



Licenciatura em Gestão de Sistemas e Computação

**Biblioteca Digital – Migração de colecção digital para o DSpace: caso
de estudo da Universidade Atlântica**

Projecto Final de Licenciatura

Elaborado por Leonardo Almeida Cruz

Aluno nº 20070975

Orientador: Mestre Filipa Taborda

Barcarena

Fevereiro 2011

Universidade Atlântica

Licenciatura em Gestão de Sistemas e Computação

**Biblioteca Digital – Migração de colecção digital para o DSpace: caso
de estudo da Universidade Atlântica**

Projecto Final de Licenciatura

Elaborado por Leonardo Almeida Cruz

Aluno nº 20070975

Orientador: Mestre Filipa Taborda

Barcarena

Fevereiro 2011

O autor é o único responsável pelas ideias expressas neste relatório

Agradecimentos

O meu primeiro agradecimento é endereçado aos meus pais, que deram todo o apoio necessário para frequentar e concluir a licenciatura.

Quero de seguida agradecer a toda a comunidade académica da Universidade Atlântica. Aos meus colegas com quem convivi muitas horas e onde se criou um espírito de partilha de conhecimento e amizade recomendável a todas as turmas. Aos professores que transmitiram a sua sabedoria e conhecimento, dentro e fora das aulas. Aos auxiliares que, não só durante o horário de expediente como também fora dele, se esforçaram por proporcionar as melhores condições possíveis.

Deixo para o fim um especial agradecimento à Professora Filipa Taborda pela orientação, apoio e motivação inculcada durante a realização de todo o trabalho e a Ricardo Gusmão que sempre que necessário esteve disponível na disponibilização de informação.

Resumo

Biblioteca Digital – Migração de colecção digital para o DSpace: caso de estudo da Universidade Atlântica

Este trabalho tem por objectivo identificar as mais-valias na migração de uma biblioteca digital sobre a plataforma Greenstone para a plataforma DSpace, bem como documentar o respectivo processo. O problema surgiu na perspectiva de melhorar o processo de aquisição de documentos assim como melhorar a gestão da ferramenta de suporte do ponto de vista informático.

Este trabalho foca o contexto da instalação do DSpace no *server farm* da universidade Atlântica e a migração da colecção mantendo a compatibilidade dos metadados definidos pela UAtlântica e os requisitos do RCAAP (Repositório Científico de Acesso Aberto de Portugal).

Palavras-chave: Biblioteca Digital, Migração, DSpace, Greenstone

Índice

Agradecimentos.....	iv
Resumo	vi
Índice.....	viii
Índice de figuras	x
Índice de tabelas	xi
1. Introdução.....	1
2. Enquadramento	2
2.1. Trabalhos anteriores	2
2.2. Possíveis vantagens da migração do Greenstone.....	2
3. O processo de aquisição de documentos na UAtlântica.....	4
3.1. Ambiente existente na UAtlântica	5
3.2. O Greenstone na UAtlântica.....	5
4. Implementação do ambiente de testes Greenstone	15
4.1. Configuração do ambiente de testes	15
4.2. Procedimento de replicação do repositório de produção.....	16
5. O DSpace.....	17
5.1. Implementação do ambiente de testes	17
5.2. A plataforma.....	19
5.3. Parametizações propostas.....	23
6. Migração das colecções.....	27
6.1. Metodologia.....	27
6.2. Problemas e ajustes	29

6.3. Recomendações.....	36
7. Considerações	38
Conclusão	39
Bibliografia.....	40

Índice de figuras

Figura 1 - Repositório Greenstone da UAtlântica: ferramentas utilizadas	9
Figura 2 - Repositório Greenstone da UAtlântica: hierarquia visual.....	10
Figura 3 - Repositório Greenstone da UAtlântica: estrutura hierarquica	11
Figura 4 - Repositório Greenstone da UAtlântica: fases de comunicação	12
Figura 5 - Repositório Greenstone da UAtlântica: estrutura de pastas.....	13
Figura 6 - O DSpace: estrutura do repositório	20
Figura 7 - O DSpace: ligações do item	21
Figura 8 - O DSpace: fluxos de aprovação (<i>Fonte:</i> <i>http://projecto.rcaap.pt/formar/mod5/workflows.html</i>).....	22
Figura 9 - O DSpace: organização proposta.....	24
Figura 10 - O DSpace: fluxo de aprovação proposto (<i>Fonte:</i> <i>http://projecto.rcaap.pt/formar/mod5/workflows.html</i>).....	24
Figura 11 - Exportação e Importação: estrutura necessária para o DSpace	29

Índice de tabelas

Tabela 1 - Esquemas de metadados: campos	14
--	----

1. Introdução

O presente documento não tem por objectivo demonstrar a importância das bibliotecas digitais na universalização da informação, pois esse tema já foi abordado em trabalhos anteriores.

Este documento pretende identificar as vantagens e definir um processo de migração de um repositório digital Greenstone para o DSpace. O repositório digital da Universidade Atlântica vai ser utilizado como caso de estudo.

A falta de flexibilidade ao nível da publicação, pelo facto do processo de edição dos metadados estar centrado no bibliotecário, dificulta a publicação de grandes volumes de informação produzida na universidade e considerada como relevante para o meio académico, pois a submissão de tal volume de informação como por exemplo a publicação de projectos finais de licenciatura (PFL), obrigaria a um esforço de tempo e de recursos humanos.

O problema centra-se na melhoria do processo de aquisição de documentos para o repositório digital da Universidade Atlântica. A solução proposta e que constitui objectivo deste trabalho, passa pela migração da colecção para uma ferramenta que integre todo o processo de aquisição de documentos e que permita a definição de fluxos de aprovação e qualidade de dados a verificar pelo bibliotecário.

O documento final pode servir também como ferramenta para a Universidade Atlântica ou outras instituições, na tomada de decisão em migrar a actual biblioteca digital Greenstone para DSpace e como auxílio no respectivo processo de migração.

O documento começa por identificar as vantagens teóricas de uma migração para a plataforma DSpace passando para a implementação do ambiente de testes, configuração das plataformas e definição do processo de migração. São depois identificados, por parte do autor, os processos e fluxos recomendados e as vantagens e desvantagens da migração.

2. Enquadramento

Este capítulo foca trabalhos anteriores realizados na Universidade Atlântica (UAtlântica) e apresenta as possíveis vantagens que uma migração para a plataforma DSpace pode trazer no actual contexto do caso de estudo da UAtlântica.

2.1. Trabalhos anteriores

O projecto final de licenciatura “*Protótipo de uma biblioteca digital numa universidade*” elaborado por *Luís Manuel L. Pinto da Rocha*, aborda os conceitos de uma biblioteca digital, a plataforma Greenstone e as soluções encontradas para contornar os problemas apresentados pela plataforma. O autor faz ainda uma breve descrição das plataformas Greenstone e DSpace e justifica a opção de escolha pelo Greenstone.

Este trabalho surge também por sugestão do trabalho de projecto final de licenciatura anteriormente referenciado.

2.2. Possíveis vantagens da migração do Greenstone

À medida que a investigação sobre ambas as plataformas foi avançado, tornou-se claro que não competem entre si apesar de alguns objectivos idênticos, pelo contrário, podem inclusive coexistir (Witten et al., 2005).

A identificação de vantagens não é um processo fácil, pois estas dependem largamente dos objectivos definidos a quando da implementação da biblioteca digital.

No DSpace o processo de publicação não é centrado no bibliotecário, esta descentralização permite ao autor ter o seu interface de publicação, facilitando o respectivo processo. Apesar de descentralizado o processo de submissão pode, caso seja pretendido, seguir um fluxo de aprovações. No Greenstone o processo de submissão é centralizado no bibliotecário.

A combinação das vantagens anteriormente identificadas facilita a utilização do repositório digital não só para a publicação de trabalhos do tipo projecto final de licenciatura, como também de outros trabalhos ou apresentações realizados no decorrer

do ano lectivo e que os professores considerem relevantes para o meio académico. Desta forma a universidade permite o acesso a conteúdos que podem ajudar os alunos do presente e do futuro pois a informação irá estar sempre disponível, podendo inclusive ver-se em alguns casos a evolução de estudos realizados em determinadas áreas. Esta solução permitirá ainda aos departamentos de investigação da universidade publicar os seus trabalhos e ensaios, gerindo a informação de forma independente.

A nova visão sobre o repositório científico pode levar à migração de alguma informação existente no Moodle para o DSpace.

O acesso a informação específica pode ainda ser limitado a grupos de utilizadores, dando uma maior flexibilidade na gestão da informação pública e utilização do próprio repositório digital.

Através da utilização de RSS, o DSpace permite aos seus utilizadores estarem a par das actualizadas realizadas no repositório digital.

Podemos ainda referir como vantagens o interface de gestão de conteúdos via *Web browser*, a disponibilização de estatísticas pela própria plataforma, a integração com a AD (*Active Directory*) local e a larga comunidade de utilizadores e programadores existente.

3. O processo de aquisição de documentos na UAtlântica

Não sendo o objectivo deste trabalho documentar processos de instalação e configurações das plataformas, neste capítulo irá ser abordado de uma forma superficial as configurações existentes na UAtlântica e as realizadas para a reprodução do ambiente de testes. É também descrito o actual processo usado para a publicação de informação no repositório digital de modo a perceber o seu fluxo.

Actualmente o processo de publicação de informação na biblioteca digital inicia com o envio de um e-mail por parte do docente com o suporte digital a ser publicado e os metadados a este associados. Por sua vez o responsável do repositório digital recebe a informação enviada e recolhe o ficheiro para a localização de importação do repositório e enriquece também os metadados corrigindo e completando informação em falta. Os metadados ficam então disponíveis para que o bibliotecário também os reveja e os enriqueça novamente. Após os passos anteriormente realizados, alguns em diferentes aplicações e interfaces, o responsável pelo repositório digital tem de compilar toda a informação submetida no repositório e torna-la pública e acessível ao público.

