

Licenciatura em Gestão de Sistemas e Computação
Universidade Atlântica

**Protótipo de uma biblioteca digital numa universidade:
*Suporte ao processo de inicialização do repositório***

Projecto Final de Licenciatura

Elaborado por Luís Manuel L. Pinto da Rocha
Aluno nº 20060916
Orientador: Mestre Filipa Taborda

Barcarena
Julho 2009

O autor é o único responsável pelas ideias expressas neste relatório.

Índice

1	Introdução	9
2	Enquadramento	10
2.1	Conceito de Biblioteca Digital	10
2.2	Aspectos Socioeconómicos das Bibliotecas Digitais	11
2.3	Tecnologia	12
2.4	A Plataforma Greenstone entre outras	13
2.5	Metadados	16
2.6	Outros esquemas de metadados.....	17
2.7	Metadados para Imagens e Multimédia	17
2.8	Acesso	18
2.9	Preservação dos objectos digitais	18
2.10	Propriedade Intelectual.....	19
2.11	Interoperabilidade.....	20
2.12	Iniciativas Relevantes.....	20
2.12.1	Iniciativas Internacionais	20
2.12.2	Iniciativas Nacionais	22
3	Estudo de solução	23
3.1	Metodologia	23
3.2	Arquitectura da solução.....	25
3.3	Pressupostos	27
4	Construção de uma colecção utilizando o protótipo.....	29
4.1	Interface Gráfico do Bibliotecário (GLI)	29
4.2	Recolha dos documentos	30
4.3	Enriquecimento e Metadados	31
4.3.1	O Editor de Metadados (GEMS).....	32
4.3.2	Operação do protótipo.....	33
4.4	Personalização do Interface do Utilizador	36
5	Trabalhos futuros	37
6	Conclusões.....	38
	Bibliografia.....	39

Índice de figuras

Figura 3-1: Ecrã principal do protótipo	26
Figura 3-2 : Exemplo de estrutura XML de metadados criada pelo protótipo	27
Figura 3-3: Formato base do ficheiro <i>metadata.xml</i>	28
Figura 3-4: Estrutura de definição de tipo de documento (DTD).....	28
Figura 4-1: Modos de funcionamento do GLI.....	29
Figura 4-2: Opções de recolha remota de elementos digitais	30
Figura 4-3: Organização dos documentos incluídos na colecção	31
Figura 4-4: Selecção da especificação de metadados	33
Figura 4-5: Estrutura Dublin Core antes do processo de preenchimento	35
Figura 4-6: Campos de Metadados após preenchimento automático	35

“The pen is mighty: Wield it wisely”

In: (Bainbridge & Witten, How to Build a Digital Library, 2003)

Resumo

O presente trabalho tem como objectivo principal criar suporte ao processo de inicialização de uma colecção digital na plataforma Greenstone, apresentando contributos práticos que visam automatizar o procedimento de enriquecimento do repositório.

Com base nos requisitos iniciais e limitações de recursos da instituição foram avaliadas diversas soluções, tendo-se optado por desenvolver uma aplicação externa à plataforma para inicialização do repositório.

Foi objecto de estudo documentação respeitante a directivas relativas à construção de bibliotecas digitais em geral, documentação essa classificada e compilada em colecção digital.

Palavras-chave: Biblioteca digital, Metadados, Protótipo, Enriquecimento, Greenstone

Abstract

This work has as main objective to create support for the process of starting a digital collection using Greenstone platform, making practical contributions that aim to automate the process of enrichment of the repository.

Based on initial requirements and limitations of resources of the institution were evaluated several solutions and have chosen to develop an application external to the platform to support the repository initialization.

Was the subject of study documentation regarding the directives on the construction of digital libraries in general, which has been classified and compiled in a digital repository.

Key-Words: Digital Library, Metadata, Prototype, Enrich, Greenstone

1 Introdução

As bibliotecas digitais representam actualmente um dos meios privilegiados de distribuição de informação e podem facilmente ser encontrados exemplos de aplicação nos mais variados domínios, tais como; na promoção do desenvolvimento humano; no desenvolvimento da divulgação científica; ou na preservação das culturas tradicionais e populares.

O problema que está na origem deste trabalho é muito específico, ocorre na inicialização de um repositório ou biblioteca digital a partir de uma lista de documentos classificados e tem como objectivo principal automatizar um processo moroso e sujeito a erros de digitação.

O estudo de solução, dirigido à plataforma Greenstone e ainda com os metadados indefinidos, envolveu um trabalho de investigação sobre a temática das bibliotecas digitais em geral e suas especificações e a plataforma seleccionada, exigido como pressuposto para a concretização da solução na instituição.

Para ilustração da funcionalidade do protótipo desenvolvido e como contributo válido da compilação bibliográfica realizada, construiu-se uma colecção inicialmente especificada numa lista classificada construída por inspecção directa aos documentos seleccionados.

São apresentados caminhos possíveis para trabalhos futuros nesta temática e as conclusões do projecto.

2 Enquadramento

Neste capítulo pretende-se efectuar um enquadramento geral sobre as bibliotecas digitais, descrever os principais conceitos associados, a sua importância e contribuição no contexto do desenvolvimento de repositórios digitais.

Pretende-se contribuir para o entendimento geral sobre o tema, as suas principais vertentes tecnológicas, socioculturais e económicas, seus problemas derivados, referenciando algumas iniciativas relevantes, pretendendo construir a base de conhecimento considerada necessária para suporte ao desenvolvimento deste protótipo e de outros projectos nesta área.

2.1 Conceito de Biblioteca Digital

A investigação em torno do conceito de biblioteca digital tem vindo a evoluir, desde um início relativamente polarizado entre a comunidade de investigadores, com uma visão estreita e focada na resolução dos seus problemas específicos enfatizando as tecnologias, e a comunidade de bibliotecários mais focada nos aspectos institucionais e de serviço das bibliotecas (Borgman, 1999).

O desenvolvimento tecnológico registado nas últimas décadas potenciou o uso deste tipo de aplicações e têm despertado a reflexão de novos autores e o surgimento de definições mais abrangentes.

Para Michael Lesk (Lesk, 2005), uma biblioteca digital é uma colecção de informação digitalizada e organizada.

David Bainbridge (Bainbridge & Witten, How to Build a Digital Library, 2003) engloba os conceitos de serviço e manutenção, considerando uma biblioteca digital como uma colecção de informação organizada, focada em objectos digitais, texto, vídeo e áudio, incluindo métodos de pesquisa e disponibilização, selecção, organização e manutenção.

A Comissão das Comunidades Europeias, na sua Comunicação da Comissão ao Parlamento Europeu em 2005 afirma que *“As bibliotecas digitais são colecções organizadas de conteúdos digitais. Podem consistir em material que foi digitalizado, como cópias digitais de livros e de outro material “físico” de bibliotecas e arquivos. Alternativamente podem basear-se em informações produzidas originalmente em formato digital.”* (Comissão das Comunidades Europeias, 2005).

Nesta definição são evidenciadas as possíveis origens dos objectos digitais que compõem a biblioteca, digitalizados ou “nascidos digitais”.

É perceptível a diferença entre a internet, enquanto repositório de informação dotado de um motor de pesquisa, e uma biblioteca digital, carecendo a primeira de organização, selecção criteriosa e classificação dos objectos digitais.

Uma biblioteca digital também não é uma “biblioteca digitalizada”. O seu objectivo não é a substituição dos tradicionais livros pelas suas versões digitais. Tão pouco servem o propósito de desprestigiar as tradicionais instituições.

As bibliotecas digitais apresentam, na sua génese, características significativamente distintas das bibliotecas tradicionais.

Além da evidente distinção quanto às fronteiras e características do acesso aos seus conteúdos, nas bibliotecas tradicionais é facilmente perceptível a sua persistência, formalidade e dimensão, traduzindo os valores que presidiram à sua construção e selecção do repositório, antagónico com a opacidade de uma página inicial de uma biblioteca digital, escondendo a sua dimensão e qualidade documental.

Contudo, a inexistência de fronteiras não implica a inexistência de limites, essenciais para a definição do seu âmbito e objectivo e para uma social e eticamente responsável selecção de materiais. (Bainbridge & Witten, *How to Build a Digital Library*, 2003)

Outro termo que habitualmente surge associado a este tipo de repositórios é o de biblioteca virtual.

Uma biblioteca virtual não contém os documentos, é constituída por apontadores para os locais onde aqueles residem apresentando semelhanças com um portal.

O processo de criação de uma biblioteca virtual apresenta duas vertentes; a recolha dos apontadores para os documentos relevantes para o tema e objectivo da biblioteca; e o enriquecimento com metadados.

O valor acrescentado das bibliotecas virtuais é evidente, sobretudo tendo em conta a crescente abrangência dos conteúdos disponibilizados na Web, exacerbada pelo potencial que apresenta em termos comerciais. Uma pesquisa sobre quase qualquer tema devolve um número considerável de apontadores. (Lesk, 2005)

Um exemplo de biblioteca virtual é a Biblioteca do Conhecimento Online (b-on) que tem como missão “*Garantir o acesso a um vasto número de publicações de natureza científica e serviços electrónicos à comunidade académica e científica nacional.*” (Fundação para a Computação Científica Nacional; UMIC Agência para a Sociedade do Conhecimento, IP, 2007)

2.2 Aspectos Socioeconómicos das Bibliotecas Digitais

A criação de uma biblioteca digital é um processo que pode facilmente ser subestimado quando observado numa perspectiva exclusivamente tecnológica. As funcionalidades disponibilizadas pelo *software* e seu nível de abstracção podem facilmente esconder um processo bastante complexo e abrangente.

Aspectos sociais, éticos, legais e económicos são determinantes na definição de âmbito, objectivo e selecção documental e podem condicionar a decisão de avançar ou não com o projecto.

O processo de recolha da informação e a sua disponibilização de forma abrangente obriga a uma responsabilidade social e ética acrescida para quem se encontra num processo de construção de uma biblioteca digital. A selecção de conteúdos não se esgota nas opções tomadas relativamente ao tipo de informação que se pretende disponibilizar, devendo ser observado se o conteúdo documental não é social e culturalmente ostensivo para a comunidade que potencialmente o consultará e para a comunidade de onde provém.

Deve ainda existir especial atenção à divulgação de determinados conteúdos que não devam ser do domínio público por conterem informação de complexidade não acessível a leigos na área de conhecimento em que se inserem, e que por tal, possam vir a ser descontextualizados, gerar interpretações e generalizações erradas podendo servir interesses contrários ao bom funcionamento das comunidades.

Aspectos legais também não podem ser descurados. Os direitos de propriedade intelectual têm de ser garantidos e revestem-se de grande complexidade no caso particular das bibliotecas digitais, podendo constituir um bloqueio incontornável no desenvolvimento de um projecto. (Bainbridge & Witten, *How to Build a Digital Library*, 2003)

Os aspectos económicos estão relacionados principalmente com três questões distintas, a origem do material que se pretende incluir no repositório, os direitos de uso de propriedade intelectual e a sustentabilidade económica da biblioteca.

A origem do material influencia directamente os custos de inicialização do repositório. Se a opção recair sobre a construção de uma biblioteca a partir de materiais físicos já existentes recorrendo à digitalização pode tornar-se num projecto consideravelmente oneroso, potenciado pelo grau de qualidade e fidelidade ao original que se pretender atingir. As necessidades legais de protecção dos direitos de autor face à capacidade tecnológica adequada a essa protecção são custos a ter em conta. Por fim a sustentabilidade económica da biblioteca, a manutenção e actualização dos sistemas de informação que a suportam ou a aquisição de novos documentos são custos que têm de ser assegurados, directamente indexados à quantidade de subscritores, à dimensão dos conteúdos (capacidade de armazenamento) e à disponibilidade (alta disponibilidade normalmente está associada a redundância e conseqüentemente a custos elevados em sistemas de informação complexos).

