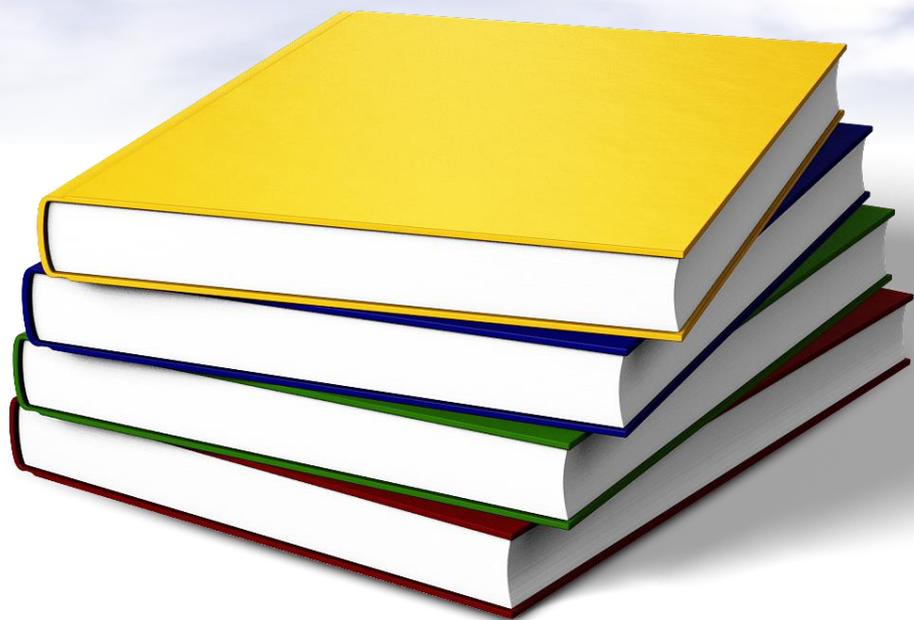


# Biblioteca Digital

## Migração de colecção digital para o DSpace: Caso de estudo da Universidade Atlântica



Licenciatura em Gestão de  
Sistemas e Computação

Projecto Final de Licenciatura

Autor: Leonardo Almeida Cruz  
(20070975)

# 0. Agenda



1. Introdução
2. Enquadramento
3. Processo de aquisição documentos na UAtlântica
4. Implementação do ambiente de testes
5. O DSpace
6. Migração das colecções
7. Considerações
8. Conclusão

# 1. Introdução



- Problema:
  - Melhoria do processo de aquisição de documentos
    - Integração de todo o processo de aquisição
    - Definição de fluxos de aprovação
    - Maximizar qualidade dos metadados
- Objectivos:
  - Identificar vantagens de migração de Greenstone para DSpace
  - Definir processo de migração

# 2. Enquadramento



- Trabalhos anteriores (PFL):
  - *Protótipo de uma biblioteca digital numa universidade*
  - *Autor: Luís Manuel L. Pinto da Rocha*
- Possíveis vantagens na migração para DSpace:
  - Processo de publicação descentralizado
  - Fluxo de aprovações
  - Flexibilidade de objectivos no utilização do repositório
  - Políticas de acesso à informação
  - RSS, estatísticas de utilização, integração AD, interface Web

# 3. Processo de aquisição documentos na UAtlântica



- Processo de aquisição
  - Definido pelo volume de informação
- Processo de aquisição de poucos documentos
  - É possível utilizar apenas o interface do Greenstone
- Processo de aquisição em lote
  - Fases de intervenção manual e automática
  - Fases com diferentes graus de dificuldade
  - Utilização de diferentes aplicações em cada fase

# 3. Processo de aquisição de documentos na UAtlântica



- Estrutura do Greenstone (exemplo)



# 3. Processo de aquisição documentos na UAtlântica



- Estrutura do Greenstone
  - Pasta por colecção
    - Todos os documentos submetidos
    - Ficheiro *'metadata.xml'* com todos os metadados
  - Subpasta *'archives'*
    - Subpasta por cada documento publicado
    - Documento em si e ficheiro de metadados
  - Subpasta *'metadata'*
    - Ficheiro por cada esquema de metadados (\*.mds)
    - Ficheiro *'profile.xml'* com mapeamento entre esquemas