O DSpace permite implementar o processo sucintamente descrito anteriormente numa única ferramenta e interface, tornando todo o processo mais amigo do utilizador. O DSpace permite ainda definir e espelhar todo fluxo do processo, desta forma todos os intervenientes serão avisados via e-mail de que tem uma acção realizar num processo de submissão iniciado e o depositante saberá em qualquer momento em ponto está o seu processo de submissão e caso a sua submissão seja recusada, existe a possibilidade do interveniente que a recusou personalizar o e-mail a ser enviado.

A flexibilidade do DSpace, permite definir diferentes fluxos de submissão para diferentes colecções. Com estas características o DSpace consegue garantir de uma forma harmoniosa a qualidade dos metadados da informação publicada, pois a sua validação pode ser incluída dentro do fluxo do processo de submissão. A descentralização da submissão, com a possibilidade de existirem vários utilizadores no sistema com permissões de submissão e também a existência de vários utilizadores com

permissões de aprovação, permite que a qualquer altura seja iniciado ou concluído um processo de submissão.

Os utilizadores terão as suas permissões de submissão e aprovação configuradas à colecção caso seja necessário, podendo assim haver utilizadores com permissões diferentes nas colecções às quais têm acesso. Pode ainda ser limitado o acesso de visualização de determinadas colecções que se pretenda, a utilizadores específicos.

3.1. Ambiente existente na UAtlântica

Foi realizada uma pequena entrevista a Ricardo Gusmão, administrador de sistemas sénior da UAtlântica. Com esta entrevista foi possível perceber como se encontra implementada a plataforma Greenstone.

A plataforma Greenstone pode ser instalada sobre uma plataforma Web existente como o IIS ou Apache, ou utilizando o seu próprio motor Web que é assente em Apache. Na UAtlântica foi utilizada a instalação utilizando o motor Web do Greenstone. Esta opção levou à existência de alguns problemas na disponibilidade do serviço, este costuma encontrar-se algumas vezes em baixo, sendo necessário reiniciar o serviço para que se possa voltar a ter acesso.

O servidor onde a plataforma se encontra instalada está assente sobre o sistema operativo Windows 2003 e está apenas dedicado ao Greenstone.

3.2. O Greenstone na UAtlântica

A identificação do processo de publicação e fluxo foi documentado através de entrevista realizada com a professora Filipa Taborda, responsável pela publicação do repositório digital.

O actual processo de publicação do repositório digital da UAtlântica passa por diferentes fases de intervenção manual e automática, intervenientes e aplicações. Este processo, como iremos ver, pode ser simples em algumas fases mas em outras o nível técnico de intervenção manual implica que os seus intervenientes possuam conhecimentos técnicos mais elevados para além de as de um utilizador normal.

O processo descrito a seguir é despoletado pela necessidade de se publicar documentos de projecto final de licenciatura.

Processo de publicação (Fluxo):

1ª Fase:

- Responsável: docente da cadeira;
- Descrição: preenchimento e envio por e-mail de formulário em PDF com dados sobre o trabalho, juntamente com o respectivo documento de projecto final de licenciatura.

2ª Fase:

- Responsável: responsável pela publicação do repositório digital;
- Descrição: o formulário enviado no ponto anterior é aberto no Acrobat Writer e colocado no *Form Dataset* e o documento na pasta *import* dentro da estrutura de pastas do Greenstone.

3ª Fase:

- Responsável: funcionário da biblioteca.
- Descrição: acede ao formulário no *Form Dataset* através do Acrobat Writer e completa/corrigue os metadados de modo a que fiquem compatíveis com as normas de classificação das bibliotecas e com o RCAAP (Repositório Científico de Acesso Aberto de Portugal).

4ª Fase:

- Responsável: responsável pela publicação do repositório digital;
- Descrição: acede ao formulário no *Form Dataset* através do Acrobat Writer e exporta os dados para XML;

5ª Fase:

- Responsável: responsável pela publicação do repositório digital;
- Descrição: importa o ficheiro XML para o Excel e organiza os campos de forma hierárquica através de um dos campos dos metadados, espelhando assim a estrutura hierárquica das colecções existentes no Greenstone.

6ª Fase:

- Responsável: responsável pela publicação do repositório digital;
- Descrição: executa a aplicação ‘Serialização de metadados em Greenstone’, desenvolvida pelo aluno Luís Rocha no âmbito do trabalho de projecto final de licenciatura, e gera um ficheiro XML com formato compatível com o Greenstone.

7ª Fase:

- Responsável: responsável pela publicação do repositório digital;
- Descrição: edita o XML gerado na fase anterior e copia o seu conteúdo para o fim do ficheiro *metadata.xml* existente na pasta *import* dentro da estrutura de pastas do Greenstone.

8ª Fase:

- Responsável: responsável pela publicação no repositório digital;
- Descrição: acede ao interface de publicação do Greenstone e executa a fase *Compile*, verificando se as fases *Import* e *Gather* estão correctas.

O fluxo descrito é particularmente útil a quando da publicação de vários documentos ao mesmo tempo, permitindo reduzir o tempo de publicação. A quando da publicação de apenas um documento não existem ganhos de tempo em contrapartida de utilizar apenas o interface de publicação do Greenstone.

Na descrição do fluxo, o responsável pela publicação do repositório digital corresponde à figura de bibliotecário do Greenstone, sobre o qual o processo se encontra centralizado.

Existem algumas particularidades que merecem destaque no fluxo descrito anteriormente. As ferramentas utilizadas, a forma como a informação aparece estruturada de forma hierárquica no Greenstone, sendo este ponto particularmente importante para o processo de importação da colecção para o DSpace e perceber também de que forma se encontra o Greenstone actualmente organizado e estruturado. A forma e os momentos em que a comunicação entre os intervenientes é feita durante todo o fluxo também é importante para perceber as melhorias que o DSpace pode trazer em termos de comunicação. A estrutura de pastas do Greenstone e o esquema de metadados utilizado merecem também uma breve descrição e explicação para melhor compreender algumas fases do fluxo descrito.

Apesar de alguns pontos deste subcapítulo poderem ter sido apresentados anteriormente, foram propositadamente apresentados após a descrição do fluxo pois desta forma o leitor não se encontrava preocupado em compreender detalhes técnicos mas sim focado apenas no fluxo descrito.

É importante perceber a forma distribuída em termos de ferramentas utilizadas como o processo de submissão decorre. Desde o seu início até ao momento em que é concluído são utilizadas diversas ferramentas, como é visível na Figura 1.

No processo actualmente definido na UAtlântica existem ferramentas extras usadas no processo de modo a permitir a publicação da informação em lote, diminuindo assim o número de passos a repetir. O processo simples de publicação do Greenstone apenas necessitaria das ferramentas: cliente de e-mail, formulário em PDF e o interface gráfico do bibliotecário. O DSpace consegue simplificar o processo todo unificando tudo num único interface Web.

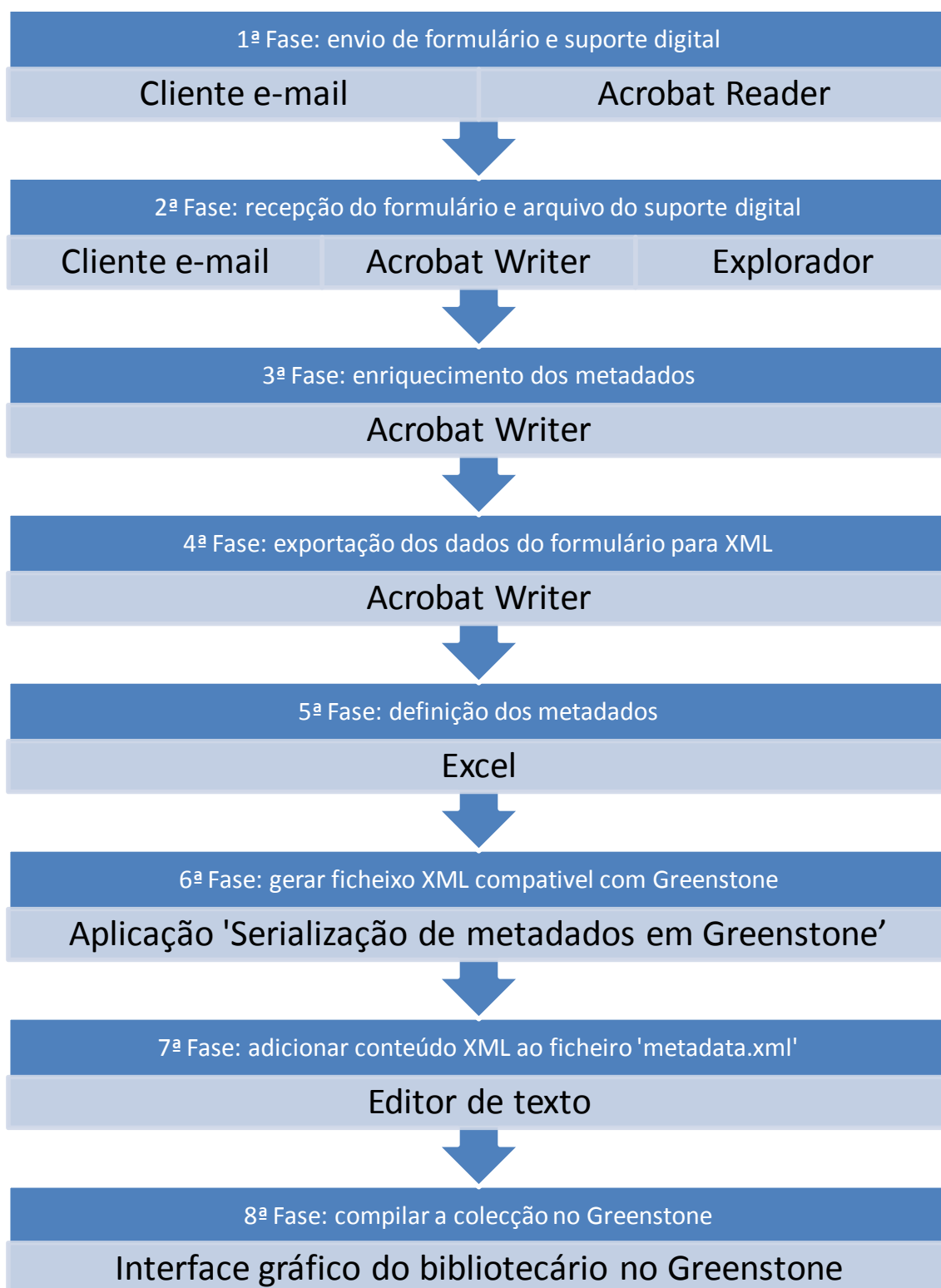


Figura 1 - Repositório Greenstone da UAtlântica: ferramentas utilizadas

O actual repositório digital em Greenstone da UAtlântica, não se encontra organizado em várias colecções ou subcolecções apesar de ao navegarmos no mesmo ser transmitida a noção de hierarquia, o repositório é constituído apenas por uma única colecção. Na fase 5 do fluxo de publicação, o responsável pela publicação do repositório enriquece os metadados preenchendo o campo que nos transmite esta sensação de hierarquia (Figura 2). O campo do esquema de metadados utilizado para implementar a funcionalidade descrita é o *Area e Curso* (Tabela 1 - Esquemas de metadados: campos). Para completar o efeito pretendido é ainda utilizada a propriedade formatação por HTML do Greenstone.