2.3 Tecnologia

A escolha da plataforma tecnológica é provavelmente o processo mais simples. Existem diversos trabalhos e estudos que permitem estabelecer critérios de avaliação de software com capacidade de endereçar a construção de uma biblioteca digital, adaptando-se aos requisitos mais comumente enunciados.

Neste trabalho a escolha da plataforma ficou determinada logo no seu início. Estiveram na base desta selecção a relação entre o custo de aquisição e licenças, a sua adequabilidade ao objectivo e recursos disponíveis para garantir as fases de implementação, configuração, disponibilização e manutenção do repositório.

Partindo desta base, a escolha entre plataformas ao abrigo da licença GNU (*General Public License*), surgiu naturalmente. Adicionalmente foram considerados:

- Os trabalhos desenvolvidos na área do software “*open source*”;
- A possibilidade de possuir uma plataforma para investigação e desenvolvimento de novas soluções;
- A simplicidade do processo de instalação e configuração, deverá ser suficientemente intuitivo para poder ser executado por pessoal não especializado em ciência bibliotecária;
- Suporte para os *standards* mais usuais, nomeadamente no que concerne aos formatos de documentos, conteúdos multimédia, metadados, interface e interoperabilidade;
- Possibilitar a integração com repositórios nacionais e internacionais, em particular com o Repositório Científico de Acesso Aberto de Portugal (RCAAP);
- Permitir adequabilidade aos requisitos específicos do projecto, em especial na perspectiva dos metadados e interface do utilizador;
- Ter uma considerável base instalada e existir acesso fácil a uma boa base de conhecimento para suporte técnico;
- Ser flexível quanto à portabilidade do repositório.

2.4 A Plataforma Greenstone entre outras

O Greenstone foi desenvolvido pelo projecto *New Zeland Digital Library Project* da Universidade de Waikato, distribuído pela UNESCO e Human Info NGO.

Esta plataforma é suficientemente escalável para responder ao extenso número de requisitos das bibliotecas digitais, beneficia de uma alargada comunidade que contribui para a constante evolução e adequabilidade do sistema e uma extensa base instalada onde se incluem organizações estatais italianas, a própria UNESCO e a Universidade das Nações Unidas em Tóquio.

É usada em diversos sítios na internet e existem interfaces e colecções em diversas línguas, como por exemplo, Português, Chinês, Maori ou Árabe. (Bainbridge & Witten, *How to Build a Digital Library*, 2003)

Não obstante a escolha predeterminada, investigou-se a existência de outras plataformas no mesmo âmbito. Como resultado desta investigação apresentam-se de seguida algumas aplicações que poderiam igualmente garantir os requisitos de criação de uma biblioteca digital.

EPRINTS

EPRINTS foi desenvolvido pela Universidade de Southampton tendo sido lançada a primeira versão no ano 2000. (School of Electronics and Computer, University of Southampton, UK, 2009). O objectivo inicial do EPrints foi proporcionar o acesso via OAI (*Open Archives Initiative*) a documentos de investigação e literatura académica. Permite também ser usado como um arquivo de outros documentos electrónicos. (EPrints)

FEDORA

Flexible and Extensible Digital Object and Repository Architecture (FEDORA) foi desenvolvido conjuntamente pela Universidade de Virginia e Universidade de Cornell, tendo sido lançada a primeira versão em 2003. Suportado numa arquitectura por camadas, usa XML e serviços Web. Foi desenhado com o intuito de disponibilizar uma base para construção de repositórios digitais e interoperabilidade entre bibliotecas digitais suportadas na Web. (Fedora Commons, 2009)

D-SPACE

D-SPACE foi desenvolvido pelo Massachusetts Institute of Technology Libraries (MIT) em associação com a Hewlett-Packard Labs.

Foi criado com o objectivo de preservar a extensa produção intelectual do corpo académico do MIT. A sua génese determina uma forte orientação para o uso institucional e para a preservação a longo prazo dos objectos digitais, possuindo acordos com organizações proprietárias de determinados formatos as quais garantem o acesso aos documentos por um alargado período temporal.

De todas as plataformas anteriormente enumeradas, o Greenstone e o D-Space são as mais amplamente usadas para construção de bibliotecas digitais. (Tam, Wat, & Kennedy, 2007).

Não sendo o âmbito deste trabalho elaborar um estudo comparativo, considerou-se útil evidenciar algumas diferenças fundamentais entre estas duas plataformas, reforçado pela intenção do projecto da instituição para o qual este trabalho pretende contribuir de migrar o repositório para o D-Space.

Enunciam-se as diferenças consideradas mais óbvias entre as duas plataformas. (Ian H. Witten, 2005)

Preservação dos documentos digitais

O D-Space é especificamente orientado para a preservação a longo prazo dos objectos digitais, inclui um esquema onde as instituições se comprometem a manter determinados formatos. O Greenstone não foi desenhado com base neste objectivo.

Infra-estrutura de suporte

O D-Space foi pensado para ser usado em instituições que possuem instalações centralizadas e suporte competente.

O Greenstone pode ser instalado em qualquer plataforma, desde um simples computador portátil até ao sistema corporativo exigindo apenas conhecimentos básicos de computação.

Operação em multi-plataforma

O Greenstone é compatível com todas as plataformas Windows (a partir da versão 95), Unix e Mac OS/X. O D-Space apenas pode ser executado em Unix e OS/X.

Orientação (Autor vs Bibliotecário)

O Greenstone possui um interface orientado ao bibliotecário. Disponibiliza para o efeito um interface gráfico intuitivo que permite a selecção de elementos digitais, enriquecimento e construção de uma colecção digital. Não permite que os utilizadores finais (tipicamente os autores) possam adicionar objectos à colecção.

O D-Space é claramente orientado ao autor. Permite a submissão de objectos e preenchimento dos metadados por utilizadores finais obrigando apenas à especificação de três campos do esquema de metadados sendo os restantes opcionais. Permite desenhar um interface para o bibliotecário mas este processo não é acessível ao típico utilizador final.

Metadados

O Greenstone permite adoptar diferentes esquemas de metadados em diferentes colecções, permitindo inclusive o desenho de esquemas *ad-hoc*.

O D-Space utiliza apenas o *Qualified Dublin Core* para toda a biblioteca.

Portabilidade da Biblioteca

O Greenstone possibilita a exportação da biblioteca para um vulgar CD-ROM permitindo a sua execução em todos os ambientes Windows (incluindo nas versões 3.1x).

Actualização Dinâmica

No Greenstone é necessário reconstruir a biblioteca para possibilitar o acesso a novos objectos.

No D-Space, os objectos recentemente adicionados são apresentados sobre o lado direito da biblioteca e podem ser acedidos de imediato. Disponibiliza uma ligação directa (*hiperlink*) ao objecto a partir da página inicial.

Linguagens

Ambas as plataformas suportam múltiplas linguagens, 35 no Greenstone e 15 no D-Space.

É evidente a diferente orientação das duas plataformas, ambas suportam a construção de bibliotecas digitais com uma complementaridade interessante. Algumas instituições optam por usar o Greenstone ou o D-Space de acordo com os objectivos específicos de cada projecto, não existindo uma plataforma comum a todos os repositórios.

O intercâmbio entre estas duas plataformas é garantido pelos protocolos utilizados, compatíveis entre ambas, pode ser encontrado na literatura de suporte de uma a metodologia de importação de repositórios construídos na outra.

2.5 Metadados

Os metadados são dados suplementares relacionados com os documentos ou objectos e permitem associar informação com o objectivo de melhor o descrever.

Na perspectiva de uma biblioteca tradicional é facilmente perceptível a utilidade dos metadados, no limite, apresentam-se como uma alternativa a ter de procurar um livro em todas as prateleiras. No meio digital onde é possível a pesquisa no conteúdo dos documentos esta vantagem já não é tão clara.

As principais funções dos metadados são:

1. Os metadados permitem expandir a descrição do objecto, quer porque o uso de palavras não se adequa ou é mesmo impossível descrever o objecto usando-as, por exemplo, pode ser uma imagem, uma animação, uma escultura, ou uma gravação sonora;
2. Os metadados podem fornecer informação sobre o objecto, como por exemplo, onde pode ser encontrado ou qual o uso que pode ser feito do mesmo, estendendo o conceito de conteúdo;
3. Os metadados podem conter informação histórica sobre o objecto, a sua proveniência, dimensões ou peso;
4. Os metadados podem resumir algumas propriedades do objecto;
5. Os metadados disponibilizam informação sobre o objecto de uma forma normalizada.

No largo espectro de estruturas de metadados a especificação MARC e Dublin Core são uma referência incontornável nesta temática. (Bainbridge & Witten, *How to Build a Digital Library*, 2003)

MARC

MARC (*MAchine-Readable Cataloging*) é um esquema compreensível, bem desenvolvido, minuciosamente controlado, pensado para ser usado em bibliotecas por profissionais. O processo de preenchimento dos campos que compõem esta especificação é demorado e exigente, sendo regulado pela *Anglo-American Cataloging Rules* (AACR) e pela *Classificação Decimal de Dewey* (DDC) também conhecido como *Sistema Decimal de Dewey*, e adoptada como padrão internacional pela ISO 2709. (Library of Congress Network Development and MARC Standards Office, 2008)

Dublin Core

A especificação Dublin Core (DC) é um conjunto de elementos de metadados desenhado com o objectivo de ser utilizado por não especialistas e é constituído por apenas quinze elementos.

Deve o seu nome à cidade de Dublin, Ohio onde ocorreu a primeira reunião sobre o tema em 1995. Recebeu a aprovação da ANSI (*American National Standards Institute*).

Com a utilização do meio electrónico para disponibilizar documentos e utilização universal da Web, rapidamente se percebeu que a especificação MARC era demasiado pesada para descrever documentos digitais, além de conter alguns campos inadaptados ao novo ambiente.

Por outro lado o HTML na sua génese incluía etiquetas <meta> que permitiam que fosse adicionada mais informação ao documento, mas cedo esta funcionalidade foi aproveitada abusivamente para intuits comerciais depreciando a sua utilização como metadados. (Lesk, 2005)

É neste contexto, que a especificação Dublin Core surge como uma proposta de esquemas de metadados que respondessem à descrição de documentos na Web de forma simples, intuitiva, flexível e escalável. (Lesk, 2005).

2.6 Outros esquemas de metadados

BibTeX

BibTeX é uma particularização do mais alargado sistema TeX (pronuncia-se *Tech*), é preferido pela maioria dos autores de trabalhos científicos e tecnológicos que usam com frequência notação matemática.

Possui características especiais na formatação do estilo e conteúdos dos seus elementos que permitem, por exemplo, uma pesquisa muito afinada sobre um repositório com abreviaturas dos conteúdos dos elementos muito usado por académicos que frequentemente criam colecções bibliográficas BibTeX na sua área de conhecimento, acumulando referências durante anos e produzindo extensos repositórios de metadados, os quais em conjunto com heurísticas de reconhecimento de entradas duplicadas, constituem uma útil ferramenta usada em bibliotecas digitais nas áreas técnico-científica. (Bainbridge & Witten, *How to Build a Digital Library*, 2003).

2.7 Metadados para Imagens e Multimédia

A ideia de incluir metadados não se restringe aos documentos de texto. Em conteúdos gráficos e multimédia, a dificuldade de pesquisar determinado conteúdo numa imagem áudio ou outro conteúdo multimédia face à facilidade de uma pesquisa *full-text* sobre um documento de texto, torna ainda mais relevante a possibilidade de associar de forma flexível metadados a este tipo de recursos.

