Correspondence of James McNeill Whistler' online archive. At the top, there is a navigation bar with links for 'A-Z', 'STAFF', 'ACADEMIC DEPARTMENTS', 'DEGREES', and 'SERVICES', along with a search box. The main header reads 'THE CORRESPONDENCE OF JAMES McNEILL WHISTLER' and includes a small portrait of Whistler. Below the header, a breadcrumb trail indicates the current location: 'Home > Faculty of Arts > History of Art > Whistler Correspondence > On-line Edition'. A left-hand navigation menu lists various options: Home, J. McN. Whistler, On-line Edition, Selected letters, Search for..., People, Letters by date, Subjects, Works of Art, Exhibitions, Institutions, Places, Freetext, Record No, Contacts, Resources, Miscellany, and Links. The main content area is titled 'The On-line Edition, University of Glasgow, 2003-2007' and lists the Project Director as Professor Nigel Thorp. It mentions editors Margaret F. Macdonald, Patricia de Montfort, and Nigel Thorp, and notes the inclusion of letters from Anna McNeill Whistler (1855-1880), edited by Georgia Toutziaris. A section titled 'Currently available: 1855-1903' lists several bullet points: Research Funding, Whistler Collections in Glasgow and Washington D.C., The Edition, Editorial Practice, Abbreviations, Citing the Edition, Staff, and Copyright Notice. Below this is a 'Research Funding' section detailing support from the British Academy, the Arts and Humanities Research Council, the Getty Grant Program, the John Sloan Memorial Trust, and Patricia Cornwell Enterprises. The 'Whistler Collections in Glasgow and Washington D.C.' section mentions the University of Glasgow and the Freer Gallery of Art at the Smithsonian Institution. On the right side, there is a small image of a letter with the caption: 'Letter from J.M. to Charles James Whistler Haason, Glasgow University Library, HSS'.

Figure 5: The Correspondence of James McNeill Whistler online archive homepage

3.3 Preservation planning and activities

The working data set is still in use and the website is regularly updated. It was originally expected that the working data set would be deposited with AHDS (Arts and Humanities Data Service)³⁰ on completion of the main phase of the project which was partly funded by AHRC (Arts and Humanities Research Council). The agreement between AHDS and AHRC having lapsed, the University of Glasgow could claim exemption from this requirement if a suitable facility for archiving the material was available at the University.

3.4 Technical infrastructure

The resource consists of two data sets: a working data set and a website. The relational database application used for the working data set is MS Access 2000. This data set is located with text documents and images (works of art) on a shared network drive at the University of Glasgow. Transcriptions of over 10,000 letters are encoded in SGML/XML and stored in the file system of the network drive. Each transcription is stored as a separate file and each is linked to its record in the main table of the database. All content on the website is updated periodically from the working data set: there is no restriction on access to the website and the resources there are freely available to the public.

3.5 Audit scope and process

The online archive has been audited by Andrew McHugh and Giuseppina Vullo from HATII on 19th and 23rd June 2008. The audit interviews were held with:

- Graeme Cannon, System developer & Learning Support Services;
- Prof. Margaret MacDonald, co-editor.

Prior to the audit visit, they provided with access to a range of articles and presentations describing the activities related to the online archive. Further information are available on the archive website.

At the end of the audit, a risk registry was drafted upon the audit results, and discussed further with the staff of Enlighten.

3.6 Audit findings

3.6.1 Organisational commitment

The project goals are found in the online edition, which has been freely available from 2003. The primary function of the website is to allow users to locate transcriptions of correspondence documents in response to a wide range of criteria used in the search mechanisms available on the site. The database searches and links from annotations in the transcriptions also allow users to locate information on people (including biographies), subjects, artworks, exhibitions, institutions, places and bibliographic data.

3.6.2 Organisational fitness

The edition of Whistler's correspondence has been supported principally by: the University of Glasgow, the Arts and Humanities Research Board, the British Academy Committee on Academy Research Projects (CARP), the Getty Grant Program, the John Sloan Memorial Trust, and the Patricia Cornwell Enterprises. A staff member from HATII is responsible for technical maintenance. An editorial handbook was provided to solve all the doubts about transcriptions, which have been overseen by

30 <<http://ahds.ac.uk/performingarts/>>.

the co-editors during the team editing work. Although the project is finished, there are individual corrections or additions still being made.