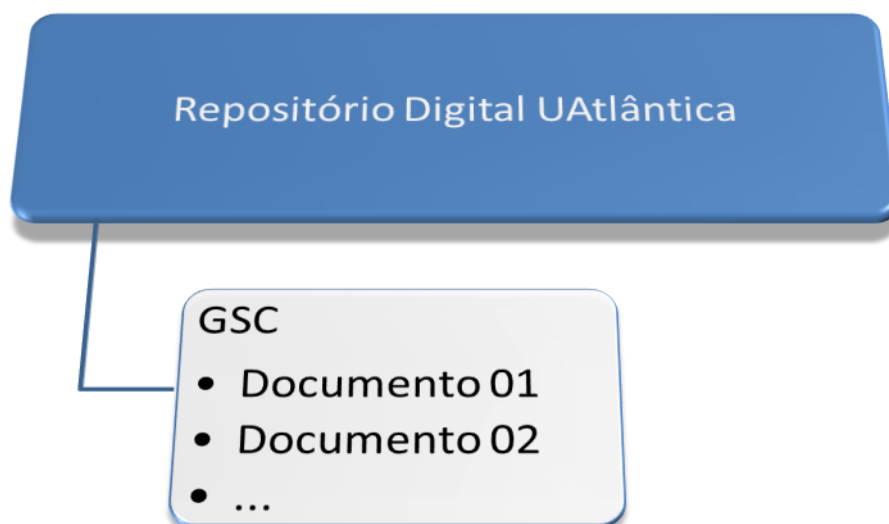


Figura 2 - Repositório Greenstone da UAtlântica: hierarquia visual

A Figura 3 mostra a estrutura hierárquica existente no repositório actual. Existe apenas uma única colecção e logo abaixo estão todos os documentos publicados. Apesar da configuração visual definida (Figura 2), o Greenstone permite a definir hierarquias entre colecções utilizando subcolecções, criando desta forma uma estrutura hierárquica. Como iremos ver mais tarde o DSpace tem uma estrutura hierárquica mais rica, permitindo a definição de fluxos e permissões de acesso, submissão e aprovação independentes.



Figura 3 - Repositório Greenstone da UAtlântica: estrutura hierárquica

O fluxo do processo de submissão de informação no repositório passa por vários intervenientes, onde a celeridade do processo e a própria conclusão depende também da eficiência da comunicação existente. A plataforma actual não tem qualquer automatismo de comunicação, alertando dessa forma os utilizadores de que têm de intervir num processo de submissão. Não existe também nenhum local onde os intervenientes possam visualizar o estado de cada pedido de submissão.

A Figura 4 apresenta a sequência de comunicação existente, não estando incluídas todas as comunicações possíveis, como a comunicação de que a informação foi publicada ou de que foi recusada, ou mesmo a solicitação de mais dados necessários para completar os metadados existentes.

O DSpace por sua vez automatiza a comunicação substituído, não na sua totalidade, a intervenção humana e garante assim a comunicação efectiva nas várias fases do fluxo e diminui assim a falhas de comunicação por esquecimento ou falta de tempo do interveniente. A comunicação no DSpace vai ainda mais longe, através da utilização de RSS, alertando os utilizadores para a publicação de nova informação nas colecções nas quais estão inscritos.

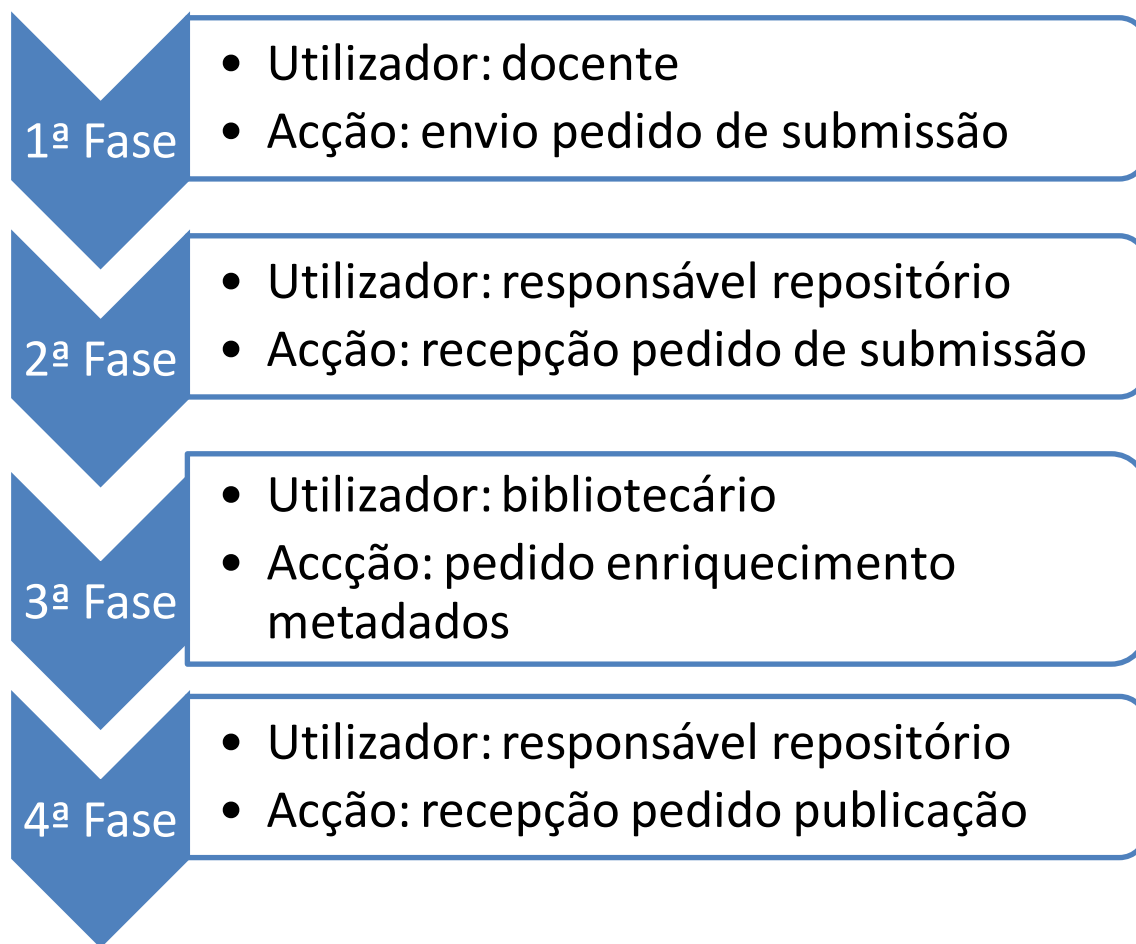


Figura 4 - Repositório Greenstone da UAtlântica: fases de comunicação

O Greenstone é composto por uma estrutura de pastas hierárquica. Nesta existe a pasta *import* que por sua vez está dentro da pasta da respectiva colecção e é a mais importante em todo o processo de migração para o DSpace e implementação do ambiente de testes. É nesta pasta que estão todos os suportes digitais correspondentes aos pedidos de publicação realizados, bem como o ficheiro *metadata.xml* que contem todos os metadados de todos os itens. Esta pasta permite reconstruir o repositório as vezes que forem necessárias mas não corresponde à pasta onde os ficheiros compilados são armazenados para disponibilização ao utilizador, essa pasta é a *archives*, onde existe uma pasta por cada item publicado e dentro desta estão os suportes digitais e um ficheiro que contem os metadados correspondentes bem como outras configurações. Existe ainda a pasta *metadata* que contem os ficheiros de esquema de metadados (*.mds) usados na colecção e ainda o ficheiro *profile.xml* onde é realizado o

mapeamento entre esquemas de metadados usados dentro da colecção, desta forma caso algum dos esquemas seja removido e existam campos preenchidos, estes são transferidos para os campos mapeados. Este ficheiro, como iremos ver, vai ser útil no processo de exportação para o DSpace. Na Figura 5 podemos ver a estrutura de pastas descrita.



Figura 5 - Repositório Greenstone da UAtlântica: estrutura de pastas

O esquema de metadados utilizado no repositório da UAtlântica é baseado no esquema de metadados Dublin Core Não Qualificado ou Simple. Este esquema é constituído por 15 campos base. A este esquema base foram adicionados campos extra, considerados necessários a quando da implementação do repositório na UAtlântica.