Considerou-se importante referenciar dois formatos de metadados que, pela sua extensa usabilidade e especificações, são referência incontornável nesta área. (Bainbridge & Witten, *How to Build a Digital Library*, 2003)

TIFF (*Tagged Image File Format*)

TIFF é um formato que se tornou extensamente usado e que permite a descrição de imagens tipicamente provenientes de scanners e programas de fotografia. Actualmente é o esquema usado para gravar imagens e documentos gráficos nas bibliotecas digitais.

MPEG-7

MPEG-7 é um esquema mais complexo, sofisticado e ambicioso que permite definir e gravar metadados associados a qualquer conteúdo multimédia. Actualmente está em fase de definição do *standard*. A sua aplicação excede em larga escala o âmbito das bibliotecas digitais, cruzando áreas tão distintas como o jornalismo, a vigilância ou as aplicações biomédicas. (Bainbridge & Witten, *How to Build a Digital Library*, 2003)

2.8 Acesso

O objectivo das bibliotecas é proporcionar o acesso à informação e neste campo, as bibliotecas digitais detêm um potencial que alavanca significativamente este objectivo.

Actualmente é possível aceder virtualmente de qualquer local do planeta à informação, não é necessário deslocarmo-nos até ao edifício da biblioteca, ela chega até nós, no nosso PDA ou computador pessoal.

A grande maioria das bibliotecas digitais encontram-se disponíveis via Web e actualmente algumas já disponibilizam recursos para desenho do acesso a partir de um dispositivo móvel (*Smart Phone* ou PDA)

2.9 Preservação dos objectos digitais

Ao longo dos anos a necessidade crescente da humanidade de obter informação e conhecimento condicionou; quer a quantidade de informação produzida; quer o tipo de suporte onde se encontrava registada.

Os nossos antepassados registavam as suas mensagens em pedra, mais tarde surgiu o barro, seguidamente o papiro (4500 anos) que permitiu a representação mais rica dos conteúdos e o papel há cerca de 2000 anos que se assumiu como a grande invenção da humanidade. Na Grande Biblioteca de Alexandria coexistiram os três suportes (barro, papiro e papel) persistindo o papel até à actualidade.

Com a descoberta da electricidade e a possibilidade de desenvolver todo o tipo de equipamento electrónico actual o meio digital tornou-se, nos últimos 50 anos, o principal meio de divulgação de informação.

Cada estágio, desde a pedra até ao meio digital tem determinado um aumento na capacidade de transporte da informação, na capacidade de a produzir e integrar.

Em contraste, o meio mais resistente era sem dúvida a pedra, resistindo milhares de anos a todo o tipo de intempéries (hieróglifos), o meio digital é nitidamente o mais frágil, perdendo-se informação devido a falta de persistência do meio e estratégias inconsistentes de preservação. (Berkman, 2006)

Uma das mais recentes técnicas que tem vindo a ser utilizada na preservação de documentos digitais é a emulação.

A emulação apresenta entre outras as seguintes vantagens:

- Existem menos plataformas para emular comparado com os formatos e aplicações;
- O sigilo comercial dificulta a obtenção dos formatos de muitos programas dificultando a conversão dos documentos neles produzidos;
- Pode ser muito complexo imitar o comportamento de um programa, sendo mais fácil executar o código original;
- A evolução de um emulador de um dispositivo traduz-se na valorização dos objectos que se pretendem armazenar.

A grande dificuldade da emulação é conseguir reproduzir fielmente e eficazmente um sistema antigo.

A volatilidade da informação digital conduz-nos a uma reflexão sobre que conteúdos queremos ter ou quais devemos manter, orientando a selecção e a construção de colecções com maior qualidade, mais adequadas ao seu propósito e às expectativas das comunidades que pretendem servir. (Lesk, 2005)

Uma visão interessante neste âmbito é a de Moxley, Joseph M., professor na Universidade de *South Florida* em Tampa, USA que afirma:

“Será sempre melhor que um trabalho possa ser consultado por milhares durante alguns anos do que estar disponível durante milhares de anos mas apenas consultado por alguns.”
Traduzido pelo autor (Moxley, 2001)

2.10 Propriedade Intelectual

As questões legais decorrentes do uso dos direitos de autor são o maior problema que as bibliotecas digitais enfrentam.

A questão fundamental não reside no facto dos utilizadores não estarem dispostos a suportar o custo da utilização da propriedade intelectual mas antes no melhor e mais justo meio de o fazer.

Existem algumas tecnologias de suporte para protecção de propriedade intelectual:

- Acesso Fraccionado (inibe o acesso total ao documento)
- Cadeados Físicos (*Hardware Locks* ou *Dongles*)

- Estenografia
- Criptografia
- Políticas Económicas (esmagamento de preços destrói a rentabilidade da cópia ilegal)
- Cintilação (*Flickering*) (suportada na frequência da visão humana, alterna o documento legível com um ecrã de pixéis negros ou brancos tornando o texto ilegível, no entanto o olho humano não se apercebe da cintilação ao contrário de uma tentativa de cópia (*print screen*)).

O enquadramento legal em Portugal está consignado na Lei nº16/2008 de 1 de Abril (Código dos Direitos de Autor e dos Direitos Conexos). Esta e outra legislação relacionada podem ser consultada a partir do site do Gabinete de Planeamento Estratégia, Avaliação e Relações Internacionais (GPEARI) do Ministério da Cultura (www.gpeari.pt).

Encontram-se definidas as regras de submissão e cedência de direitos sobre os trabalhos publicados na instituição decorrente do repositório digital básico de trabalhos científicos já em uso.

2.11 Interoperabilidade

Uma parte importante da cultura tradicional estabelecida entre bibliotecas é o intercâmbio de informação. A possibilidade de localizar um recurso entre várias bibliotecas e beneficiar de um empréstimo inter-bibliotecas é um factor importante da divulgação cultural. As bibliotecas juntam esforços para divulgar informação e o grau de cooperação é significativamente extenso e conhecido.

Para que exista um entendimento e comunicação efectivos entre as bibliotecas são necessários protocolos de representação de documentos, metadados e *queries*.

Destacam-se pela sua alargada adopção das bibliotecas digitais, a iniciativa OAI (Open Archives Initiative) e o *standard Z39.50*.

O *standard Z39.50* define um protocolo de recolha de informação num ambiente cliente-servidor de base de dados.

A iniciativa OAI nasceu sustentada na forte motivação da comunidade académica no intercâmbio dos trabalhos académicos e de investigação e no aumento do potencial de acesso aos mesmos e encontra-se descrito com maior detalhe no parágrafo seguinte. (Bainbridge & Witten, *How to Build a Digital Library*, 2003)

2.12 Iniciativas Relevantes

Existem actualmente iniciativas internacionais e nacionais que pelas suas características e contribuições para o desenvolvimento das bibliotecas digitais, assumem um papel relevante nesta temática.

2.12.1 Iniciativas Internacionais

Networked Digital Library of Theses and Dissertations (NDLTD)

A NDLTD é uma organização internacional criada exclusivamente com finalidades educacionais e de apoio que se dedica a promover a adopção, criação, uso, divulgação e preservação do equivalente electrónico das tradicionais Teses e Dissertações em suporte de papel. (Networked Digital Library of Theses and Dissertations, 2008).

Em 1996 a SURA (*Southeastern Universities Research Association, Inc.*) iniciou o “*Virginia Tech's electronic thesis and dissertation project* “ com o objectivo de divulgar e disseminar a partir das Universidades um método normalizado de publicação e disponibilização em linha dos trabalhos de graduação.

Os principais objectivos desde projecto são:

- Permitir aos estudantes a aprendizagem dos temas publicação electrónica e livrarias digitais, aplicando este conhecimento no processo de pesquisa, criação e submissão do seu próprio trabalho;
- Para as universidades conhecerem e aprenderem sobre livrarias digitais à medida que recolhem, catalogam, arquivam e tornam acessíveis globalmente as teses e dissertações produzidas pelos seus alunos;
- Para as universidades do Southeast e outras entenderem como podem potenciar a sua propriedade intelectual e produção de trabalhos;
- O desenvolvimento do conhecimento académico através da maior partilha e;
- O desenvolvimento tecnológico e do conhecimento à medida que os resultados de investigação se tornam mais facilmente disponíveis e acessíveis.

Traduzido pelo autor de (The Networked Digital Library of Theses and Dissertations, 1996)

Open Archives Initiative (OAI)

A OAI desenvolve e promove normas de interoperabilidade que visam facilitar uma disseminação eficiente de conteúdos. Pedra basilar do programa da OAI é o trabalho inicialmente desenvolvido no sentido de potenciar o acesso a arquivos digitalizados como meio para incrementar a comunicação académica.

Actualmente a OAI é uma organização em transição comprometida em explorar e viabilizar o crescente número de novas aplicações. Com base no entendimento das características tecnológicas subjacentes e normas desenvolvidas e do crescente conhecimento da estrutura cultural das várias comunidades aderentes a OAI acredita que deverá evoluir e desenvolver alterações continuadas na sua missão e organização. (Open Archives Initiative, s.d.)

Principais projectos em curso:

- **Protocol for Metadata Harvesting (OAI-PMH)**

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) é um mecanismo que permite a interoperabilidade entre repositórios.

Estão subjacentes dois actores:

- Os *Data Providers* que disponibilizam metadados estruturados via OAI-PMH.
- Os *Service Providers* que efectuem pedidos de serviço ao OAI-PMH para recolha dos metadados.

Resumidamente o OAI-PMH é um conjunto de seis serviços invocados dentro do protocolo HTTP.

- **Object Reuse and Exchange (OAI-ORE)**

Define normas para descrição e troca de agregação de recursos Web. O objectivo é disponibilizar o conteúdo variado destas agregações (objectos digitais compostos) os quais compreendem vários tipos de suporte (texto, imagens, dados e vídeo). Por exemplo, habitualmente referenciamos um conjunto de informação obtida na Web pelo URI da primeira página, mas este identificador apenas se refere à primeira página e não ao agregado de informação (texto, imagens, vídeo) que efectivamente se pretende endereçar. O objectivo deste projecto é desenvolver normas para generalização de toda a informação baseada na Web incluindo a cada vez mais popular rede social “Web 2.0” (Open Archives Initiative, s.d.)

2.12.2 Iniciativas Nacionais

Repositório Científico de Acesso Aberto de Portugal (RCAAP)

O RCCAP é uma iniciativa da UMIC - Agência para a Sociedade do Conhecimento, IP, concretizada pela Fundação para a Computação Científica Nacional (FCCN). Este projecto conta com a colaboração técnica da Universidade do Minho e visa aumentar a visibilidade, acessibilidade e difusão dos resultados da actividade académica e de investigação científica nacional e facilitar o acesso à informação sobre a produção científica nacional em regime de “*open access*”, bem como integrar Portugal num conjunto de iniciativas internacionais neste domínio. (Ministério da Ciência, Tecnologia e Ensino Superior, 2006)

Depósito de Teses e Dissertações na BNP (DiTeD)

“ *O serviço DiTeD destina-se a gerir teses e dissertações em formato digital, entregues como depósito voluntário na BNP - Biblioteca Nacional de Portugal. A BNP assegurará para essas obras o seu registo, preservação a longo prazo, divulgação e ainda o acesso nas condições que forem determinadas.*”

“*O serviço DiTeD está associado a outros projectos semelhantes, através da iniciativa NDLTD - Networked Digital Library of Theses and Dissertations. As obras registadas no DiTeD beneficiarão assim de um elevado potencial de divulgação internacional.*” (Biblioteca Nacional Digital, 2006)

3 Estudo de solução

Como suporte ao estudo da solução, desenvolveu-se uma investigação sobre a plataforma Greenstone, caracterizando os principais aspectos técnicos directamente relacionados com a gestão de metadados e com o processo de inicialização de colecções digitais.