3.6.3 Legal and regulatory legitimacy

The online archive letters are protected by copyright and a statement is available online. The University of Glasgow holds the copyright for a large percentage of the data set. Other institutions and individuals also own copyright for some material that has been digitised as part of this project. Before an item is made available it is established whether permission is required from a copyright holder or owner of the document for the transcribed content to appear on the website. If permission is granted the transcription is made available. If permission is refused or the holder cannot be contacted or established the document is not made available.

3.6.4 Policy framework

The policy is traced by several AHDS and Faculty reports when the Centre was active. Now that the project is finished, short term costs for the Whistler Correspondence have been funded for maintaining the network drive, website and backup systems by the British Academy and the University of Glasgow, but a new policy is not yet formalized. There is some quality assurance policy implicit within the editorial handbook issued to each individual responsible for transcribing letters, and manual checking has been relied upon too, with checks undertaken for each record documented within the database. Currently the editors are the only contributors to the content of the project. They are responsible for the manual checking.

3.6.5 Acquisition and ingest

The database has data on over 10,000 documents and it also includes separate tables describing people, works of art, exhibitions, places, subjects, institutions and places that relate to the correspondence. There is also a separate bibliographical database, linked to transcriptions of the correspondence and biographies. The working data set currently is located and stored on a network drive and it can be accessed and developed by editors using VPN. Work on the active data set is primarily concerned with accurately transcribing what are in many cases scarcely legible handwritten documents. Research is carried out on the document texts to correctly identify people, places, works of art, events and institutions referred to in the text and the results of this research is included as annotations. The dating of individual documents is often the result of research carried out on other documents in the data set. It is possible to verify the editorial history of each letter in the database, with named and dated checks of details.

3.6.6 Maintenance of digital objects

The working data set is currently located on a RAID 5 network drive maintained by the Arts Support team in the Faculty of Arts at the University of Glasgow. Arts Support manages backups of this network drive and website and provides network support. Other backups to local hard drives and external backup devices are also carried out by the staff. The XML dataset, MS Access database and images which form the working data set are located on a RAID 5 network drive. In 1997 RTF and Word based transcriptions were re-developed as SGML file fragments that allowed either single documents or large collections of documents to be searched using an SGML browser (Panorama). The design of these SGML structures later allowed the same fragments to be processed as XML files. The SGML/XML content (transcriptions of correspondence and biographies) is validated against a DTD (Document Type

Definition). For display on the website the XML content is transformed to HTML using XSLT (EXtensible Stylesheet Language Transformations).

3.6.7 Metadata management

In light of some duplication, the existing collection now consists of 10,000 unique letters, and 13,000 records. TEI was initially considered as a potential encoding format, but later rejected as the complexity of the header information exceeded the needs of the project. SGML was eventually chosen due to the simplicity of the core markup and interoperability potential. The letters themselves are marked up in Author-Editor format using Softquad's program. Mark up corresponds to a bespoke schema with minimal elements. The collection consists of both XML and SGML marked up documents, but by using a simple AE header template, each marked up letter can be referenced as an entity, and subsequently transformed and deployed as HTML. Most of the information relating to each letter, and facilitating its discovery is recorded within the associated database, and not marked up intrinsically within the file.

3.6.8 Access and dissemination

The delivery mechanism, an Internet accessible website, does not interface directly with the Access database. Instead, a MySQL database that resides on the web server contains a duplicate set of tables to support content discovery within this application. The MySQL database is manually rebuilt on a periodic basis to remain in sync with the master Access database. Much of the website consists of static pages. For discovery, ASP scripts are deployed, querying the MySQL database and generating links to the static HTML pages that correspond to the appropriate XML representation of each letter. The content is made freely available via the website, but each document contains a copyright statement. There are many worldwide users, historians and private individuals, but there is no quantitative information about usage. It will be considered as an option to justify a formalised University financial commitment.

3.6.9 Preservation planning and action

The working data set is the primary resource that needs to be preserved. The website does not include live data and is currently a little dated in terms of style, HTML formatting and methods of data delivery. The Department of Art History, University Library, Hunterian Art Gallery and HATII are developing plans for long-term site maintenance and development. The working data set can currently be archived in file formats that make the data reasonably safe from obsolescence.