O nome do esquema de metadados da UAtlântica é ‘*ua*’, este nome é também o prefixo de cada campo e ajuda a identificar o esquema de metadados, por exemplo o esquema Dublin Core tem o nome ‘*dc*’. É importante perceber a forma como o nome do campo é composto, pois irá ser importante no processo de exportação e importação para o DSpace. Podemos tomar como exemplo: o campo com o nome *titulo* vai ter como nome

completo: *dc.titulo*, se criado no esquema Dublin Core ou *ua.titulo* se criado no esquema da UAtlântica. A Tabela 1 descreve os campos existentes nos esquema da UAtlântica e Dublin Core, mostrando também o mapeamento entre ambos. É também possível ver que apesar de o mapeamento existir os nomes dos campos são diferentes.

Esquemas de metadados	
UATLA	Dublin Core
UA.CDU	DC.IDENTIFIER
UA.Origem	DC.SOURCE
UA.AssuntoOrig	
UA.Area e Curso	
UA.SubArea	
UA.Ciclo	
UA.Tipo	
UA.Ano	DC.COVERAGE
UA.Data	DC.DATE
UA.Titulo	DC.TITLE
UA.Autor	DC.CREATOR
UA.EMAutor	
UA.Orientador	DC.CONTRIBUTOR
UA.PalavrasChave	DC.SUBJECT
UA.Formato	DC.FORMAT
UA.Descricao	DC.DESCRPTION
UA.Editor	DC.PUBLISHER
UA.Type	DC.TYPE
UA.Lingua	DC.LANGUAGE
UA.Relacao	DC.RELATION
UA.Direitos	DC.RIGHTS

Tabela 1 - Esquemas de metadados: campos

4. Implementação do ambiente de testes Greenstone

Este capítulo descreve o modo como foi implementado o ambiente de testes Greenstone e os passos realizados para replicar o repositório do ambiente de produção.

4.1. Configuração do ambiente de testes

A instalação do Greenstone no ambiente de testes foi realizada sobre o sistema operativo Windows 2008 ao contrário do identificado no ambiente da UAtlântica utilizando o motor Web do próprio Greenstone, tal como identificado no ambiente da UAtlântica. O facto de se utilizar o motor de Web do Greenstone levou a que a escolha da versão do sistema operativo Windows fosse irrelevante.

A versão instalada do Greenstone foi a 2.83, a mais recente na altura da elaboração deste documento. Tal como listado nas características da plataforma, a instalação é simples, directa e rápida. O processo de instalação pode ser realizado por um utilizador com conhecimentos básicos de informática.

Com a plataforma de testes instalada é agora necessário exportar a colecção existente no repositório digital da UAtlântica para de seguida importar a mesma no ambiente de testes criado. Este procedimento irá permitir identificar os problemas que poderão surgir a quando da migração da colecção do Greenstone para o DSpace, isto porque os metadados definidos na colecção da UAtlântica são diferentes dos existentes numa colecção de exemplo de uma instalação nova do Greenstone, não permitindo assim maximizar a identificação de possíveis erros no processo de migração.

Como vimos anteriormente, os metadados do repositório digital da UAtlântica são compatíveis com os metadados Dublin Core e a estes foram adicionados outros considerados necessários a quando da implementação do repositório. Sabemos também que o nome da estrutura de dados criada no Greenstone tem o nome de ‘*ua*’, compondo assim os nomes dos campos na forma *ua.<nome do campo>*.

Não esquecendo a definição anterior, passemos ao processo de exportação do Greenstone.

4.2. Procedimento de replicação do repositório de produção

O procedimento definido neste documento corresponde a um processo bastante simples, pois não é necessário reproduzir o interface gráfico, a navegação, a indexação, etc... a única característica necessária reproduzir é o conteúdo da colecção e os seus metadados. Para isso foi realizada uma cópia directa da pasta *import* e do ficheiro *ua.mds* para a pasta *metadata*. O processo de exportação e importação foi definido da seguinte forma:

1. Criar colecção nova no ambiente de testes;
2. Substituir a pasta *import* da colecção criada no ponto anterior pela pasta *import* da colecção no ambiente de produção;
3. Copiar o ficheiro *ua.mds* da colecção do ambiente de produção para a colecção do ambiente de testes;
4. Abrir colecção criada no ambiente de testes e:
 - a. Confirmar a existência dos ficheiros no separador *Gather*;
 - b. Confirmar a existência da estrutura de metadados ‘*ua*’ no separador *Enrich*;
 - c. Opcionalmente no separador *Design* corrigir os parâmetros *Partition Indexes* e *Browsing Classifiers* alterando os campos usados para os correspondentes na estrutura de dados ‘*ua*’;
5. No separador *Create*, realizar o *Complete Rebuild* por forma a compilar a colecção por completo.

Este processo pode levar à compilação de alguns campos com caracteres ‘estranhos’ como símbolos asiáticos, mas para o âmbito deste documento é um problema que não tem qualquer impacto para a posterior definição do processo de exportação para o DSpace. Este problema está relacionado com o modo de funcionamento do Greenstone que utiliza um conjunto de metadados de sistema (‘*ex.<nome do campo>*’) e que mapeia com alguns campos da estrutura de dados usada no item. Em alguns itens o mapeamento não é feito correctamente. A exportação para o DSpace do ambiente de produção será directa e este problema não será replicado.

5. O DSpace

Neste capítulo são abordadas as características do DSpace consideradas como principais pelo autor no âmbito do trabalho, passando por parametrizações sugeridas. Este capítulo aflora ainda de uma forma superficial o processo de instalação e configuração do DSpace sobre um ambiente semelhante ao existente na UATLA.

5.1. Implementação do ambiente de testes

Ao contrário do Greenstone o DSpace não pode ser facilmente instalado por qualquer utilizador. O processo de instalação do DSpace obriga a que o utilizador tenha conhecimentos sólidos na administração de sistemas e preferencialmente também bases de dados T-SQL para poder realizar a instalação, configuração, despiste e resolução de problemas que possam ocorrer durante todo o processo. Esta particularidade obriga a que a universidade tenha o capital humano com os conhecimentos necessários.

A solução de testes foi instalada sobre o sistema operativo Windows 2008 e sobre a plataforma Web Apache.

A versão do DSpace instalada foi a 1.7, correspondente à última versão estável disponível à data da realização deste trabalho.

Como o processo de instalação e configuração não faz parte do âmbito deste trabalho, deve ser consultada a documentação necessária do DSpace para melhor identificar e perceber os pré-requisitos. Na bibliografia deste trabalho podem ser encontradas as referências para os documentos utilizados para a instalação da solução de testes. Existem passos que não se encontram descritos nesses documentos e que foram identificados como sendo necessários para o bom funcionamento da aplicação e que estão subjacentes como fazendo parte dos conhecimentos técnicos de quem realiza a instalação e configuração da solução.

Para a implementação da solução foram instaladas e configuradas as seguintes aplicações necessárias como pré-requisitos:

- Oracle Java JDK 6 ou superior;
- Apache Maven 2.0.8 ou superior;
- Apache Ant 1.7 ou superior;
- PostgreSQL 8.2, 8.3 ou 8.4;
- Apache Tomcat 5.5 ou superior;
- DSpace 1.7.

Os três primeiros requisitos são necessários para a compilação e configuração da plataforma DSpace, bem como para futuras actualizações que sejam necessárias realizar.

A plataforma de dados seleccionada para o ambiente de testes foi o PostgreSQL em detrimento do Oracle. Esta opção foi tomada de modo a manter o requisito inicial a quando da implementação da plataforma na UAtlântica referente a custos reduzidos. As duas plataformas de dados anteriormente mencionadas são as únicas suportadas pelo DSpace. Para facilitar a gestão da plataforma de dados escolhida foi ainda instalada a ferramenta de gestão pgAdmin III.

A plataforma Web escolhida foi Apache TomCat. As outras plataformas recomendadas para utilização com o DSpace são: Jetty e Caucho Resin. Foi escolhida a Apache TomCat por ser a que o autor melhor conhece e domina.

A versão do DSpace escolhida foi a última, podendo obviamente ter sido escolhida qualquer uma das outras, optou-se no entanto por realizar testes com a última versão disponível de modo a se poder testar, caso necessário, todas as funcionalidades disponíveis no DSpace à altura da elaboração deste trabalho.

Todos os requisitos listados referem as versões mínimas com as quais a versão do DSpace 1.7 funciona. Para a solução de testes foram obtidas, instaladas e configuradas as últimas versões estáveis de todos os requisitos para o Windows:

- Oracle Java SDK 6 actualização 23;
- Apache Maven 3.0.2;
- Apache Ant 1.8.2;
- PostgreSQL 8.3.14-1;
- Apache Tomcat 7.0.8.

A instalação dos requisitos foi feita utilizando as pré-definições dos pacotes de instalação e realizada com a seguinte ordem:

1. Oracle Java SDK;
2. Apache Maven;
3. Apache Tomcat;
4. Apache Ant;
5. PostgreSQL.

5.2. A plataforma

Ao contrário do Greenstone o DSpace não está apenas organizado em colecções e itens. A estrutura hierárquica do DSpace é mais rica mas ao mesmo tempo simples de compreender e permite uma maior flexibilidade na utilização do repositório.

O DSpace está organizado em comunidades, colecções, ficheiros e item. As comunidades podem corresponder a departamentos, órgãos, unidades, etc.. Dentro de cada comunidade são publicados ficheiros identificados por um item que contém os metadados de todos os ficheiros a ele associado. Os itens encontram-se por sua vez agrupados em colecções (Figura 6).

Aliado à estrutura do DSpace estão as permissões. As permissões vão desde acesso apenas para leitura à permissão de publicação ou aprovação. As permissões de acesso podem ser dadas verticalmente permitindo, caso seja necessário, dar permissões específicas ao nível dos itens publicados.