O actual site estático a partir do qual são acedidas as publicações científicas está alojado na instituição e os ficheiros em pastas temáticas. Existe também um ficheiro *Microsoft Excel* de classificação dos documentos que representaria os metadados dos documentos armazenados.

A arquitectura do protótipo desenvolvido permite uma utilização flexível de diferentes especificações de metadados, permitindo à instituição utilizar a ferramenta após a estabilização da especificação de metadados.

3.1 Metodologia

O Greenstone possui basicamente duas formas de criar colecções; executando sequencialmente os *scripts* escritos em Perl, passando os parâmetros adequados; ou utilizando o interface gráfico do bibliotecário (GLI). A primeira metodologia, apesar de ser menos intuitiva e expedita é mais detalhada na utilização dos recursos da plataforma e foi a abordagem inicialmente escolhida na procura de uma solução.

Na literatura de suporte foram encontradas algumas orientações do sentido de gerir a origem dos metadados bibliográficos a associar à colecção:

“ O standard Plug-in RecPlug incorpora a possibilidade de assignar metadados a documentos manual ou automaticamente a partir de ficheiros XML. <...>. Se a opção use_metadata_files estiver especificada, RecPlug usa um ficheiro auxiliar com o nome metadata.xml.

<...>

Quando for usada a opção use_metadata_files, RecPlug verifica em cada directoria a existência de um ficheiro XML com o nome metadata.xml e aplica o seu conteúdo a todas as directorias e subdirectorias.

<...>

O mecanismo metadata.xml embebido no Plug-in RecPlug é apenas uma das vias de especificar os metadados para os documentos. É fácil escrever Plug-in diferentes que aceitem formatos completamente diferentes de metadados.” (Bainbridge, McKay, & Witten, Greenstone Digital Library : Developers Guide, 2004)

Existe uma opção que possibilita o controlo do ficheiro de metadados, mas para utilizar uma origem diferente da estrutura de metadados obriga ao desenvolvimento de um *plug-in* específico.

Esta abordagem foi abandonada por ser de difícil acessibilidade tendo em conta os recursos disponíveis, exige a compreensão detalhada do encadeamento de alguns processos *core* do Greenstone e conhecimentos de linguagem Perl.

O próximo passo foi estudar o interface gráfico (GLI) na perspectiva de encontrar uma funcionalidade que viabilizasse uma solução para o problema.

Também aqui, a única alternativa seria a alteração do código fonte do próprio interface, desenhando a funcionalidade pretendida.

O desenvolvimento da investigação sobre os processos internos do Greenstone clarificou aspectos fundamentais para o desenho da solução.

O processo de enriquecimento está suportado na leitura de uma estrutura relacionada com o esquema de metadados seleccionado, escrita num ficheiro em formato XML residente no mesmo directório onde estão colocados os objectos da colecção que se pretende construir.

Esse ficheiro de nome *metadata.xml* é lido durante o processo de importação (*import*) o qual é responsável pela conversão do formato original dos documentos no formato de arquivo usado internamente pela plataforma. Este formato inclui os metadados fornecidos com a colecção explicitados numa estrutura XML no ficheiro *metadata.xml* particular dessa colecção. (Bainbridge & Witten, How to Build a Digital Library, 2003) .

Neste ponto já era claro que a solução teria de passar pelo desenvolvimento de uma aplicação externa ao Greenstone que construísse o código XML adequado no ficheiro *metadata.xml* a partir da lista de classificação dos documentos.

Subsequente investigação não introduziu novas perspectivas ou evidenciou outras soluções possíveis para o problema, de facto, a consulta ao suporte da comunidade do Greenstone, fortaleceu esta decisão, tendo sido identificada uma questão que descreve um problema muito similar a qual se transcreve parcialmente:

*"> [Context: we have digitalised documents whose metadata are
> directly created in Greenstone using Dublin Core Set (enrich). We
> have furthermore a lot of digitalised documents described in an
> other tool and which metadata ca be imported in CSV and converted
> in XML. We want to insert them in Greenstone too for complete our
> collection.*

B. Delgado

*.> documentaliste resp.
> doCip
> Centre de documentation, de recherche
> et d'information des peuples autochtones
> www.docip.org"*

“ Hi,

The best way would be to convert the current metadata into a Metadata.xml file. Metadata.xml files store metadata about files within your import directory in a format that Greenstone can read easily. Within the Metadata.xml file you can format the metadata as Dublin Core (with the colons replaced by full stops, I believe) and Greenstone will import it as Dublin Core.<...>

John Rowe

Business Development Manager

DL Consulting

Greenstone Digital Library and Digitisation Specialists

johnr@dlconsulting.com

www.dlconsulting.com

“(Rowe, 2007).

3.2 Arquitectura da solução

Seleccionou-se a plataforma de desenvolvimento *Microsoft Visual Studio 2008* para desenvolver a solução. Criou-se um projecto baseado em *Windows Forms Application* e escolheu-se a linguagem *Visual Basic* para desenvolver o código do protótipo.

Estas escolhas devem-se; por um lado ao baixo nível de experiência do autor em matéria de programação; por outro à possibilidade de desenvolver um protótipo intuitivo e de fácil utilização.

Os projectos *Windows Forms Application* são adequados para a criação de uma aplicação baseada num interface de utilizador similar ao usado no sistema *Windows*. (Microsoft, 2008)

Seleccionou-se a especificação *Dublin Core* para suportar a implementação uma vez que a instituição ainda não estabilizou os metadados numa variante desta norma.

O protótipo desenvolvido apresenta a seguinte arquitectura:

Componentes principais

- Ecrã principal representado por uma classe agregadora de todos os objectos que constituem a aplicação. (Figura 3-1)
- Objectos de diálogo que permitem a navegação no sistema de directórios do sistema
- A função de conversão da lista de metadados em estruturas XML (*GeraXML*)

Entradas

- Localização da lista de metadados dos documentos em formato MS Excel (origem)

- Localização do ficheiro *metadata.xml* (destino)

Saída

- Estruturas XML de acordo com a especificação do Dublin Core não qualificado escritas no ficheiro *metadata.xml* adequadas ao processo de importação do Greenstone.

Função de conversão (*GeraXML*)

- Determina a dimensão da lista de classificadores (linhas x colunas)
- Faz a leitura para uma lista interna da primeira linha para construir os nomes dos campos do esquema de metadados.
- Em ciclo balizado pela dimensão calculada constrói uma estrutura de metadados por cada linha no ficheiro de destino.

Na Figura 3-2 pode ser observado uma amostra das estruturas XML criadas pelo protótipo. Entre as etiquetas `<FileSet>` `</Fileset>` está representada uma linha do ficheiro de classificação, ou seja, o conjunto de metadados para um documento.

Entre as etiquetas `<Description>` `</Description>` encontram-se os campos da especificação de metadados Dublin Core.

Cada campo está especificado entre as etiquetas `<Metadata>` `</Metadata>` aparecendo a sua designação descrita no atributo “name”.

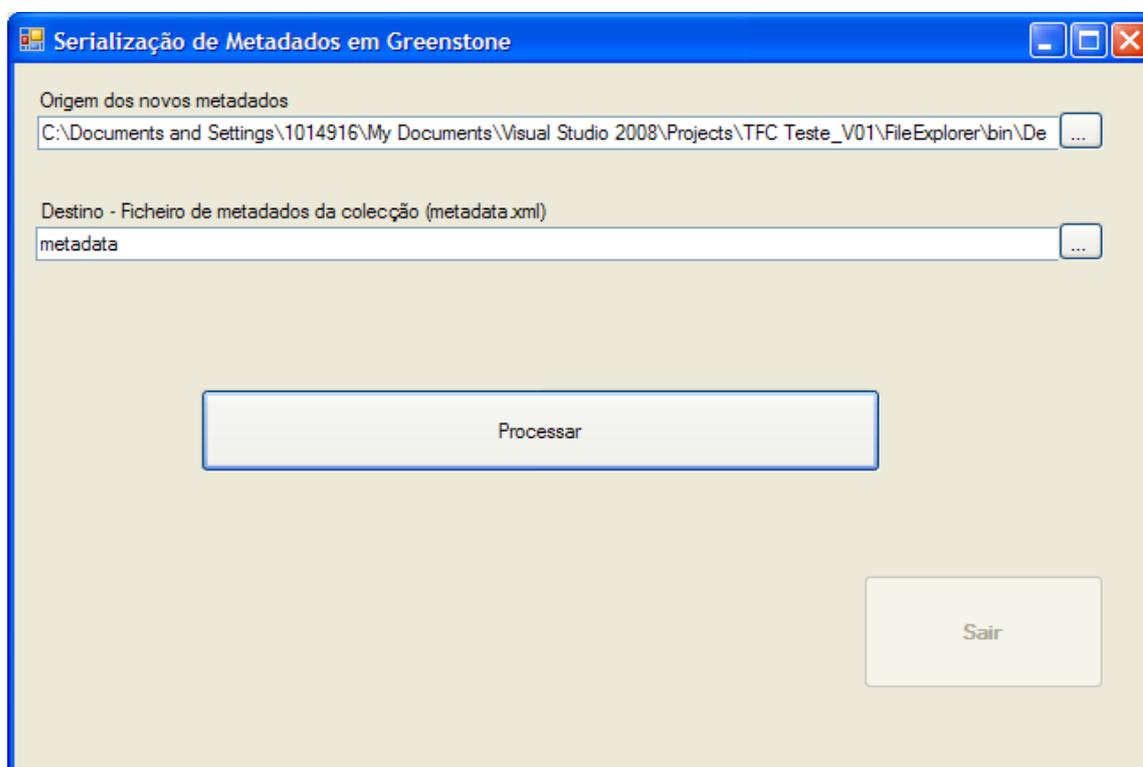


Figura 3-1: Ecrã principal do protótipo

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE DirectoryMetadata (View Source for full doctype...)>
<DirectoryMetadata>
  <FileSet>
    <FileName>generating metadata file\.pdf</FileName>
    <Description>
      <Metadata mode="accumulate" name="dc.Title">Generating Metadata
File</Metadata>
      <Metadata mode="accumulate" name="dc.Creator">Katherine Don</Metadata>
      <Metadata mode="accumulate" name="dc.Subject">metadata, greenstone,
automatic, automatico, metadados, importar</Metadata>
      <Metadata mode="accumulate" name="dc.Description">Resposta a questões
colocadas no forum de utilizadores do Greenstone</Metadata>
      <Metadata mode="accumulate" name="dc.Publisher">greenstone-users
mailing list</Metadata>
      <Metadata mode="accumulate" name="dc.Contributor">University of
Waikato, New Zealand</Metadata>
      <Metadata mode="accumulate" name="dc.Date">2005-09-20</Metadata>
      <Metadata mode="accumulate" name="dc.Type">Text</Metadata>
      <Metadata mode="accumulate" name="dc.Language">en</Metadata>
      <Metadata mode="accumulate" name="dc.Format">PDF</Metadata>
      <Metadata mode="accumulate" name="dc.Identifier">generating metadata
file.pdf</Metadata>
    </Description>
  </FileSet>
  ...
  <mais estruturas FileSet>
  ...
</DirectoryMetadata>
```

Figura 3-2 : Exemplo de estrutura XML de metadados criada pelo protótipo

Este formato de XML é usado em outras especificações de metadados disponíveis no Greenstone, a arquitectura do protótipo permite a sua utilização para outras especificações diferentes da Dublin Core utilizada nesta prova de conceito.

3.3 Pressupostos

Para o correcto funcionamento do protótipo assumiram-se os seguintes pressupostos:

1. A lista de classificação dos documentos (metadados) deve estar no formato MS Excel onde, na primeira linha devem estar representados os campos da especificação de metadados que irá ser utilizada na colecção, nas linhas subsequentes deve constar a classificação dos diversos documentos;
2. Deve existir na directoria *standard* do Greenstone, onde reside a compilação, um ficheiro *metadata.xml*, com ou sem dados mas que contenha as suas definições iniciais (Figura 3-3);
3. O sistema onde está a ser construída a colecção permite o acesso à internet, caso contrário terá de ser criado um ficheiro local com a DTD (*Document Type Definition*) com o nome de *DirectoryMetadata.dtd* e este deve estar convenientemente descrito no cabeçalho do ficheiro *metadata.xml* (Figura 3-4);
4. O nome do ficheiro não poderá conter pontos (“.”). Nestes casos a aplicação entenderá o valor após o ponto como sendo a extensão do ficheiro e desta forma os metadados associados não serão lidos convenientemente pelo Greenstone.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE DirectoryMetadata SYSTEM "C:\DirectoryMetadata.dtd">
<DirectoryMetadata>
  Aqui ficam as estruturas de Metadados
</DirectoryMetadata>
```

Figura 3-3: Formato base do ficheiro *metadata.xml*

```
<!ELEMENT DirectoryMetadata (FileSet*)>
<!ELEMENT FileSet (FileName+,Description)>
<!ELEMENT FileName (#PCDATA)>
<!ELEMENT Description (Metadata*)>
<!ELEMENT Metadata (#PCDATA)>
<!ATTLIST Metadata name CDATA #REQUIRED>
<!ATTLIST Metadata mode (accumulate|override) "override">
```

Figura 3-4: Estrutura de definição de tipo de documento (DTD)

4 Construção de uma colecção utilizando o protótipo

Como contributo prático válido construiu-se uma colecção digital compilada a partir da base bibliográfica utilizando o protótipo no processo de inicialização do repositório.