# 3. Processo de aquisição de documentos na UAtlântica



- Fluxo do processo de aquisição

1ª Fase: envio de formulário e suporte digital

Cliente e-mail

Acrobat Reader



2ª Fase: recepção do formulário e arquivo do suporte digital

Cliente e-mail

Acrobat Writer

Explorador



3ª Fase: enriquecimento dos metadados

Acrobat Writer

# 3. Processo de aquisição de documentos na UAtlântica



- Fluxo do processo de aquisição (continuação)

4ª Fase: exportação dos dados do formulário para XML

Acrobat Writer



5ª Fase: definição dos metadados

Excel



6ª Fase: gerar ficheiro XML compatível com Greenstone

Aplicação 'Serialização de metadados em Greenstone'

# 3. Processo de aquisição de documentos na UAtlântica



- Fluxo do processo de aquisição (continuação)

7ª Fase: adicionar conteúdo XML ao ficheiro 'metadata.xml'

Editor de texto



8ª Fase: compilar a colecção no Greenstone

Interface gráfico do bibliotecário no Greenstone



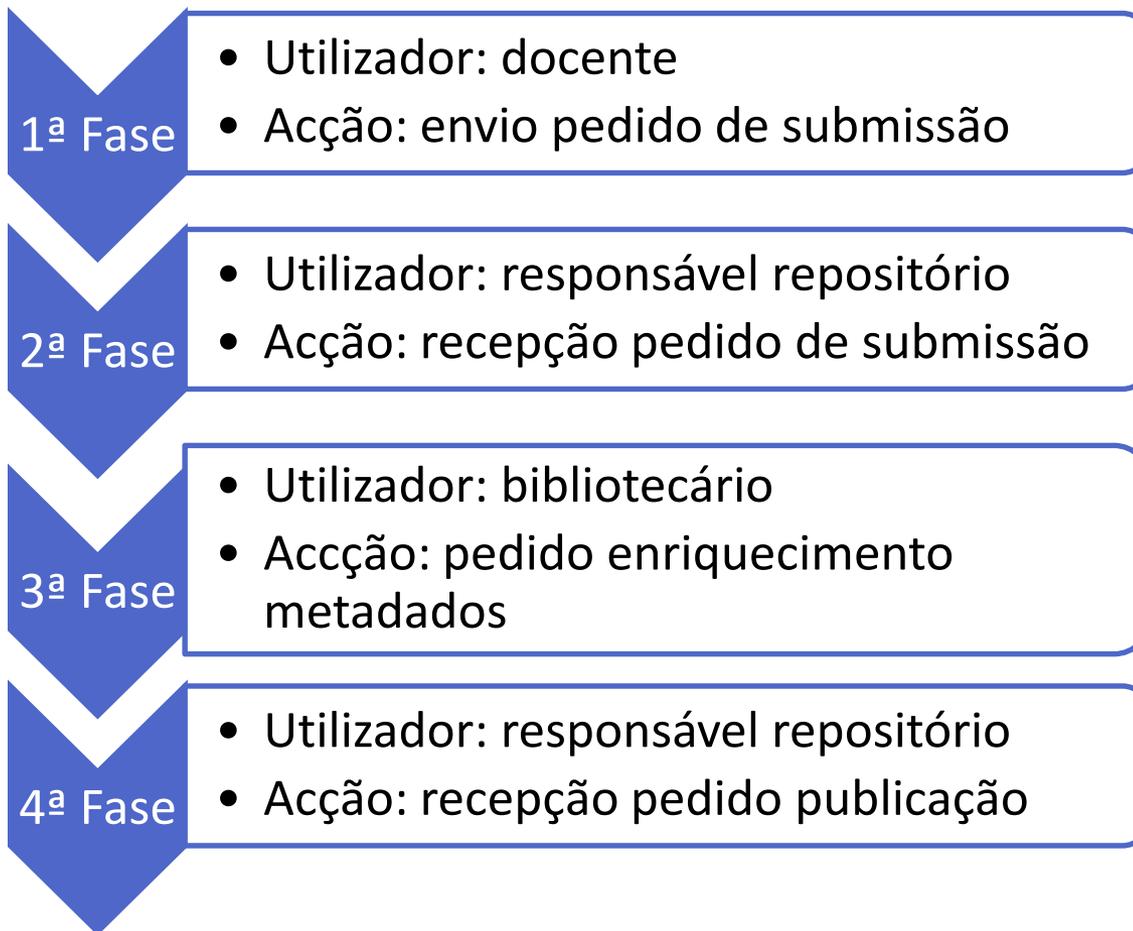
Documentos disponíveis para consulta

Interface do *Web Browser*

# 3. Processo de aquisição documentos na UAtlântica



- Comunicações no processo de aquisição



# 3. Processo de aquisição de documentos na UAtlântica



- Organização do repositório
  - Definido apenas por 1 colecção
  - Metadados baseados no esquema Dublin Core
    - Prefixo 'ua'
    - Foram criados campos extra
    - Nomes dos campos traduzidos para português
  - Hierarquia visual definida pelo metadado 'ua.area e curso'