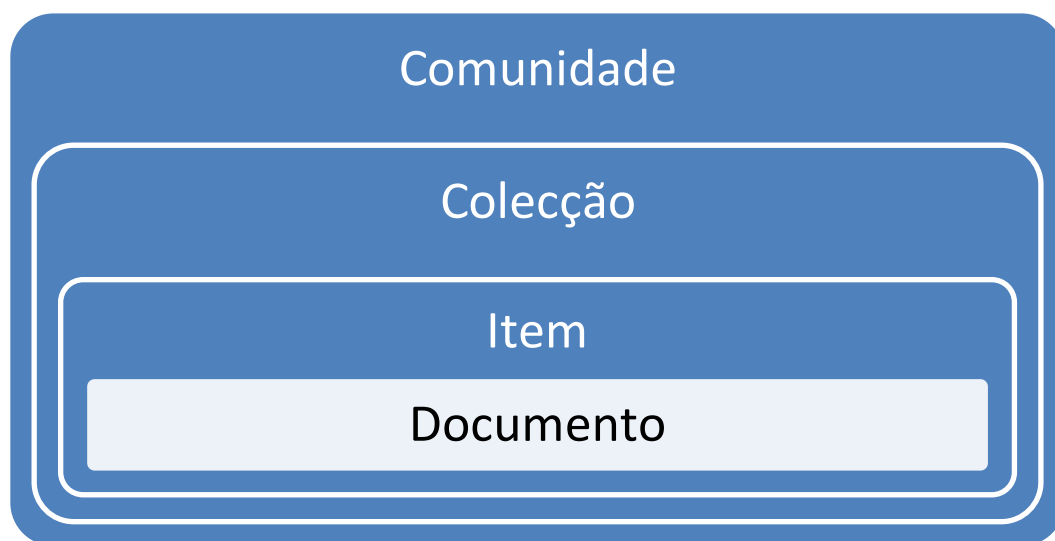


Figura 6 - O DSpace: estrutura do repositório

Um item publicado pode estar ligado a várias colecções, permitindo a partilha de informação entre utilizadores que tenham acesso a colecções diferentes (Figura 7). No Greenstone, um item para pertencer a várias colecções tem de ser publicado em cada uma, não permitindo uma real partilha do item, pois as alterações realizadas numa colecção não se reflectem nas restantes até que seja realizada uma actualização manual em cada colecção.

Cada ficheiro pertencente ao item tem um identificador único. Esta funcionalidade permite utilizar um URL persistente como referência em trabalhos, publicações, etc., pois mesmo alterando o nome do documento ou a colecção à qual pertence a ligação não se perde, esta só se perde com a remoção efectiva do documento do repositório. No Greenstone a ligação é feita com base no nome do documento, todas as alterações realizadas ao nome ou localização têm impacto nas ligações criadas para o documento, originando ligações quebradas.

Ao nível utilizadores, estes podem ter restrições e papéis diferentes dentro da plataforma e para facilitar a configuração são agrupados em grupos, sendo possível dar permissões ao grupo e indirectamente ao utilizador.

Os utilizadores podem submeter os seus próprios documentos ou delegar em outros utilizadores com permissões de submissão nas colecções em causa. É esta característica

que permite ao DSpace, ao contrário do Greenstone, gerir o processo de submissão de uma forma descentralizada. O processo de submissão é ainda composto por uma sequência de passos que obriga ao preenchimento dos metadados, aceitação de licenças e outros passos que se pretendam adicionar. De forma a minimizar os erros de preenchimento dos metadados é também possível restringir os valores de preenchimento de alguns metadados a valores existentes numa lista pré-definida. No Greenstone existe apenas a possibilidade de seleccionar o valor de uma lista de valores já utilizados, *auto-complete*, mas não é restritivo.

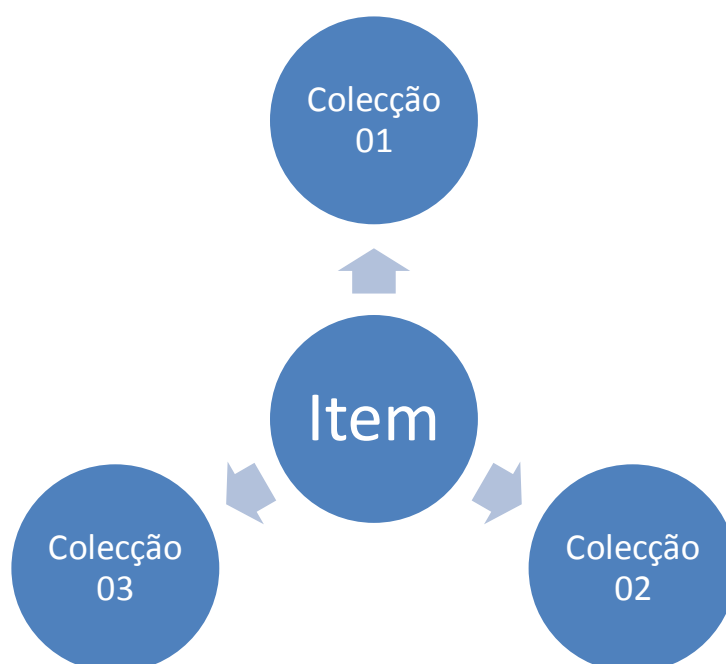


Figura 7 - O DSpace: ligações do item

Os fluxos de submissão podem ser definidos por colecção, sendo assim possível configurar um fluxo por cada colecção existente ou simplesmente reutilizar. Estes fluxos permitem aumentar a qualidade dos metadados preenchidos, pois podem ser validados pelo número de utilizadores pretendidos antes da publicação.

Os fluxos de submissão podem ir do mais simples ao mais complexo. O mais simples passa por uma submissão e aprovação final, num fluxo mais complexo podem existir passos de aprovação intermédios, enriquecimento dos metadados por outros utilizadores e por fim a aprovação final. A Figura 8 exemplifica alguns fluxos de aprovação

possíveis, indo do mais simples ao mais complexo. Os fluxos definidos devem corresponder às necessidades de cada repositório e de cada colecção, podendo ser alterados ao longo do tempo.

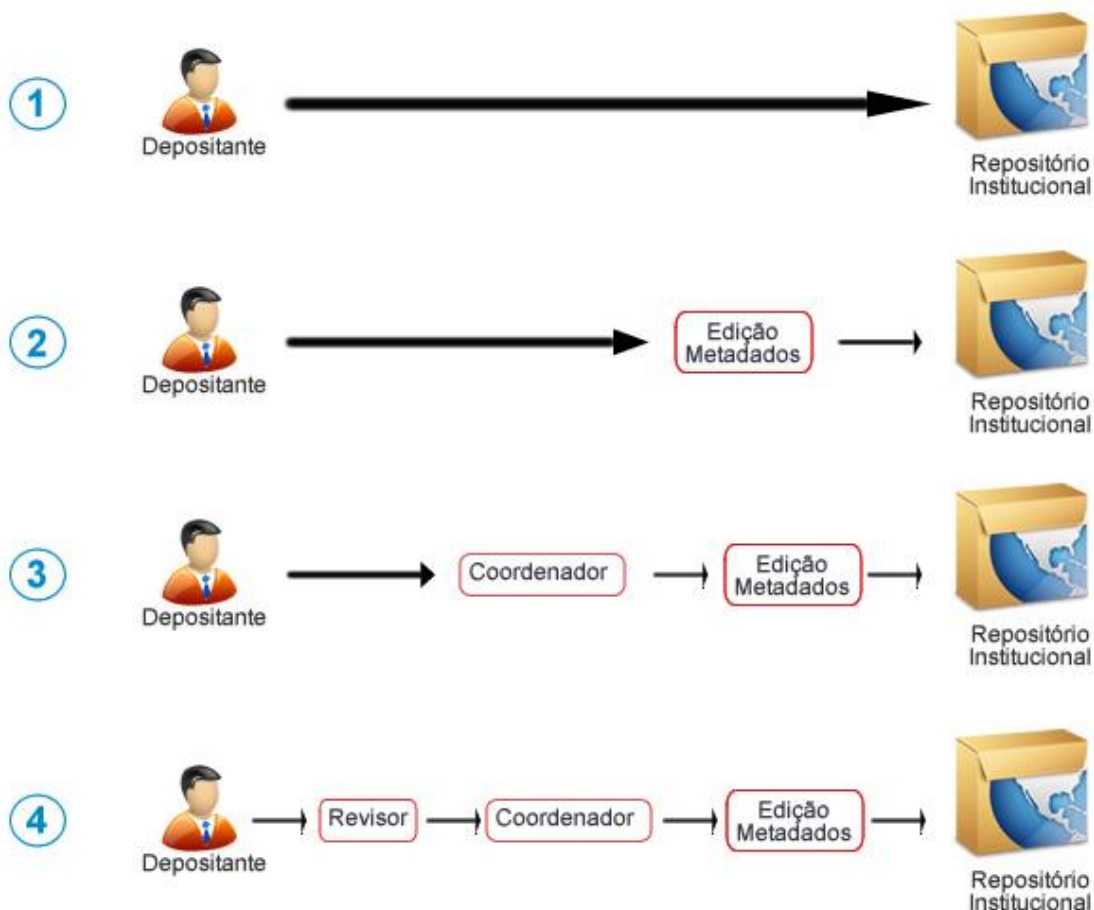


Figura 8 - O DSpace: fluxos de aprovação (Fonte: <http://projecto.rcaap.pt/formar/mod5/workflows.html>)

O fluxo de submissão é composto por vários intervenientes:

Depositante:

- O depositante corresponde a um utilizador com permissões de depósito na colecção na qual pretende iniciar o processo de submissão. O depositante pode ou não ser o autor do documento a ser depositado (ex.: aluno).

Revisor:

- O revisor pode rejeitar o pedido de submissão iniciado, interrompendo o processo de submissão e comunicando a sua decisão via e-mail (ex.: docente).

Coordenador:

- O coordenador tem as mesmas permissões que o revisor, podendo ainda editar os metadados (ex.: docente).

Editor de Metadados:

- O editor de metadados tem permissões para alterar os metadados e após a sua validação o documento passa a estar publicado no repositório (ex.: bibliotecário).

5.3. Parametrizações propostas

As parametrizações recomendadas são referentes a funcionalidades consideradas necessárias, pelo autor, à forma como o DSpace deve ser organizado. No capítulo seguinte são realizadas recomendações referentes à configuração do sistema.