Desenvolveu-se um estudo sobre as principais funcionalidades dos interfaces do Greenstone disponibilizados para a criação, configuração e gestão de colecções digitais e de especificações de metadados.

Como contributo para a base de conhecimento de operação da plataforma referem-se sempre que considerado relevante opções específicas dos diversos estágios da construção.

4.1 Interface Gráfico do Bibliotecário (GLI)

O Greenstone possui apenas dois tipos de interface, o interface do leitor (*Reader Interface*) e o interface do bibliotecário (*Greenstone Librarian Interface*).

Os leitores acedem à biblioteca digital através do interface de leitor que funciona integrado num Web browser.

O *Greenstone Librarian Interface* (GLI) é um interface gráfico que permite a construção da biblioteca digital disponibilizando as funcionalidades; recolha dos documentos, ou mais genericamente, elementos digitais, que irão compor a colecção; o seu enriquecimento associando a cada elemento metadados; especificar as funcionalidades de pesquisa e navegação que a colecção irá disponibilizar aos utilizadores; construir e distribuir a colecção.

O GLI pode funcionar em quadro modos distintos disponibilizando mais ou menos funcionalidades. Este modo de funcionamento pode ser ajustado na opção *Preferences* do menu *File* do ecrã principal.

Neste trabalho optou-se pelo modo *Expert* por permitir um maior controlo sobre todas as fases do processo e disponibilizar mecanismos de análise e despiste de erros.

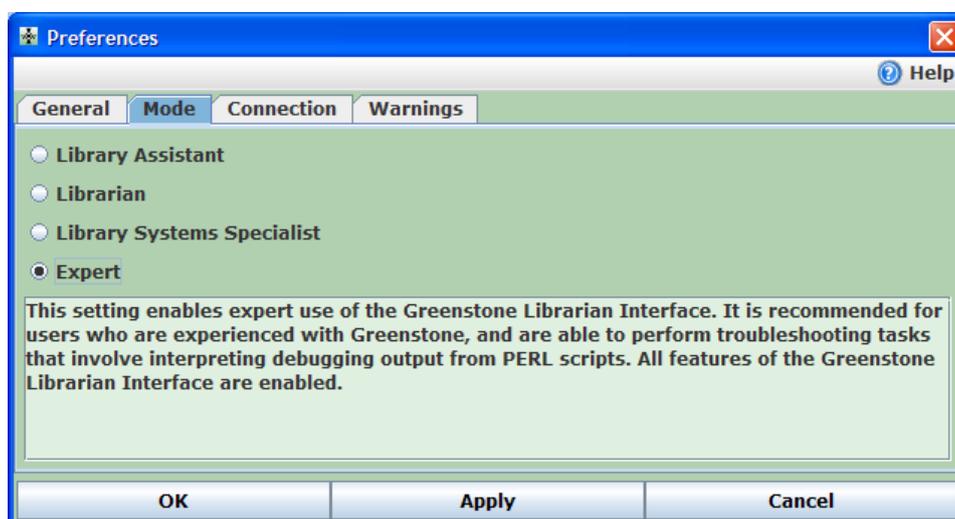


Figura 4-1: Modos de funcionamento do GLI

4.2 Recolha dos documentos

Para a construção desta colecção foram seleccionados alguns documentos da base bibliográfica do presente trabalho.

O GLI possibilita a recolha de elementos digitais directamente da Web ou com origem remota suportado em protocolos de comunicação como o Z39.50 ou o SRW, standards que especificam protocolos para pesquisa e recolha de informação em bases de dados remotas baseados na arquitectura cliente/servidor (The Library of Congress, 2007) e (Taylor, 2003); ou o OAI, uma norma de interoperabilidade que visa a disseminação eficiente de conteúdos via Web (Open Archives Initiative, s.d.).

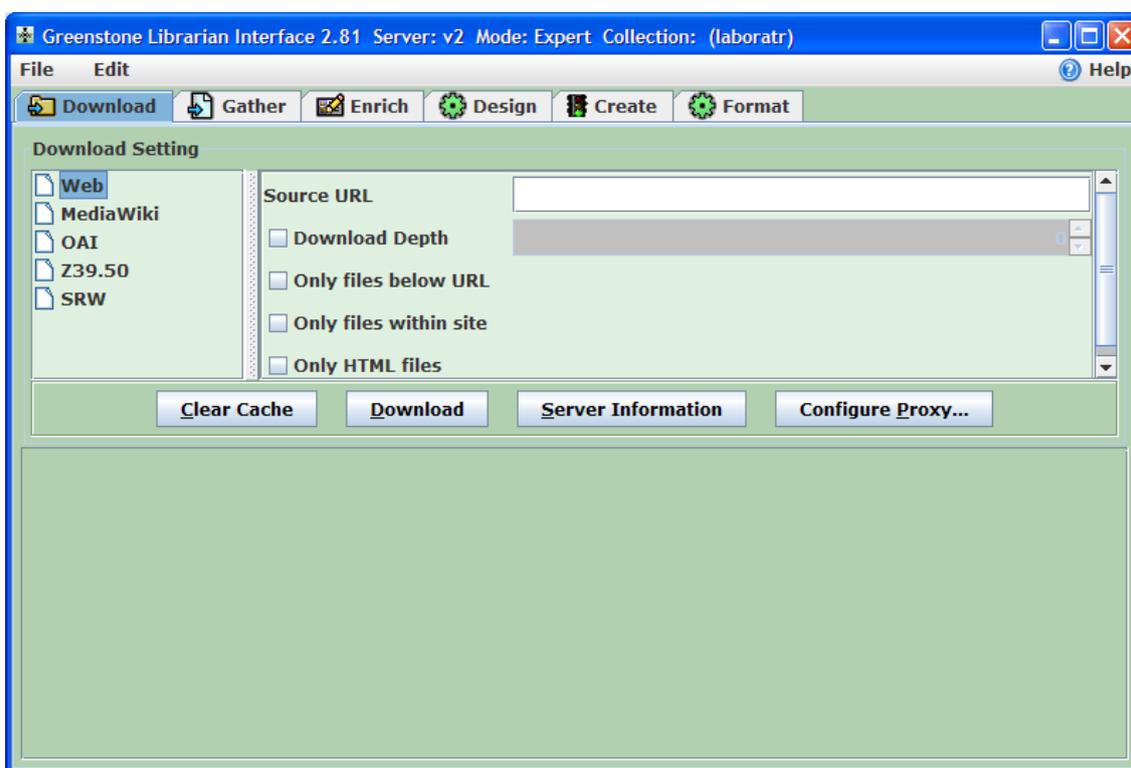


Figura 4-2: Opções de recolha remota de elementos digitais

No presente trabalho a proveniência dos documentos constituintes da colecção é local e neste caso a selecção deve ser executada a partir do quadro *Gather*, onde é apresentado na janela esquerda a estrutura de directórios e ficheiros do sistema local.

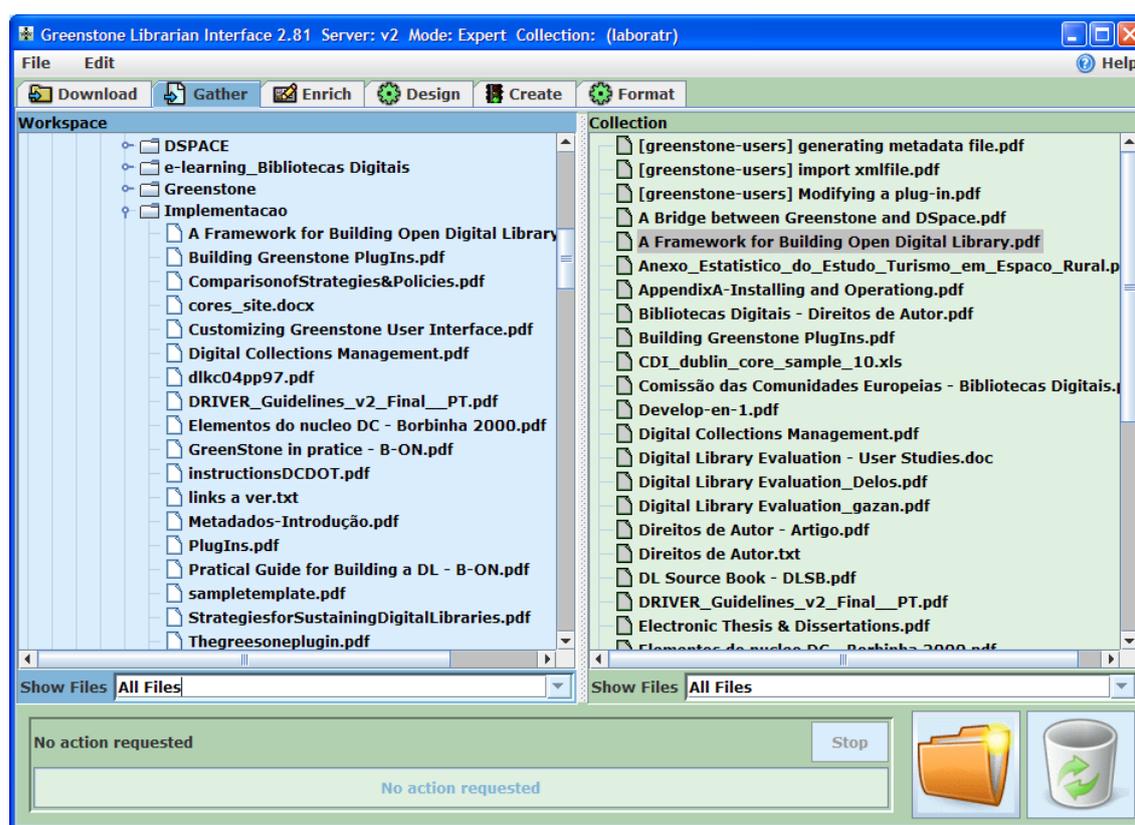


Figura 4-3: Organização dos documentos incluídos na colecção

A Figura 4-3 mostra o resultado da agregação e organização dos documentos seleccionados para este protótipo. Estes documentos são locais mas não necessitam de residir no mesmo directório.