# 3. Processo de aquisição de documentos na UAtlântica



- Organização do repositório

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



# 4. Implementação do ambiente de testes



- Utilização de máquina virtual
- Instalação e configuração do Greenstone
  - SO: Windows Server 2008
  - Utilizado motor Web da plataforma baseado em Apache
  - Utilizada última versão disponível
  - Fácil instalação
  - Replicação da colecção em Greenstone
    1. Criar colecção e substituir pasta 'import'
    2. Copiar ficheiro 'ua.mds' com esquema de metadados
    3. Validar: 'Gather' e 'Enrich'
    4. Realizar: 'Complete Rebuild'

# 4. Implementação do ambiente de testes



- Instalação e configuração do DSpace
  - SO: Windows Server 2008
  - Utilizada plataforma Web Apache
  - Utilizada última versão disponível
  - Utilizada última versão de cada requisito
    - Oracle Java JDK, Apache Maven, Apache Ant, PostgreSQL, Apache Tomcat, pgAdmin III
  - Instalação complexa

# 5. O DSpace



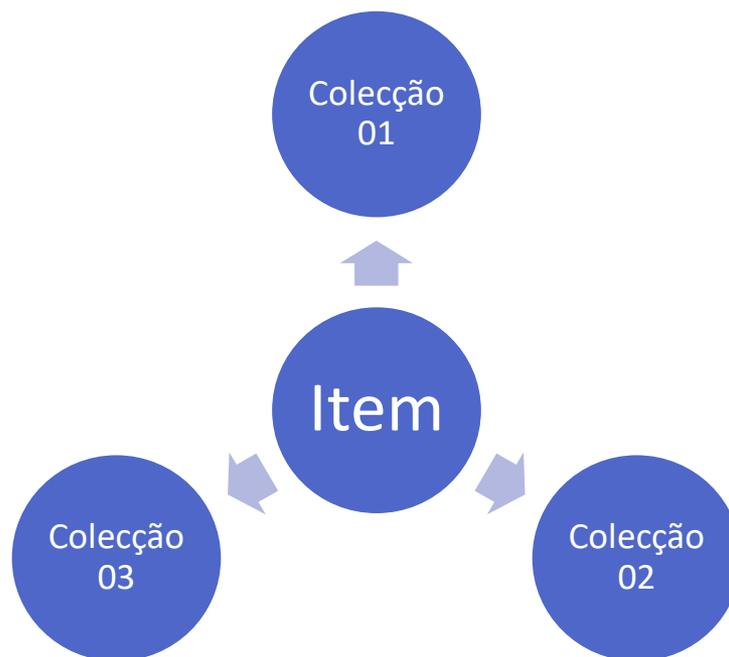
- Organização
  - Estrutura hierárquica mais rica
  - Permissões a cada nível
  - Identificador único para cada ficheiro (URL persistente)