Navegando um pouco pelos diversos repositórios digitais a nível nacional como o RCAAP e de instituições universitárias como a Universidade do Minho, parceira do projecto DSpace, propõe-se para o caso de estudo da UAtlântica a seguinte organização a nível de comunidades e colecções para a publicação de projectos finais de licenciatura:

- Uma comunidade por cada unidade de ensino;
- Uma subcomunidade por cada curso;
- Uma colecção para os projectos finais de licenciatura.

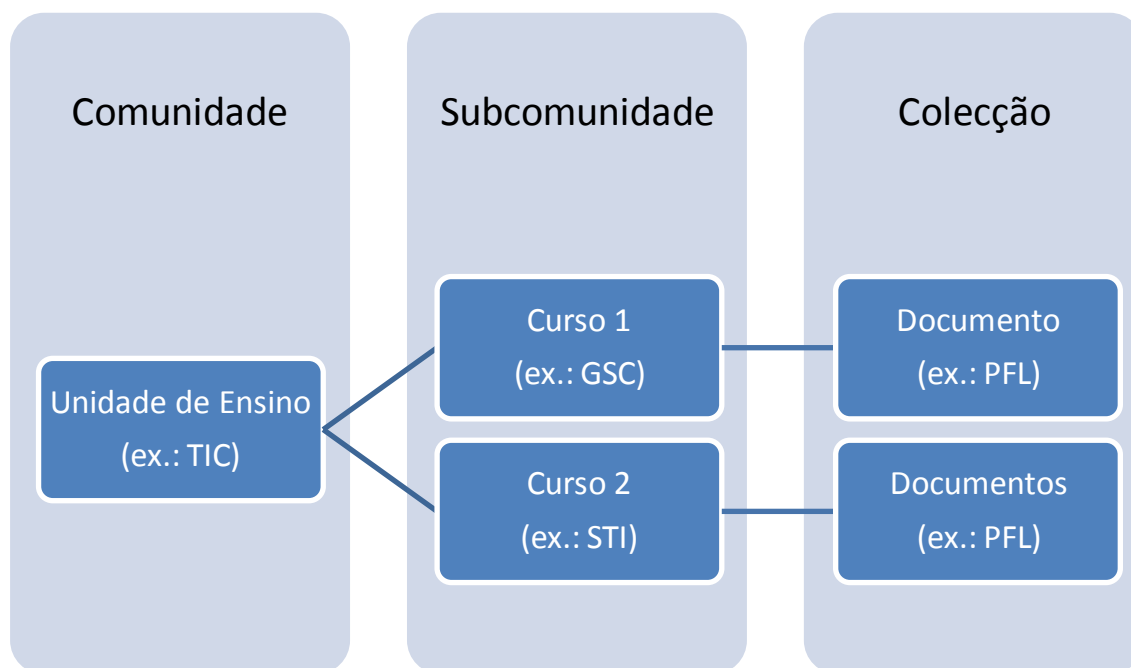


Figura 9 - O DSpace: organização proposta

A organização proposta permite espelhar o actual repositório e em termos futuros adicionar colecções para mestrados ou comunidades para departamentos de investigação, biblioteca, etc.. Permite também reutilizar fluxos de aprovação pois existem colecções com a mesma lógica de conteúdo e utilização e ainda aplicar diferentes fluxos de aprovação consoante o conteúdo da colecção.



Figura 10 - O DSpace: fluxo de aprovação proposto (Fonte: <http://projecto.rcaap.pt/formar/mod5/workflows.html>)

Uma das propostas principais está relacionada com o fluxo de aprovação. O fluxo actual da UAtlântica, descrito no capítulo **Erro! A origem da referência não foi encontrada.**, pode ser alterado para o fluxo da Figura 10. Este fluxo está relacionado com a submissão de documentos referentes a projectos finais de licenciatura. Fluxo proposto:

1. Aluno inicia pedido de submissão;
2. Docente recebe informação de pedido de submissão e avalia o documento e metadados. Caso necessário pode alterar os metadados;
3. Os utilizadores do grupo da biblioteca recebem informação de pedido de submissão e validam os metadados actualizando se necessário;
4. Após validação no passo anterior o documento passa a estar publicado no repositório.

Comparativamente ao processo actual o ganho em termos do número de fases necessárias é substancial. A possibilidade da validação final dos metadados poder ser feita pelos diversos utilizadores pertencentes ao grupo de utilizadores da biblioteca, permite agilizar o processo de submissão.

O processo proposto levanta um problema a nível de manutenção do sistema, os utilizadores que iniciam o pedido de submissão correspondem aos alunos, o que implica adicionar cada aluno que necessite de submeter um projecto final de licenciatura ao grupo de utilizadores com as permissões necessárias na colecção. Este problema pode ser contornado utilizando a integração do DSpace com a *Active Directory* e a utilização de scripts de SQL, para tal deve ser possível identificar o grupo de submissão da colecção por exemplo com uma nomenclatura de nomes onde o nome do curso e a tipo de permissão seja incluído (ex.: COLLECTION_TIC-GSC_SUBMIT). Deve também ser possível identificar o curso ao qual o aluno pertence por exemplo através da consulta de uma lista externa ao DSpace como o Moodle ou a *Active Directory* onde essa informação esteja disponível.

O fluxo de submissão inicia-se com o preenchimento da submissão em si. Este ponto não deve ser descorado e deve ser aproveitada a funcionalidade de listas de valores pré-

definidos do DSpace. Pareceu a melhor solução definir listas para os metadados onde os valores são estipulados pela biblioteca, RCAAP, etc., desta forma é possível desde o início do processo aumentar a probabilidade de o preenchimento dos metadados estarem correctos.

6. Migração das colecções

Este capítulo aborda as fases de migração da colecção do Greenstone para o DSpace e os problemas e soluções correspondentes encontrados.

6.1. Metodologia

A exportação da colecção do Greenstone para o DSpace passa por utilizar a funcionalidade de exportação existente no Greenstone e que já tem uma parametrização por defeito que permite exportar para o DSpace. No lado do DSpace a importação é realizada utilizando a importação em lote via linha de comando (New Zealand Digital Library Project, 2010).

A estrutura necessária (Figura 11), à importação de colecções no DSpace é definida da seguinte forma:

1º Ponto:

Deve existir uma pasta por cada item a ser importado e estas devem estar todos dentro de uma pasta principal.

2º Ponto:

Cada pasta identificada no ponto anterior, deve conter os seguintes ficheiros:

- Ficheiro de metadados Dublin Core, *dublin_core.xml*. Estrutura:

```
<dublin_core>
  <dcvalue element="title" qualifier="none">Titulo do documento</dcvalue>
  ...
</dublin_core>
```

- Ficheiro de metadados de outro esquema caso seja necessário, como é o caso da UAtlântica, com o nome *metadata_[nome do esquema].xml*, neste caso seria *metadata_ua.xml*. Na estrutura do ficheiro é acrescentado no elemento *dublin_core* o atributo *schema*. Estrutura:

```
<xml version="1.0" encoding="UTF-8"?>
<dublin_core schema="ua">
  <dcvalue element="titulo" qualifier="none"> Titulo do documento </dcvalue>
  ...
```

```
</dublin_core>
```

3º Ponto:

Ficheiro *contents*, que contem uma linha por cada ficheiro a ser importado.

4º Ponto:

E os respectivos ficheiros a serem importados.

No lado do DSpace deve existir pelo menos uma colecção e o esquema de metadados a ser importado deve existir.

Como referido anteriormente a exportação do Greenstone para o DSpace deve em primeiro lugar cumprir os parâmetros identificados no capítulo 4.2 e de seguida passar pelas seguintes fases:

1ª Fase:

- No interface gráfico do bibliotecário do Greenstone carregar a colecção pretendida;
- Exportar a colecção utilizando a estrutura pré-definida para o DSpace.

2ª Fase:

- Os requisitos identificados anteriormente devem ser garantidos antes de iniciar o processo de importação;
- Aceder à linha de comandos e navegar até à pasta *bin* da instalação do DSpace e executar o comando de importação em lote:

```
dspace import -a -e <utilizador> -c <ID da colecção> -s <pasta a ser importada> -m <mapfile>
```

- No comando anterior o utilizador tem de ter permissões de publicação e o parâmetro *<mapfile>* corresponde ao ficheiro de mapeamento gerado durante a importação, onde é feita a associação entre a pasta do item importado e a ligação persistente criada. Este ficheiro pode ser utilizado mais tarde para actualizar ou

remover os itens importados. Para que estas acções sejam possíveis a estrutura de pasta deve ser salvaguardada.