A selecção é bastante intuitiva, bastando para tal arrastar para a janela do lado direito os objectos previamente marcados na janela do lado esquerdo (*drag & drop*).

4.3 Enriquecimento e Metadados

Esta fase é determinante para a qualidade da colecção. A definição da estrutura de metadados e seu conteúdo condicionará as funcionalidades de pesquisa e navegação, características fundamentais do repositório bem como a sua interoperabilidade.

Basicamente o processo pode ser executado de duas formas distintas; definir explicitamente os metadados, ou extrai-los automaticamente do conteúdo dos documentos.

A primeira é mais precisa. Necessita de análise dos documentos e dependendo da estrutura de metadados adoptada pode tornar-se numa actividade significativamente demorada. A segunda é mais expedita e rápida mas menos fiável. Baseia-se em etiquetas HTML, características intrínsecas do documento (por exemplo os *fields* do MS Word) ou

identificação de frases ou palavras “*text mining*” suportadas em técnicas complexas e heurísticas.

A recolha automática de metadados é um tema vasto e objecto de larga discussão não se enquadrando no âmbito deste trabalho. O Greenstone possui uma metodologia intrínseca aos seus *plug-in* que permitem a extracção de determinados metadados no caso de não serem explicitamente preenchidos pelo bibliotecário.

Particularmente interessante no âmbito deste trabalho é a primeira metodologia, permite um maior controlo e adequação do conteúdo dos metadados. De seguida detalhar-se-á este processo.

4.3.1 O Editor de Metadados (GEMS)

O Greenstone possui um interface especialmente desenhado para a gestão das estruturas de metadados, o *Greenstone Editor for Metadata Sets* (GEMS) incluído no pacote base. Pode ser evocado directamente ou durante o processo de criação da colecção escolhendo a opção *Manage Metadata Sets* do ecrã de enriquecimento (*Enrich*).

Como anteriormente referido o Greenstone é particularmente flexível na definição dos metadados permitindo; a selecção de uma das estruturas disponibilizadas, entre as quais o largamente utilizado Dublin Core; a extensão de uma estrutura já existente, adicionando-lhe mais campos de informação; ou criar uma estrutura completamente nova.

A intenção manifestada pela instituição de vir a integrar o Repositório Científico de Acesso Aberto de Portugal (RCAAP) obriga ao cumprimento de um conjunto de requisitos, entre os quais está a definição da estrutura de metadados Dublin Core segundo as directrizes DRIVER 2.0 adoptadas por aquela iniciativa. (Digital Repository Infrastructure Vision for European Research, 2008).

Contudo a instituição ainda não estabilizou a especificação dos metadados que irá ser utilizada. Para a construção desta colecção optou-se por seleccionar a especificação de metadados Dublin Core v1.1 disponibilizada de base com o Greenstone.

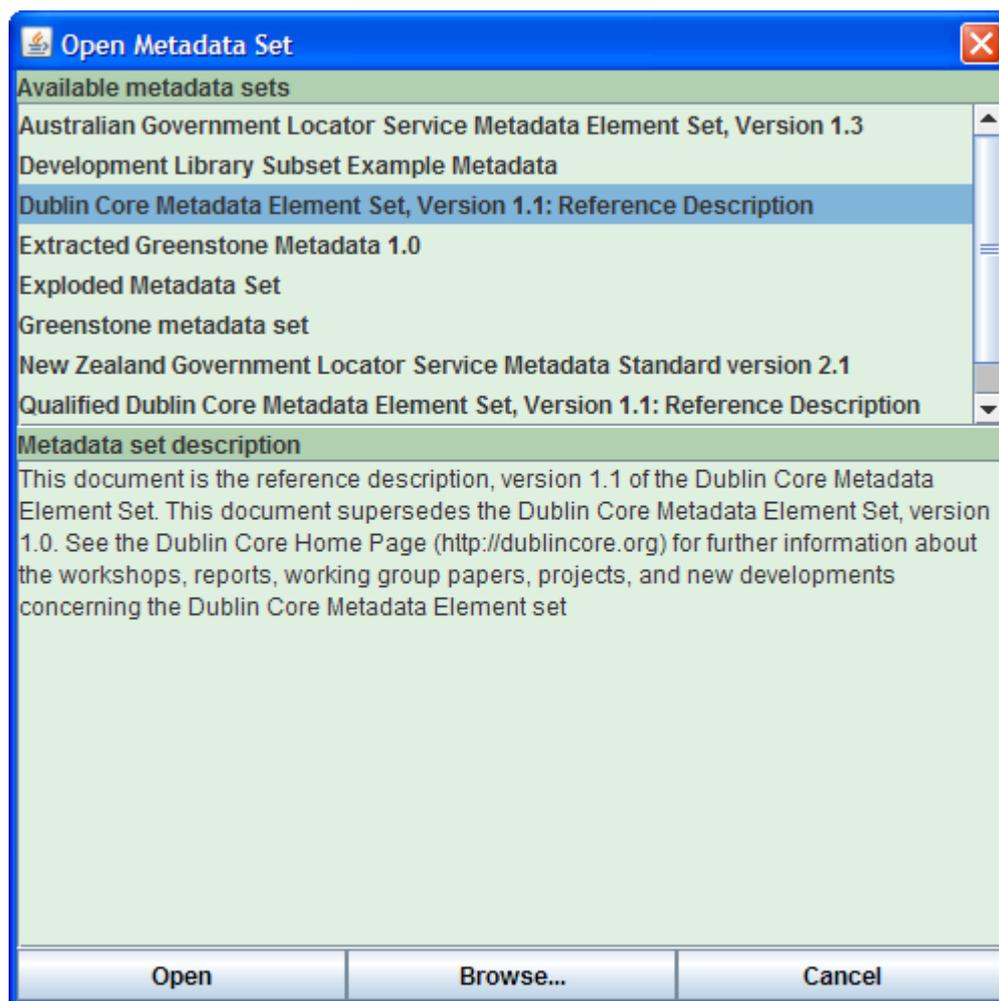


Figura 4-4: Seleção da especificação de metadados

O Dublin Core é uma estrutura de dados bibliográficos composta por quinze campos de dados. Uma descrição comentada pode ser consultada no Anexo B (Borbinha, 2000).

Para o preenchimento dos campos da estrutura Dublin Core será usado o protótipo desenvolvido e um ficheiro *Microsoft Excel* com a classificação dos documentos segundo a especificação Dublin Core. O conteúdo dos campos foi obtido por inspeção directa dos documentos e transposto manualmente para a lista de classificadores.

No preenchimento dos campos da especificação Dublin Core, apesar de não ter sido extremamente rigorosa, foram observadas algumas regras consideradas essenciais para manter a coerência do resultado.

4.3.2 Operação do protótipo

A operação do protótipo é fácil e intuitiva e é quase auto-explicativa.

Para se proceder ao preenchimento automático dos metadados devem ser executados os seguintes passos:

1. Seleccionar o ficheiro *Excel* que contém os metadados a associar aos documentos que irão constituir a colecção pretendida, navegando nos directórios do sistema local ou nos sistemas que lhe estejam ligados de uma forma similar ao explorador do Windows.
2. Seleccionar o ficheiro *metadata.xml*, que deverá estar na mesma directoria onde reside a colecção (requisito 3).
3. Garantir que nenhum dos ficheiros está bloqueado devido a estar a ser utilizado por outro utilizador.
4. Premir o botão “Processar”.
5. Quando o processamento terminar surge uma janela informativa e é disponibilizado o botão “Sair”.

O processamento, compreende:

- A abertura automática do ficheiro *Excel*;
- A selecção dos dados elegíveis (todos os documentos têm de ter pelo menos o nome atribuído, se existir uma linha em branco são descartados todos os dados subsequentes);
- A interpretação dos dados do Excel com base nos campos do Dublin Core;
- A abertura automática do ficheiro *metadata.xml*;
- A construção das estruturas em XML adequadas ao processo de importação do Greenstone;
- O encerramento automático de ambos os ficheiros;
- Reinicializar a compilação executando as opções “close” seguido de “open” no menu “File” do GLI.

Na Figura 4-5 encontram-se representados os campos da especificação Dublin Core antes do seu preenchimento. Sem a utilização do protótipo, esta actividade teria de ser assegurada individualmente para cada campo da estrutura e para cada documento a incluir na colecção listada na janela do lado esquerdo.

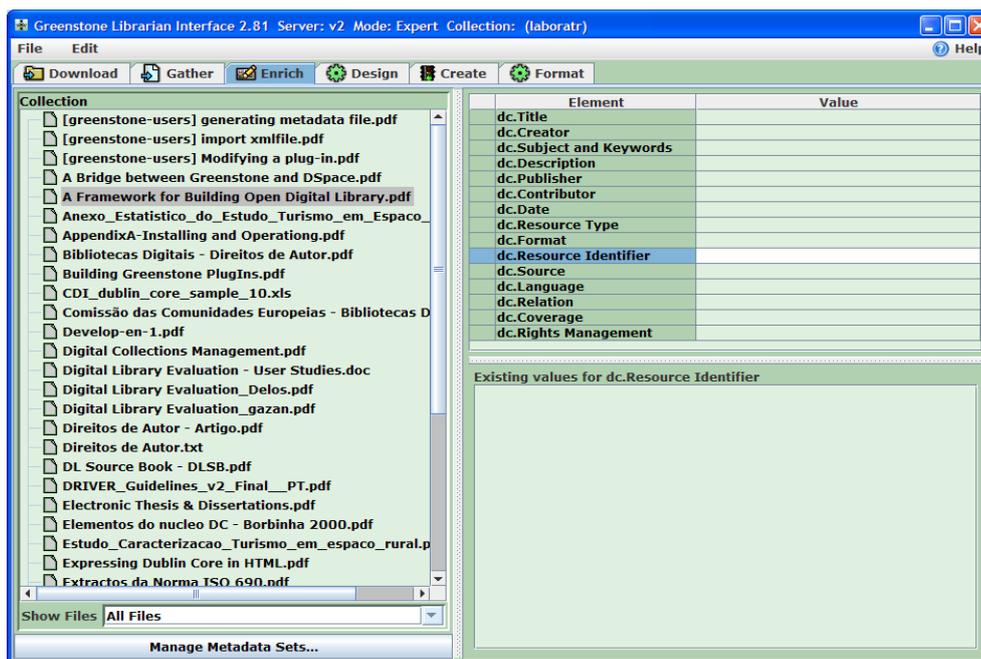


Figura 4-5: Estrutura Dublin Core antes do processo de preenchimento

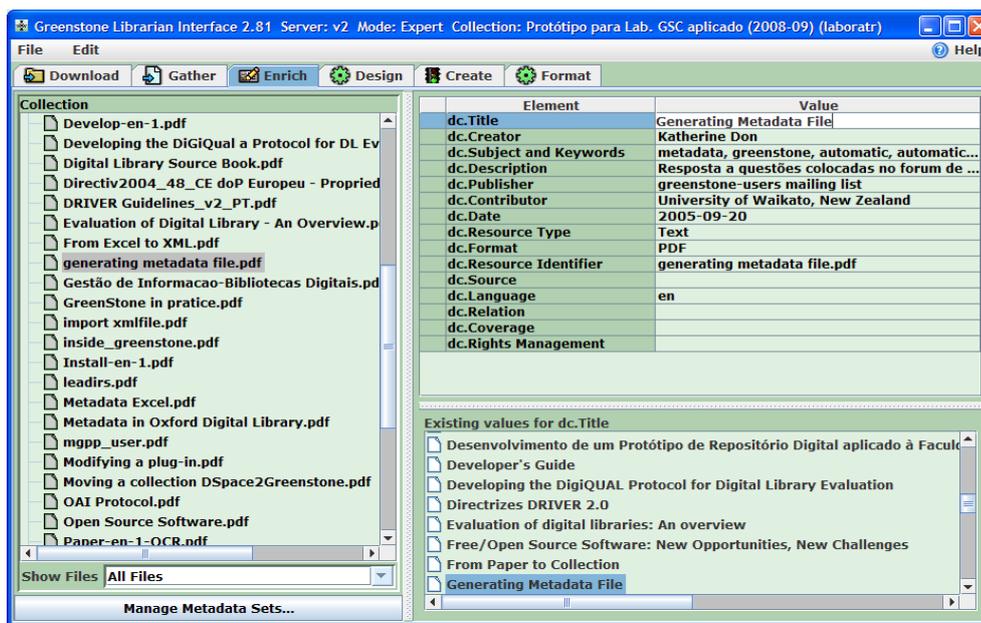


Figura 4-6: Campos de Metadados após preenchimento automático

Na Figura 4-6 encontram-se representados os campos da especificação Dublin Core após o preenchimento recorrendo ao protótipo.