# 5. O DSpace



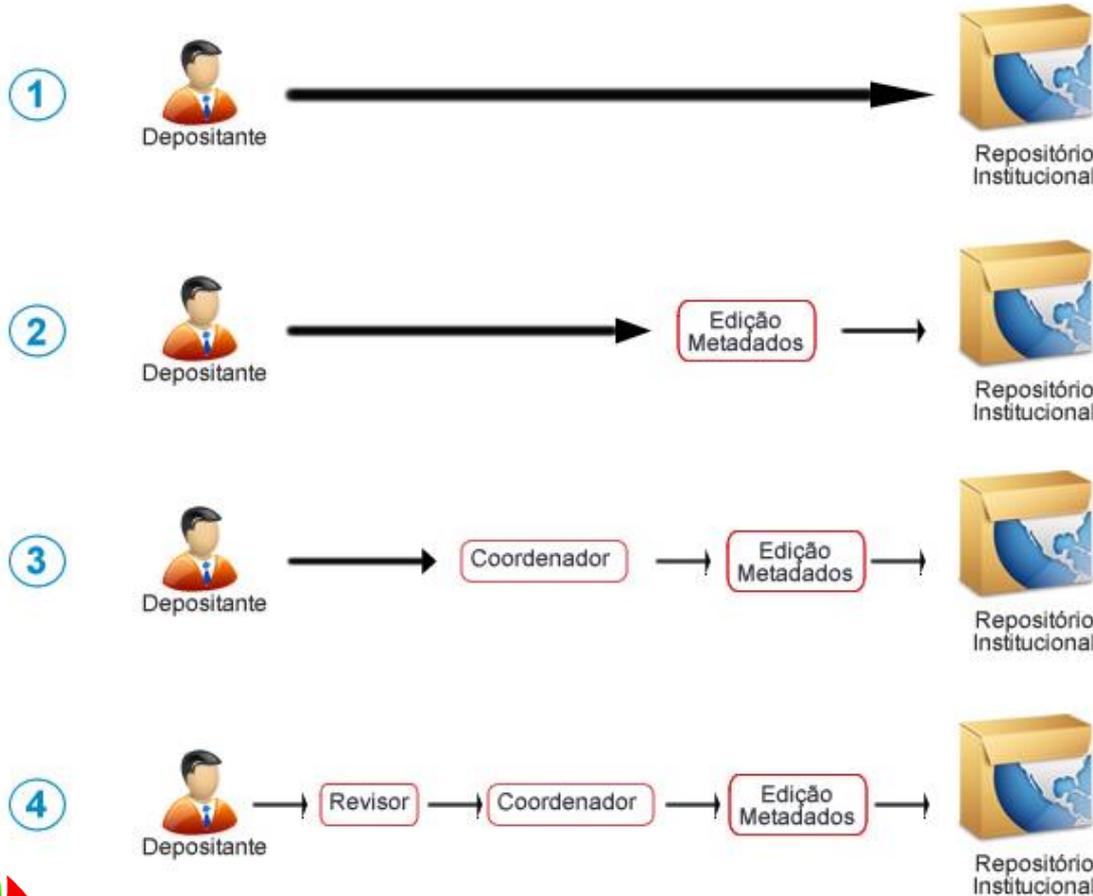
- Organização
  - Item pode estar ligado a várias colecções



# 5. O DSpace



- Fluxos de aprovação (exemplos)



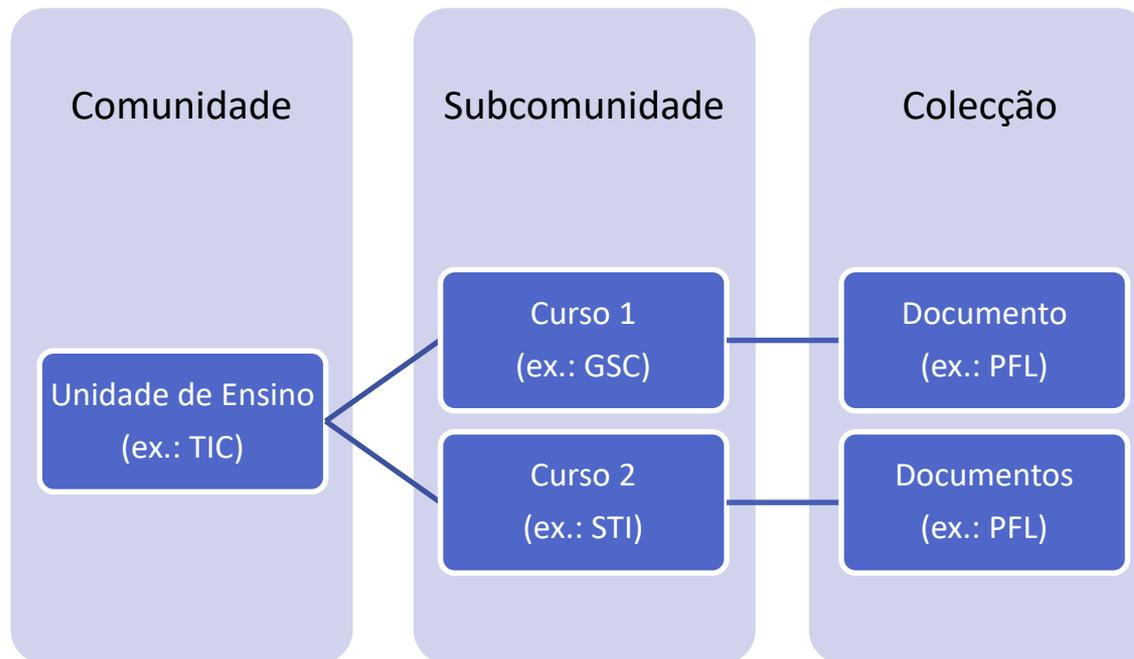
Actor DSpace	Actor UAtlântica (exemplo)
Depositante	Aluno
Revisor	Docente
Coordenador	Docente
Editor de Metadados	Bibliotecário