3.6.10 Technical infrastructure and security

The technical development commenced in the early 1990s; the new system developer in 1996 consolidated the initial FoxPro and Paradox databases into the Access structure currently in use.

New content into the collection is created with the transcription of individual letters. A Microsoft Access database record is initially created to correspond with each new letter. This has an automatically generated unique identifier, which is simply an incremental numeric value. This value links the database record with associated tables and the marked up files containing the content of each letter. The database contains rich information about each record in a variety of tables, including information about related individuals, places, institutions and artworks. All project members may access this database with no tiered permissions, which is stored on an Arts Faculty network drive. Content is backed up regularly, in accordance with

Arts Faculty network storage policy, and technical contributors periodically recreate content on their own local workstations to facilitate straightforward recovery. Anti-virus checks on the network accessible content are not guaranteed, although most Arts Faculty work stations have an anti-virus installed in order to conform with central University IT recommendations and requirements.

4 Conclusions

As in previous audits, DRAMBORA has revealed itself as a powerful and effective assessment methodology which, starting from a digital preservation evaluation framework, allows a general assessment of the curation and maintenance of digital library aggregated resources. Both case studies presented in this paper are emblematic of two characteristic repositories involved in digital preservation and in quality maintenance: a growing institutional repository (Enlighten) and a closed digital collection (The Correspondence of Whistler online edition). But while Enlighten includes different types of digital content and is a dynamic part of the University Library, the Whistler online edition is an electronic edition of a unique type of documents.

Enlighten represents an important service at the University of Glasgow. The assessment with DRAMBORA helped the repository staff to understand not only the shortcomings of Enlighten, but also its strengths and opportunities for improvement. The audit highlighted the effort invested in actively promoting Enlighten and its OA policy, but also the need for sufficient technical resources for current and future developments, as well as the necessity to rationalise the number of supported platform. Focussing on the creation of more documentation, planning ahead and developing a preservation policy were the key management strategies arising from the assessment. And the prioritization of the risk registry was deemed very useful to communicate with the other library staff and library director.

The Correspondence of James McNeill Whistler online edition is a relevant and large edition of digitised content. Perhaps the most immediate shortcoming of the system is the lack of automation that exists to link each discrete and somewhat static part of the system. It is felt that an integrated and dynamically updated solution would be both easier to manage, and more streamlined. The co-editors and the system developer are aware that this resource needs to be preserved in the long term, and a formal agreement about its technical and legal maintenance is felt to be necessary. This is currently being formalised by the departments involved and the audit results will support the process. The formal preservation agreement will be a starting point for the future maintenance of this high value resource: the Department of Art History, Faculty of Arts, Hunterian Art Gallery, HATII and Special Collections in Glasgow University Library will be the actors involved in curating it and are developing a coordinated policy to share their specific responsibilities.

From a more general view, DRAMBORA has now been deployed in a range of evaluative contexts, and the processes of self assessment and facilitated assessment continue to yield considerable insights into both preservation activities, and the state of preservation assessment. Work associated with DRAMBORA will continue a variety of ways, from training activities to international audits and collaborations. The developers of DRAMBORA have or have had active collaborations with the following international initiatives and projects: Trustworthy Repository Audit and Certification (TRAC) Criteria and Checklist Working Group, Center for Research Libraries (CRL) Certification of Digital Archives Project, Network of Expertise in Long-term storage of Digital Resources (nestor), DELOS Digital Preservation Cluster

(WP6), International Audit and Certification Birds of a Feather Group, the EU-funded project SHAMAN (Sustaining Heritage Access through Multivalent ArchiviNg).

The DCC and DPE are committed to training a generation of DRAMBORA auditors through a number of planned events taking place in 2008 and 2009. Facilitated audits will continue both interactively and through physical visits, with new organisations registering their repositories and completing self-assessments every week. DRAMBORA Interactive was released in early 2008 and the procedure to submit DRAMBORA as the basis of an ISO standard has been initiated (ISO TC46 /SC 11). DPE and the Digital Curation Centre intend to continue to develop DRAMBORA to support the longer term management of repositories and ensuring that they are auditable and continue to develop in ways that enable them to consistently improve their levels of service and the longer term sustainability. They will also support its widest acceptance within the United Kingdom, Europe and broader international contexts.