4.4 Personalização do Interface do Utilizador

O Greenstone disponibiliza mecanismos que possibilitam a sua configuração a vários níveis de forma bastante flexível.

O interface do utilizador não é excepção sendo possível a sua a adaptação aos objectivos particulares de cada colecção. Este interface é influenciado por vários factores entre os quais o conteúdo da colecção, o tipo de formato dos documentos e a recolha dos metadados dos recursos constituintes da compilação.

As funcionalidades de pesquisa e navegação no repositório que se pretendam disponibilizar estão directamente ligados ao conteúdo dos metadados. Podem ser criados índices, classificadores e listas baseadas nos campos do esquema de metadados.

De igual modo, a personalização do aspecto das páginas apresentadas ao utilizador pode ser configurado de acordo com as intenções da instituição.

Não está no âmbito deste trabalho a descrição e implementação destes mecanismos, aconselhando-se para tal a literatura de suporte do Greenstone, em particular a descrição da colecção de demonstração incluída no pacote inicial. (Bainbridge, Witten, & al, Greenstone Digital Library : Inside Greenstone Collections, 2003).

5 Trabalhos futuros

Tendo em conta as intenções da instituição de integrar o RCAAP aconselha-se a consulta das directrizes DRIVER 2.0 para a recolha dos metadados. (Digital Repository Infrastructure Vision for European Research, 2008)

Sugerem-se como trabalhos futuros, a melhoria da robustez da ferramenta; tornar o processo de construção da biblioteca mais transparente integrando funcionalidades avançadas e personalizando o interface *standard*; estudar e desenvolver estratégias de migração entre as plataformas D-Space e Greenstone; desenvolver o projecto da instituição de uma forma sustentada, considerando aspectos de persistência dos recursos compilados no repositório e eventualmente a coexistência de ambas as plataformas endereçando requisitos específicos, por exemplo, o D-Space será mais adequado para suportar o repositório de trabalhos científicos devido às suas funcionalidades de gestão de acessos e submissão de conteúdos, enquanto o Greenstone poderá ser mais adequado para colecções mais estáveis de conteúdos com carácter mais social, histórico ou patrimonial, derivado das possibilidades de desenho e adaptação dos esquemas de metadados, flexibilidade de distribuição do repositório e operação em diferentes plataformas.

6 Conclusões

No enquadramento das bibliotecas digitais, o objectivo deste trabalho foi o de apresentar contributos práticos para o desenvolvimento de um repositório digital numa universidade, alinhado com a intenção manifestada pela instituição de promover os trabalhos académicos e científicos elaborados pelos seus graduados utilizando a plataforma Greenstone, partindo de um conjunto de documentos já classificados em ficheiro próprio numa lista em *Microsoft Excel*.

O constrangimento na criação do repositório devido ao procedimento de preenchimento dos metadados bibliográficos para cada documento levou ao desenvolvimento de uma ferramenta auxiliar de migração que veio resolver o problema que deu origem a este trabalho.

Optou-se por abandonar todas as soluções que apontassem para uma alteração ao funcionamento *standard* do Greenstone por exigirem alterações nas rotinas *core* da plataforma e conhecimentos específicos de linguagem de programação (Perl).

Foi consultado o suporte da comunidade de utilizadores do Greenstone tendo sido identificadas orientações úteis que serviram de apoio à investigação, comprovando a efectiva utilidade deste canal de suporte e comunicação.

O protótipo criado apresenta uma arquitectura flexível e pode ser adaptado para outros tipos de metadados. A utilização da especificação Dublin Core foi no entanto suficiente para estudar os principais problemas quando da construção de uma ferramenta de migração de metadados para o Greenstone na realidade da instituição em causa. Serve também de referência a futuros desenvolvimentos.

Como prova de conceito criou-se e inicializou-se em Greenstone, uma biblioteca digital constituída por uma compilação de documentos da base bibliográfica deste trabalho.

O enquadramento das práticas e conceitos consideram-se indispensáveis ao entendimento e desenvolvimento dos objectivos e daí também o tempo empenhado no referido processo.

Bibliografia

Bainbridge, D., & Witten, I. H. (2003). *How to Build a Digital Library*. San Francisco: Morgan Kaufmann.

Bainbridge, D., McKay, D., & Witten, I. H. (2004). *Greenstone Digital Library : Developers Guide*. University of Waikato, New Zeland: New Zeland Digital Library Project.

Bainbridge, D., Witten, I. H., & al, S. B. (2003). *Greenstone Digital Library : Inside Greenstone Collections*. University ok Waikato, New Zeland: New Zeland Digital Library Project.

Benz, B., & Durant, J. D. (2003). *XML Programming Bible*. New York: Wiley.

Biblioteca Nacional Digital. (2006). *Depósito de Dissertações e Teses Digitais*. Obtido em 6 de Junho de 2009, de DiTed- Depósito de Dissertações e Teses Digitais: <http://dited.bn.pt/>

Borgman, C. L. (1999). What are digital libraries? Competing visions. In C. L. Borgman, *Information Processing and Managment* (Vol. 35, pp. 227-243). Los Angeles, USA: Elsevier Science Ltd.

Comissão das Comunidades Europeias. (2005). i2010 : Bibliotecas Digitais. *Comunicação da Comissão ao Parlamento Europeu, ao Conselho, ao Comité Económico e Social Europeu e ao Comité das Regiões*. Bruxelas: Comissão das Comunidades Europeias.

Corporation for National Research Initiatives. (2009). *D-Lib Magazine*. Obtido em Maio, Junho de 2009, de D-Lib Magazine: <http://www.dlib.org/>

Fedora Commons. (2009). *Fedora Commons*. Obtido em Junho de 2009, de Fedora Commons Repository: <http://www.fedora-commons.org/>

Fundação para a Computação Científica Nacional; UMIC Agência para a Sociedade do Conhecimento, IP. (2007). *Biblioteca do Conhecimento Online*. Obtido em 6 de Junho de 2009, de b-on Biblioteca do Conhecimento Online: <http://www.b-on.pt/>

Lesk, M. (2005). *Understanding Digital Libraries*. San Francisco: Morgan Kaufmann.

Library of Congress Network Development and MARC Standards Office. (21 de March de 2008). *MARC STANDARDS*. Obtido em Junho de 2009, de MARC Standards: <http://www.loc.gov/marc/marc.html>

Ministério da Ciência, Tecnologia e Ensino Superior. (2006). *UMIC - Agência para a Sociedade do Conhecimento*. Obtido em 6 de Junho de 2009, de UMIC - Agência para a Sociedade do Conhecimento: <http://www.umic.pt/>

Moxley, J. M. (Number 3 de 2001). Universities Should Require Electronic Theses And Dissertations. *Educause Quarterly* .

Networked Digital Library of Theses and Dissertations. (27 de Julho de 2008). *Networked Digital Library of Theses and Dissertations*. Obtido em 6 de Junho de 2009, de Networked Digital Library of Theses and Dissertations: <http://www.ndltd.org/>

Open Archives Initiative. (s.d.). *Mission Statement*. Obtido em 6 de Junho de 2009, de Open Archives Initiative: <http://www.openarchives.org/OAI/OAI-organization.php>

Rowe, J. (30 de May de 2007). *Greenstone mailing list archives*. Obtido em Junho de 2009, de Greenstone Archives: http://nzdl.sadl.uleth.ca/cgi-bin/library?e=d-00000-00---off-0gsarch-00-0---0-10-0---0---0direct-10---4-----0-11--11-en-50---20-about---00-0-1-00-0-0-11-1-OutfZz-8-00&a=d&c=gsarch&cl=CL1.9.116&d=HASH0187a9866181e84322818b0d_440

Santos, A. M., & Barreto, A. (22 de Fev. de 2007). *Bibliotecas digitais e Direito de Autor: até onde podemos ir?* Lisboa, Lisboa, Portugal.

School of Electronics and Computer, University of Southampton, UK. (2009). *Open Access and Institutional Repositories with EPrints*. Obtido em Junho de 2009, de Open Access and Institutional Repositories with EPrints: www.eprints.org

Tam, S. Y., Wat, S. T., & Kennedy, D. M. (5 de Maio de 2007). *An Evaluation of Two Open Source Digital Library Software Systems*. University of Hong Kong, China.

Tamaro, A. M. (2008). User perceptions of digital libraries: a case study in Italy. (www.emeraldinsight.com/1467-8047.htm, Ed.) *Performance Measurement and Metrics*, Vol. 9 (Nº 2), pp. 130-137.

Taylor, M. (24 de November de 2003). *ZeeRex: The ZeeRex DTD Commentary*. Obtido em 19 de Junho de 2009, de ZeeRex: The ZeeRex DTD Commentary: <http://explain.z3950.org/dtd/commentary.html#5.3>

The Dublin Core Metadata Initiative. (8 de June de 2009). *Dublin Core Metadata Initiative (DCMI)*. Obtido em 15 de Junho de 2009, de About the Dublin Core Metadata Initiative: <http://dublincore.org/>

The Library of Congress. (1 de October de 2007). *Z39.50 Maintenance Agency Page*. Obtido em 6 de Junho de 2009, de Z39.50 International Standard Maintenance Agency: <http://www.loc.gov/z3950/agency/>

The Networked Digital Library of Theses and Dissertations. (s.d.). *Bylaws of The Networked Digital Library of Theses and Dissertations*. Virginia, USA.

The Networked Digital Library of Theses and Dissertations. (Spring de 1996). *ETD: SURA Press Release*. Obtido em 6 de Junho de 2009, de Dissertations, The Networked Digital Library of Theses and: <http://www.ndltd.org/Members/husseini/news/sura>

Witten, I. H., & Boddie, S. (2004). *Greenstone Digital Library : Installer's Guide*. University of Waikato, New Zealand: New Zealand Digital Library Project.

Zhang, A. B. (2008, Novembro 25). Customizing Greenstone User Interface. Washington, USA: Washington Research Library Consortium.

Zhang, A. B., Witten, I. H., Olson, T. A., & Sheble, L. (2006). Greenstone in Practice : Implementations of an Open Source Library System. n.d., USA.