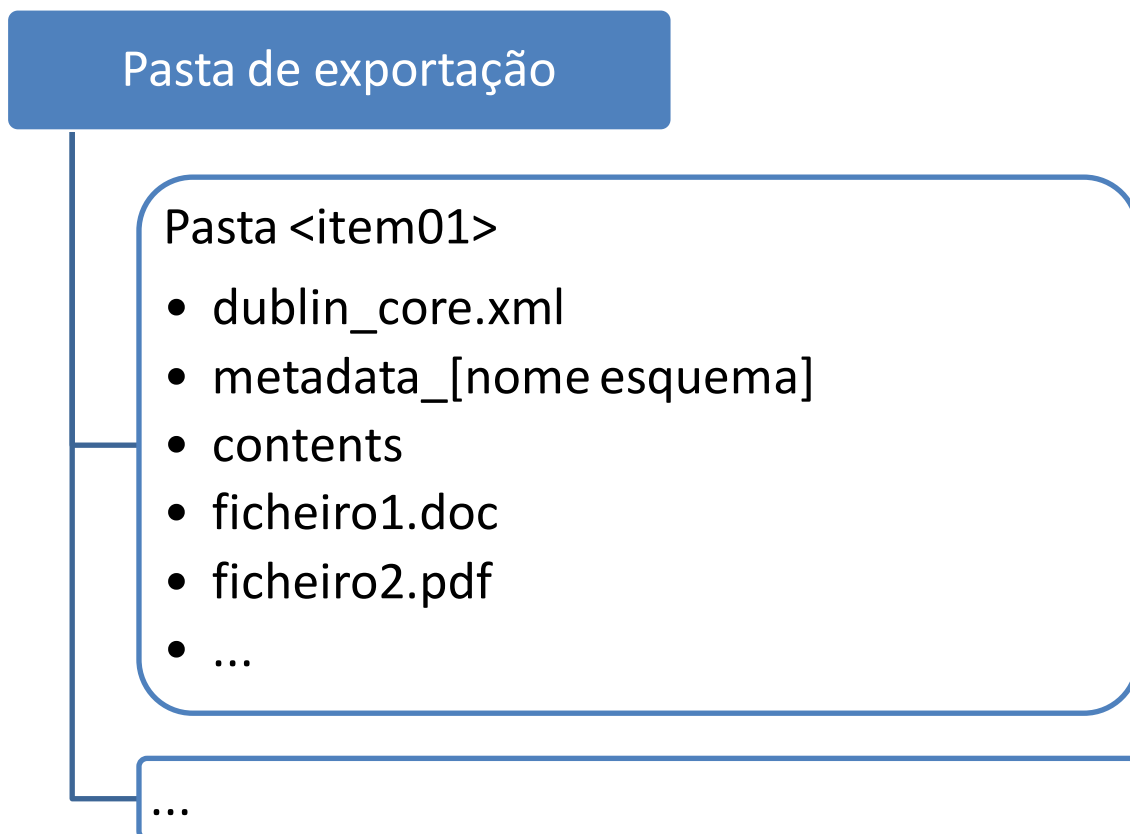


Figura 11 - Exportação e Importação: estrutura necessária para o DSpace

6.2. Problemas e ajustes

Foram realizados vários testes à exportação e importação da colecção. Através destes testes foi possível identificar e eliminar problemas apresentados até chegar a um procedimento final.

No lado do Greenstone, a exportação utilizando a parametrização já existente para o DSpace, exporta a colecção seleccionada utilizando o esquema de metadados Dublin Core ou seja, ao exportar este apenas vai buscar os metadados com o prefixo 'dc'. Caso o esquema de metadados associado à colecção não seja o Dublin Core o Greenstone só exporta alguns campos do esquema de metadados de sistema com o prefixo 'ex', e que

permitem o mapeamento para o sistema Dublin Core. No caso da UAtlântica o prefixo do esquema de dados é ‘ua’ o que levou a que não fosse realizada a exportação dos metadados pretendidos. De forma a contornar a situação, foram realizados os seguintes passos identificados após a realização de vários testes via tentativa erro:

1º Passo:

Na pasta *metadata* da colecção no Greenstone foi copiado o ficheiro *ua.mds* para *dc.mds*.

2º Passo:

O ficheiro anterior foi editado e o atributo *namespace* do elemento *metadataset* foi alterado de ‘ua’ para ‘dc’ e o elemento *name* foi alterado para ajudar a diferenciar os esquemas na lista de esquemas no interface gráfico do bibliotecário no Greenstone:

- Alteração do atributo de *namespace="ua"* para *namespace="dc"*;
- Alteração do atributo de *<Name>Metadados UATLA VI</Name>* para *<Name>Metadados UATLA Export DSpace VI</Name>*.

3º Passo:

No interface gráfico do bibliotecário do Greenstone foi carregada a colecção criada para testes e no separador *Enrich* foi removido o esquema ‘ua’, ficando apenas o ‘dc’, criado nos passos anteriores. A remoção do esquema actual irá despoletar um processo de mapeamento onde os metadados do esquema removido, ‘ua’, que contenham valores são mapeados, conforme selecção do utilizador, com os campos do esquema novo ‘dc’. Este mapeamento é guardado no ficheiro *profile.xml* na pasta *metadata*. Para evitar erros, deve-se fechar o interface gráfico do bibliotecário e editar o ficheiro *profile.xml* e garantir que os mapeamentos estão correctos.

4º Passo:

Tornar a abrir o interface gráfico do bibliotecário e carregar novamente a colecção e no separador *Enrich* validar aleatoriamente se os metadados se encontram preenchidos conforme o esperado. No separador *Create* realizar o *Complete Build*, reconstruindo assim toda a colecção.

5º Passo:

Realizar a exportação da colecção utilizando a parametrização para o DSpace.

Ao concluir os passos anteriores é criada a estrutura de pastas já descrita. A partir deste momento não é necessário realizar qualquer operação na plataforma Greenstone.

No DSpace deve ser criada uma colecção para albergar os itens exportados da colecção do Greenstone, no caso da UAtlântica só é necessário criar uma colecção pois no Greenstone só existe uma. Após a importação e criadas as colecções necessárias os itens devem ser movidos manualmente. Deve ainda ser replicado esquema de metadados da UAtlântica, 'ua'.

A definição dos metadados no DSpace processasse de forma diferente. No DSpace os metadados são definidos por um elemento e um qualificador. O elemento corresponde ao nome do campo e o qualificador permite definir o elemento como estando hierarquicamente abaixo de outro elemento. Este modo de definição permite espelhar esquemas de metadados como o Dublin Core Qualificado. O parâmetro qualificador não é de preenchimento obrigatório e para o esquema da UAtlântica não deve ser preenchido.

Casos os passos anteriores não tenham sido realizados no DSpace, o processo de importação em lote falha.

Após termos criado uma colecção no DSpace e definido o respectivo esquema de metadados, é necessário realizar algumas alterações aos ficheiros existentes na estrutura de pastas de importação criadas a partir do Greenstone. Passemos a descrever os passos necessários e respectivas explicações:

1º Passo:

Alterar o nome do ficheiro *dublin_core.xml* para *metadata_ua.xml*. Esta alteração é necessária porque o ficheiro exportado contem o esquema de metadados da UAtlântica e não o Dublin Core. Como vimos anteriormente, para o processo de importação em

lote do DSpace os esquemas de metadados que não sejam Dublin Core, devem ser guardados com outro nome e estrutura.

2º Passo:

Criar o ficheiro *dublin_core.xml*. Este passo é necessário porque apesar de não existirem itens a serem importados que utilizem o esquema Dublin Core o DSpace procura por este ficheiro no processo de importação em lote. O ficheiro não deve conter entradas de itens.

3º Passo:

Editar e alterar o ficheiro *metadata_ua.xml* para incluir o atributo *schema="ua"* no elemento *dublin_core*. Este passo é obrigatório porque faz parte da estrutura do ficheiro XML em causa, caso contrário o processo de importação em lote não reconhece o ficheiro como pertencente a um esquema de metadados diferente do Dublin Core.

Neste passo são também removidos os elementos criados por defeito pelo Greenstone e que correspondem a elementos do esquema de metadados de sistema 'ex'. Os elementos são: *language*, *date*, *creator*, *title*, *identifier* e *source*. Estes elementos têm de ser removidos pois não existem no esquema de metadados 'ua' criado no DSpace.

4º Passo:

Apagar o ficheiro *handle* gerado pelo Greenstone. Este ficheiro contém a referência de ligação persistente para o item importado. Apesar de a importação ocorrer sem erros e o acesso ao recurso funcionar correctamente, a remoção do ficheiro *handle* é realizada para forçar o DSpace a criar a referência.

Para facilitar a implementação dos passos anteriores, pois o processo de percorrer pasta a pasta seria bastante moroso, foi elaborado um script usando a linguagem VBScript que permite automatizar os passos anteriores. Definição do script:

```
' # Este script percorre a estrutura de ficheiros da pasta de exportação do Greenstone  
' # e define a estrutura de ficheiros necessária bem como o respectivo conteúdo.  
' # O script deve ser colocado na pasta de exportação.
```

```
' #  
' # Elaborado por Leonardo Almeida Cruz no ambito do projecto final de licenciatura  
  
' Definição de parâmetros necessários  
strPrefix = "ua" ' Nome do esquema de metadados criado no DSpace  
  
' Criar os objectos e constantes necessários  
Set objShell = CreateObject("Wscript.Shell")  
Set objFileSystem = CreateObject("Scripting.FileSystemObject")  
Const ForReading = 1  
Const ForWriting = 2  
  
' Obter pasta onde o script se encontra  
strPath = Wscript.ScriptFullName  
Set objFile = objFileSystem.GetFile(strPath)  
strMainFolder = objFileSystem.GetParentFolderName(objFile)  
  
' Percorrer as subpastas da pasta de exportação  
Set objFolder = objFileSystem.GetFolder(strMainFolder)  
Set objSubFolderList = objFolder.SubFolders  
  
For Each objSubFolder in objSubFolderList  
    strFolder = strMainFolder & "\" & objSubFolder.name  
  
    ' Renomear ficheiro dubli_core.xml para metadata_[prefix].xml  
    On Error Resume Next  
    objFileSystem.MoveFile strFolder & "\dublin_core.xml", strFolder & "\metadata_" &  
strPrefix & ".xml"  
    On Error Goto 0  
  
    ' Criar ficheiro dublin_core.xml  
    If Not objFileSystem.FileExists(strFolder & "\dublin_core.xml") Then  
        Set objFile = objFileSystem.CreateTextFile(strFolder & "\dublin_core.xml")  
        Set objFile = Nothing
```

```
Set objFile = objFileSystem.OpenTextFile(strFolder & "\dublin_core.xml",  
ForWriting)  
objFile.WriteLine("<?xml version=" & Chr(34) & "1.0" & Chr(34) & " encoding=" &  
& Chr(34) & "UTF-8" & Chr(34) & " standalone=" & Chr(34) & "no" & Chr(34) & "?>")  
objFile.WriteLine("<dublin_core>")  
objFile.WriteLine("</dublin_core>")  
objFile.Close  
End If  
  
' Alterar ficheiro metadata_[prefix].xml  
If objFileSystem.FileExists(strFolder & "\metadata_" & strPrefix & ".xml") Then  
Set objFile = objFileSystem.OpenTextFile(strFolder & "\metadata_" & strPrefix &  
".xml", ForReading)  
  
' Remover elementos referentes ao esquema de metadados DC exportados por defeito  
pelo Greenstone  
strText = ""  
Do Until objFile.AtEndOfStream  
strLine = objFile.ReadLine  
  
If InStr(strLine, "language") = 0 And InStr(strLine, "date") = 0 And InStr(strLine,  
"creator") = 0 And _  
InStr(strLine, "title") = 0 And InStr(strLine, "identifier") = 0 And InStr(strLine,  
"source") = 0 Then  
If Trim(strText) = "" Then  
strText = strText & strLine  
Else  
strText = strText & vbCrLf & strLine  
End If  
End If  
Loop
```

```
' Incluir o atributo "schema={prefix}" no elemento <dublin_core>
strNewText = Replace(strText, "<dublin_core>", "<dublin_core schema=" &
Chr(34) & strPrefix & Chr(34) & ">")

objFile.Close

Set objFile = objFileSystem.OpenTextFile(strFolder & "\metadata_" & strPrefix &
".xml", ForWriting)
objFile.WriteLine strNewText
objFile.Close
End If

' Apagar ficheiro handle
On Error Resume Next
objFileSystem.DeleteFile(strFolder & "\handle")
On Error Goto 0
Next

wscript.Echo "Processo concluido."
```