Acknowledgements

The development of DRAMBORA has been enabled by the European Union through its Framework Six support for Digital Preservation Europe under the IST Programme (IST-2005-2.5-10) through grant IST-2005-034762. The Digital Curation Centre (DCC) in the UK was the other key funder of DRAMBORA development.

We are grateful to the staff of Enlighten and The Correspondence of James McNeill Whistler online edition – and in particular to Susan Ashworth and Prof. Margaret MacDonald - for their active engagement in the DRAMBORA assessment, and their authorisation to publish these overview of the audits.

Valutare la conservazione dei repository istituzionali con DRAMBORA: casi di studio dall'Università di Glasgow

di Perla Innocenti, Giuseppina Vullo

Nel 2007 il Digital Curation Centre (DCC), <<http://www.dcc.ac.uk/>>, organizzazione attiva nella gestione e conservazione delle risorse digitali di istituzioni britanniche, insieme a Digital Preservation Europe (DPE) <<http://www.digitalpreservationeurope.eu/>>, progetto finanziato dalla Commissione Europea nell'ambito del Sesto Programma Quadro, hanno rilasciato DRAMBORA (Digital Repository Audit Method Based on Risk Assessment), un metodo ed un sistema online di valutazione dei depositi digitali basato sulla valutazione del rischio <<http://www.repositoryaudit.eu/>>.

DRAMBORA può essere utilizzato come strumento di autovalutazione da un'ampia gamma di repository, biblioteche e archivi digitali, inclusi quelli il cui mandato non prevede una gestione specifica per la conservazione digitale a lungo termine. Il suo sviluppo ha portato alla realizzazione di DRAMBORA Interactive, un sistema online disponibile gratuitamente in rete a <http://www.repositoryaudit.eu/>, finalizzato ad ottimizzare e a tenere aggiornato il processo valutativo.

Il metodo DRAMBORA adotta un approccio misto *top-down* e *bottom-up* che, partendo dal concetto di rischio e dai fondamenti del *risk management*, identifica il successo del repository digitale e ne traccia il potenziale sviluppo nel tempo, attraverso l'identificazione strutturata dei suoi punti di forza e di debolezza.

Lo sviluppo e l'evoluzione di DRAMBORA sono stati modellati con casi di studio: circa venti depositi digitali in ambito internazionale sono stati oggetto di sperimentazione, e hanno permesso la sua validazione a livello metodologico, offrendo spunti per il suo potenziamento ed estendendo la sua applicabilità in diversi ambiti conservativi. Tali applicazioni hanno inoltre consentito l'avvio di ricerche per la creazione di profili del rischio dei repository digitali, volte a facilitare il confronto tra i repository e basate sull'identificazione di denominatori comuni in vari contesti.

Ad oggi sono attivi in DRAMBORA Interactive i profili di circa centotrenta repository, biblioteche e archivi digitali, provenienti da un ampio ventaglio di istituzioni e paesi (Stati Uniti, Europa, Asia, Australia e Nuova Zelanda).

Nella prima parte, quest'articolo descrive il metodo DRAMBORA, concentrandosi sui benefici derivanti dal suo utilizzo e sui suoi sviluppi, e delinea le caratteristiche di DRAMBORA Interactive.

PERLA INNOCENTI, Preservation Research Fellow at HATII presso l'University of Glasgow.

GIUSEPPINA VULLO, Preservation Research Fellow at HATII presso l'University of Glasgow.

La seconda parte dell'articolo presenta i risultati di due applicazioni di DRAMBORA presso l'Università di Glasgow, Gran Bretagna, con l'intento di mettere a confronto il suo utilizzo per la valutazione di un deposito istituzionale in crescita (Enlighten, the University of Glasgow Institutional Repository Service) e per quella di una collezione digitale chiusa (The Correspondence of James McNeill Whistler online edition): <<http://www.lib.gla.ac.uk/enlighten/>>, <<http://www.whistler.arts.gla.ac.uk/>>.