Licenciatura em Gestão de Sistemas e Computação
Universidade Atlântica

**Protótipo de uma biblioteca digital numa universidade:
*Suporte ao processo de inicialização do repositório***

ANEXOS

Projecto Final de Licenciatura

Elaborado por Luís Manuel L. Pinto da Rocha
Aluno nº 20060916
Orientador: Mestre Filipa Taborda

Barcarena
Julho 2009

1

#

2 Índice

ANEXO A: Código fonte do protótipo.....	46
ANEXO B: Campos da estrutura Dublin Core - comentado.....	53
ANEXO C: Exemplo de compilação digital em Greenstone: CD auto-instalável	57

3 ANEXO A: Código fonte do protótipo

```
Imports Microsoft.Office.Interop
```

```
Imports System.Net
```

```
Imports System.Xml.Linq
```

```
Imports System.Xml
```

```
Imports System.IO
```

```
Public Class SMGSDL
```

```
Function StartExcel(Optional ByVal isVisible As Boolean = True) As Excel.Application
```

```
    Dim objExcel As New Excel.Application
```

```
    objExcel.Visible = isVisible
```

```
    Return objExcel
```

```
End Function
```

```
Private Sub Form1_Load(ByVal sender As System.Object, ByVal e As System.EventArgs) Handles MyBase.Load
```

```
    tbDirOrigem.Text = My.Computer.FileSystem.CurrentDirectory
```

```
    fbDialogo.SelectedPath = tbDirOrigem.Text
```

```
End Sub
```

```
Private Sub Button1_Click(ByVal sender As System.Object, ByVal e As System.EventArgs) Handles bProcessar.Click
```

```
    Dim objAppExcel As Excel.Application
```

```
    Dim objLivroExcel As Excel.Workbook
```

```
    Dim objFolhaExcel As Excel.Worksheet
```

```
Dim range As Excel.Range

objAppExcel = StartExcel(True)

objLivroExcel = objAppExcel.Workbooks.Open(tbDirOrigem.Text)

objFolhaExcel = objLivroExcel.Sheets("Metadados")

range = objFolhaExcel.Range("A1",
objFolhaExcel.Range("A1").End(Excel.XlDirection.xlDown).End(Excel.XlDirection.xlTo
Right))

range.Select()

GeraXML(range, tbDirDestino.Text)

objLivroExcel.Close()

objAppExcel.Quit()

MessageBox.Show("Processamento Concluido")

bSair.Enabled = True

End Sub
```

```
Private Sub Button3_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles bExplorePathOrigem.Click

Dim mStream As Stream = Nothing

Dim fichExcel As String

Dim itfFicheiro As New OpenFileDialog()

itfFicheiro.InitialDirectory = "C:\Documents and Settings\1014916\Desktop"

itfFicheiro.Filter = "Excel files (*.xlsx)|*.xls|All files (*.*)|*.*"

itfFicheiro.FilterIndex = 2

itfFicheiro.RestoreDirectory = True

If itfFicheiro.ShowDialog() = System.Windows.Forms.DialogResult.OK Then

Try

fichExcel = itfFicheiro.FileName
```

```
mStream = itfFicheiro.OpenFile()

    If (fichExcel IsNot Nothing) Then
        tbDirOrigem.Text = fichExcel
    End If

Catch Ex As Exception
    MessageBox.Show("Não é possível ler ficheiro. Erro: " &
Ex.Message)

Finally
    ' Nova Verificação de Excepções
    If (mStream IsNot Nothing) Then
        mStream.Close()
    End If

End Try

End If

End Sub
```

```
Private Sub Button2_Click_1(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles bExplorePathDestino.Click
```

```
    Dim mStream As Stream = Nothing
```

```
    Dim fichXml As String
```

```
    Dim itfFicheiro As New OpenFileDialog()
```

```
    itfFicheiro.InitialDirectory = "C:\Documents and Settings\1014916\Desktop"
```

```
    itfFicheiro.Filter = "Xml files (*.xml)|*.xml|All files (*.*)|*.*"
```

```
    itfFicheiro.FilterIndex = 2
```

```
    itfFicheiro.RestoreDirectory = True
```

```
If itfFicheiro.ShowDialog() = System.Windows.Forms.DialogResult.OK Then
    Try
        fichXml = itfFicheiro.FileName
        mStream = itfFicheiro.OpenFile()
        'If (myStream IsNot Nothing) Then
        ' Insert code to read the stream here.
        If (fichXml IsNot Nothing) Then
            tbDirDestino.Text = fichXml
        End If
    Catch Ex As Exception
        MessageBox.Show("Não é possível ler ficheiro de metadados.
Original error: " & Ex.Message)
    Finally
        ' Nova Verificação de Excepções
        If (mStream IsNot Nothing) Then
            mStream.Close()
        End If
    End Try
End If
End Sub
```

```
Function GeraXML(ByVal rngData As Excel.Range, ByVal metaDoc As String)
```

```
    'Const NODE_DELIMITER As String = "/"
```

```
    Const EXT_DELIMITER As String = "."
```

```
    Dim totColunas As Integer
```

```
    Dim totLinhas As Integer
```

```
Dim coluna As Integer

Dim linha As Integer

Dim slitNomeFicheiro() As String

Dim dcFileName As String

Dim rngCell As Excel.Range

Dim docXml As New XmlDocument

Dim elem, elem1 As XmlElement

Dim noTxt As XmlText

Dim atribMode, atribName As XmlAttribute

docXml.Load(metaDoc)

With rngData ' Região selecionada na Folha Excel

    ' Identifica as dimensões dos dados a processar

    totColunas = .Columns.Count

    totLinhas = .Rows.Count

    Dim lstNomesColunas(totColunas) As String ' Lista de cabeçalhos
    (nomes dos campos)

    If totLinhas > 1 Then ' Existem dados ...

        'Processar(cabeçalhos)

        For coluna = 1 To totColunas

            rngCell = .Cells(1, coluna)

            ' Nao aceita celulas agregadas .. Sai

            If Not rngCell.MergeArea.Address = rngCell.Address Then

                MsgBox("!! Celula Agregada ... Formato inválido !!")

            End If

        Next coluna

    End With

End With
```

```
Return 0

End If

lstNomesColunas(coluna) = rngCell.Text

Next

' Processar Metadados

For linha = 2 To totLinhas

    elem = docXml.CreateElement("FileSet")

    elem.InnerXml =
"<FileName></FileName><Description></Description>"

    For coluna = 1 To totColunas

        rngCell = .Cells(linha, coluna)

        ' Nao aceita celulas agregadas .. Sai

        If Not rngCell.MergeArea.Address = rngCell.Address Then

            MsgBox("!! Celula Agregada ... Formato inválido !!")

            Return 0

        End If

        If coluna = 1 Then

            slitNomeFicheiro = Split(rngCell.Text, EXT_DELIMITER)

            dcFileName = slitNomeFicheiro(0) & "\" &
slitNomeFicheiro(1)

            noTxt = docXml.CreateTextNode(dcFileName)

            elem.FirstChild.AppendChild(noTxt)

            elem.AppendChild(docXml.CreateWhitespace(ControlChars.CrLf))

            docXml.DocumentElement.AppendChild(elem)

        Else

            elem1 = docXml.CreateElement("Metadata")

            atribMode = docXml.CreateAttribute("mode")

            atribMode.Value = "accumulate"
```

```
        atribName = docXml.CreateAttribute("name")
        atribName.Value = lstNomesColunas(coluna)
        elem1.Attributes.Append(atribMode)
        elem1.Attributes.Append(atribName)
        elem1.InnerText = rngCell.Text
        elem.FirstChild.NextSibling.AppendChild(elem1)
    End If
Next
elem.AppendChild(docXml.CreateWhitespace(" "))
docXml.DocumentElement.AppendChild(elem)
Next
End If
docXml.Save(metaDoc)
End With
Return 0
End Function
```

```
Private Sub Button4_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles bSair.Click
```

```
    Me.Close()
```

```
End Sub
```

```
End Class
```

--- // ---

4 ANEXO B: Campos da estrutura Dublin Core - comentado

Identificador	<i>Title</i>
Nome	Titulo
Definição	O nome dado ao recurso.
Comentário	O nome pelo qual o recurso é conhecido.

Identificador	<i>Creator</i>
Nome	Criador
Definição	A entidade responsável em primeira instância pela existência do recurso
Comentário	Tipicamente deve indicar uma entidade, pode indicar uma pessoa ou um serviço.

Identificador	<i>Subject</i>
Nome	Assunto
Definição	Tópicos do conteúdo do assunto.
Comentário	Tipicamente, um Assunto deverá ser expresso por palavras-chave, frases ou códigos de classificação que descrevem o conteúdo do assunto. Como boa prática recomenda-se a utilização de vocabulários controlados ou sistemas de classificação formais.

Identificador	<i>Description</i>
Nome	Descrição
Definição	Uma descrição do conteúdo do recurso.
Comentário	Pode incluir um resumo, um índice, uma referência a uma representação gráfica do conteúdo ou uma descrição textual.

Identificador	<i>Publisher</i>
Nome	Editor
Definição	Uma entidade responsável por tornar o recurso acessível.
Comentário	Tipicamente deve indicar uma entidade, pode indicar uma pessoa ou um serviço.

Identificador	<i>Contributor</i>
Nome	Outro Contribuinte
Definição	Uma entidade responsável por qualquer outra contribuição para o conteúdo do recurso.
Comentário	Tipicamente deve indicar uma entidade, pode indicar uma pessoa ou um serviço.

Identificador	<i>Date</i>
Nome	Data
Definição	Uma data associada a um evento do ciclo de vida do recurso.
Comentário	Tipicamente deve estar associada à criação ou disponibilidade do recurso. Uma boa prática a codificação de valores de datas em perfis da norma ISO 8601 segundo o formato (AAAA-MM-DD).

Identificador	<i>Type</i>
Nome	Tipo
Definição	A natureza ou género do recurso.
Comentário	Tipos incluem termos descrevendo categorias genéricas, funções ou níveis de agregação do conteúdo. Recomenda-se o uso de vocabulários controlados, por exemplo “Dublin Core Types”.

Identificador	<i>Format</i>
Nome	Formato
Definição	A manifestação física ou digital do recurso.
Comentário	Tipicamente o formato deve incluir o tipo de meio e dimensões do recurso. Este elemento deve ser usado para determinar quais as aplicações informáticas ou qualquer outro tipo de equipamento necessário para reproduzir ou operar com o recurso. Como boa prática recomenda-se a selecção de valores a partir de vocabulários recomendados, por exemplo MIME.

Identificador	<i>Identifier</i>
Nome	Identificador
Definição	Uma referência não ambígua ao recurso identificada em determinado contexto.
Comentário	Recomenda-se a identificação do recurso por meio de uma cadeia de caracteres ou número de acordo com um sistema de classificação formal (URI, URL, DOI, ISBN)

Identificador	<i>Source</i>
Nome	Fonte
Definição	Uma referência a um recurso de onde o presente recurso possa ter derivado.
Comentário	Recomenda-se a referência ao recurso fonte por meio de um sistema de classificação formal

Identificador	<i>Language</i>
Nome	Língua
Definição	A língua do conteúdo intelectual do recurso.
Comentário	Recomenda-se a utilização do RFC 1766.

Identificador	<i>Relation</i>
Nome	Relação
Definição	Uma referência a um recurso relacionado.
Comentário	Recomenda-se a referência ao recurso relacionado por meio de um sistema de classificação formal.

Identificador	<i>Coverage</i>
Nome	Cobertura
Definição	A extensão ou alcance do recurso.
Comentário	Cobertura inclui tipicamente uma localização espacial (coordenadas geográficas, um local), um período de tempo (a sua designação, data ou intervalo de tempo) ou jurisdição (o nome de uma entidade administrativa). Recomenda-se o uso de vocabulários geográficos controlados (TGN).

Identificador	<i>Rights</i>
Nome	Direitos
Definição	Informação de direitos sobre o recurso ou relativos ao mesmo.
Comentário	Tipicamente este elemento deverá conter uma declaração sobre os direitos deste recurso ou uma referência a uma organização que fornecerá essa informação. Poderá compreender, direitos de propriedade intelectual, direitos de autor ou outros.

--- // ---

#

Elaborado por Luís Manuel L. Pinto da Rocha
Aluno nº 20060916

5 ANEXO C: Exemplo de compilação digital em Greenstone: CD auto-instalável

Parte integrante deste trabalho consiste num CD contendo um exemplo de uma colecção digital produzida em Greenstone constituída por uma compilação de documentos da base bibliográfica consultada.