O processo de importação em lote ocorre no momento em que a pasta do item é avaliada pelo DSpace, caso ocorra algum erro os itens já importados não são removidos. Este modo de funcionamento leva a que deva executar, em primeiro lugar, o comando de importação em modo de teste. Este teste é realizado executando a mesma linha de comando de importação mas acrescentando o parâmetro *-t* no final. Exemplo:

```
dspace import -a -e <utilizador> -c <ID da colecção> -s <pasta a ser importada> -m
<mapfile> -t
```

Para facilitar a resolução de problemas durante o processo de importação, foi identificado que o ideal é após ocorrer um erro na execução do comando de testes, se deva mover a pasta do item onde o erro ocorreu para outra estrutura de pastas, até não

existirem mais itens a originarem erros. De seguida deve ser realizado o comando de importação e testar o respectivo acesso aos itens importados.

Após os passos anteriores, pode-se realizar a importação em lote no DSpace via linha de comandos:

```
dspace import -a -e <utilizador> -c <ID da colecção> -s <pasta a ser importada> -m <mapfile>
```

6.3. Recomendações

A importação da colecção para o DSpace utilizando o actual esquema de metadados personalizado da UAtlântica baseado em Dublin Core, implica inúmeras configurações no DSpace para que o mesmo funcione correctamente.

Os campos base de indexação do DSpace são os pertencentes ao esquema Dublin Core, sendo por isso necessário identificar quais os campos no esquema da UAtlântica que se pretende usar para indexação e adicioná-los à configuração do DSpace.

Os campos usados para indexação têm também reflexo na página de navegação do DSpace, logo é também necessário alterar os campos que são apresentados no interface de utilizador que permitem a navegação por categoria, como por exemplo visualizar por título ou por autor.

É necessário também alterar os formulários de submissão para que estes incluam os campos do esquema de metadados da UAtlântica.

Durante os testes realizados neste trabalho, estes foram os pontos identificados como tendo impacto com a utilização de um esquema de metadados personalizado, podendo existir mais situações onde haja impacto e necessite de alterações na configuração do DSpace.

Para contornar os problemas anteriores e evitar problemas não detectados, é recomendação do autor a utilização do esquema Dublin Core para todos os campos que tenham mapeamento directo. Para os campos sem mapeamento directo deve ser criado

um esquema de metadados da UAtlântica apenas com esses mesmos campos. Desta forma não é necessário alterar por completo as configurações indicadas mas sim realizar apenas pequenos acréscimos e ajustes onde se identifique necessário como é o caso dos formulários de submissão e possíveis campos dos metadados da UAtlântica que se pretendam indexar.

A utilização do esquema de metadados Dublin Core juntamente com um esquema de metadados da UAtlântica personalizado, torna mais fácil a integração das colecções com outras plataformas pois os campos necessários já estão todos no esquema Dublin Core, usado em repositórios como o RCAAP. Desta forma não necessário o mapeamento entre o esquema da UAtlântica e o Dublin Core sempre que se pretende exportar ou importar colecções ou itens.

7. Considerações

No início deste trabalho a ideia de que uma plataforma teria que ser melhor que outra foi-se desvanecendo. Foi possível perceber que ambas têm os seus pontos fortes e alguns objectivos diferentes. No caso do Greenstone, a facilidade com que é possível implementar abre a possibilidade de utilização de bibliotecas digitais a utilizadores com níveis de conhecimentos informáticos baixos. A possibilidade de poder criar colecções em suportes como CD ou DVD, abre portas a imensas possibilidades de divulgação da informação disponível no repositório. Por seu lado o DSpace oferece uma panóplia de parametrizações que vão desde fluxos de submissão com validação de metadados e aprovações intermédias, formulários de submissão, permissões por utilizador ou grupo que depois podem ser associadas verticalmente na plataforma, permitem um elevado nível de detalhe e flexibilidade.

O facto de ser possível exportar e importar de cada uma destas plataformas de uma forma fácil e compatível, desde que se utilize esquemas de metadados como o Dublin Core, permite dizer que estas plataformas podem mesmo coabitar sendo, por exemplo o DSpace a plataforma principal devido à sua flexibilidade e o Greenstone ser utilizado para criar colecções em DVD ou CD, ou permitir a alguns utilizadores criarem uma colecção localmente no seu portátil que depois é exportada e importada no DSpace.

A interoperabilidade destas plataformas é garantida através do esquema de metadados, este deve ser, dentro do possível, baseado em Dublin Core e caso existam campos específicos que não tenham mapeamento deve ser criado um esquema personalizado incluindo apenas esses campos, desta forma garante-se a interoperabilidade entre repositórios, até porque por norma os campos que não são mapeados para Dublin Core, não serão necessários nos repositórios para os quais se possa querer enviar colecções ou itens como o RCAAP.

Conclusão

É possível concluir que a migração da plataforma Greenstone para DSpace permite atingir os objectivos inicialmente propostos, ganhos ao nível da publicação de documentos e na validação dos metadados. Estes ganhos são conseguidos principalmente através dos fluxos e formulários de submissão. Foram identificados outros ganhos que não os propostos no início deste trabalho mas que revelaram que o DSpace permite um crescimento sustentável do repositório consoante as necessidades da instituição e uma flexibilidade ao nível das parametrizações que possibilita espelhar na plataforma a forma de trabalhar.

No final, a decisão de migrar um repositório digital de plataforma deve ser sempre ponderada através da identificação de todos os prós e contras, para isso devem ser envolvidos no projecto todos os *stakeholders* necessários, desta forma a decisão final a ser tomada não será apenas baseada em artigos de opinião ou estudos realizados, mas também sustentada nos conhecimentos intrínsecos que cada individuo tem da organização e dos recursos disponíveis.

Bibliografia

Diggory, M. & Donohue, T., 2010. *The DSpace System Documentation*. [Online]

Available at:

<https://wiki.duraspace.org/display/DSDOC/DSpace+System+Documentation> [Accessed 18 February 2011].

Donohue, T., 2011. *Installing DSpace*. [Online] Available at:

<https://wiki.duraspace.org/display/DSPACE/Installation> [Accessed 18 February 2011].

Dublin Core Metadata Initiative, 2003. *What is the difference between "Simple" ("unqualified") and "Qualified" Dublin Core?* [Online] Available at:

<http://dublincore.org/resources/faq/#whatisthedifference> [Accessed 18 February 2008].

DuraSpace, n.d. *DSpace Documentation: 8.3. Item Importer and Exporter*. [Online]

Available at: http://www.dspace.org/1_6_2Documentation/ch08.html#N159FE

[Accessed 18 February 2011].

Kokkelink, S. & Schwänzl, R., 2002. *Expressing Qualified Dublin Core in RDF / XML*.

[Online] Available at: <http://dublincore.org/documents/dcq-rdf-xml/> [Accessed 18 February 2011].

New Zealand Digital Library Project, 2007. *GREENSTONE DIGITAL LIBRARY USER'S GUIDE: Chapter 3 Making Greenstone Collections*. [Online] Available at:

http://www.greenstone.org/manuals/gsd12/en/html/Chapter_making_greenstone_collections.htm [Accessed 18 February 2011].

New Zealand Digital Library Project, 2007. *GREENSTONE DIGITAL LIBRARY USER'S GUIDE: Chapter 4 Administration*. [Online] Available at:

http://www.greenstone.org/manuals/gsd12/en/html/Chapter_administration.htm

[Accessed 18 February 2011].

New Zealand Digital Library Project, 2010. *Moving a collection from Greenstone to DSpace*. [Online] Available at:

http://wiki.greenstone.org/wiki/gsdoc/tutorial/en/greenstone_to_dspace.htm [Accessed 18 February 2011].

RCAAP: Módulo 4 - Administração do Repositório, n.d. *RCAAP*. [Online] Available at:

<http://projecto.rcaap.pt/formar/mod4/index.html> [Accessed 18 February 2011].

RCAAP, n.d. *RCAAP: Módulo 3 - Auto-Arquivo*. [Online] Available at:

<http://projecto.rcaap.pt/formar/mod3/contents/index.html> [Accessed 18 February 2011].

Biblioteca Digital – Migração de colecção digital para o DSpace: caso de estudo da Universidade Atlântica - Licenciatura em Gestão de Sistemas e Computação

RCAAP, n.d. *RCAAP: Módulo 5 - Workflows*. [Online] Available at: <http://projecto.rcaap.pt/formar/mod5/index.html> [Accessed 18 February 2011].

Serviços Documentação da Universidade do Minho, 2009. *Directrizes DRIVER 2.0*. [Online] Available at: http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_PT.pdf [Accessed 18 February 2011].

Witten, I.H. et al., 2005. StoneD: A Bridge between Greenstone and DSpace. *D-Lib Magazine*, Setembro.