(Fonte: <http://projecto.rcaap.pt/formar/mod5/workflows.html>)

# 5. O DSpace



- Parametrizações propostas
  - Comunidades: uma por cada unidade de ensino
  - Subcomunidades: uma por cada curso
  - Coleções: uma por cada PFL



# 5. O DSpace



- Parametrizações propostas
  - Permite espelhar o actual repositório
  - Permite adicionar futuramente:
    - Colecções por mestrados
    - Comunidades para departamentos de investigação, biblioteca, etc.
  - Permite reutilizar fluxos de aprovação para conteúdos semelhantes
  - Permite utilizar diferentes fluxos de aprovação consoante conteúdo

# 5. O DSpace



- Fluxo de aprovação proposto



(Fonte: <http://projecto.rcaap.pt/formar/mod5/workflows.html>)

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

# 6. Migração das colecções



- Metodologia
  - Utilização da exportação para DSpace do Greenstone
  - No Greenstone deve ser redefinido o esquema de metadados para conter o prefixo 'dc' (Dublin Core)
  - No DSpace deve existir uma colecção e esquema de metadados a ser importado
  - Utilização da importação em lote do DSpace
    - Deve existir uma pasta por cada item a ser importado

# 6. Migração das colecções



- Metodologia
  - Importação em lote do DSpace: estrutura de cada pasta

Pasta de exportação

Pasta <item01>

- dublin\_core.xml
- metadata\_[nome esquema]
- contents
- ficheiro1.doc
- ficheiro2.pdf
- ...

...

# 6. Migração das colecções



- Problemas e ajustes
  - Realização de vários testes até procedimento final
  - Exportação no Greenstone utilizando opção DSpace
    - Exporta apenas metadados com prefixo 'dc'
    - Metadados da UAtlântica têm prefixo 'ua'
    - Conversão dos metadados existentes para novo esquema de metadados cujo prefixo é 'dc'
  - Importação em lote no DSpace
    - Criar esquema de metadados
    - Alterar estrutura de pastas para permitir importação
    - Definido script para realizar o ponto anterior automaticamente
    - Executar processo de importação em modo de testes

# 6. Migração das colecções



- Recomendações
  - Alteração ao esquema de metadados da UAtlântica
    - Utilizar o esquema Dublin Core para todos os campos que tenham mapeamento directo
    - Utilizar esquema à parte para campos extra necessários
  - O ponto anterior evita alterações à configuração do DSpace
    - Campos de indexação, campos de categoria, formulários
    - Apenas é necessário acrescentar campos aos formulários e campos de indexação extra
  - O ponto anterior permite ainda uma melhor integração com outros repositórios como o RCAAP

# 7. Considerações



- Ambas as plataformas podem coexistir
  - Greenstone pode ser usado para criar colecções
  - DSpace usado como plataforma principal do repositório
  - Greenstone pode ser usado para distribuir colecções via CD
- Esquema de metadados Dublin Core permite uma maior integração entre plataformas
- DSpace permite maior nível de detalhe e flexibilidade
  - Fluxos de aprovação com validação de metadados
  - Permissões de utilização e acesso aplicadas verticalmente
  - Formulários de submissão

# 7. Conclusão



- Migração para o DSpace permite atingir o objectivo proposto pelo problema:
  - Melhoria do processo de aquisição de documentos
    - Integração de todo o processo de aquisição
    - Definição de fluxos de aprovação
    - Maximizar qualidade dos metadados
- Outros ganhos identificados com a migração:
  - Crescimento sustentável
  - Espelhar no repositório a forma de trabalhar