

DSpace - Projecto de Repositório Digital de Documentos do LNEC

Rui Alexandre Rodrigues Gamito

Laboratório Nacional de Engenharia Civil, Departamento de Geotecnia, Núcleo de Tecnologias da Informação para a Engenharia Civil, Lisboa, Portugal
Email: rgamito@lnec.pt

Resumo

O trabalho que se apresenta foi desenvolvido no âmbito do PIP 2005-2008 (Plano de Investigação Programada) do LNEC, área Software Livre para sistemas de informação científicos.

Nesta comunicação é apresentado o trabalho de implantação do Repositório Digital de Documentos do LNEC, recorrendo à plataforma de software livre DSpace.

São apresentadas as diferentes fases do projecto, bem como as preocupações, dificuldades e soluções encontradas. São também exemplificadas algumas implementações de funcionalidades extra e de ferramentas adicionais necessárias para a adaptação da plataforma à realidade do Laboratório.

É ainda delineado um objectivo mais largo, onde o Repositório se insere, nomeadamente o de servir de repositório de backend único para todas as aplicações em funcionamento no Laboratório que manipulem documentos.

1 Introdução

Repositórios digitais são sistemas de informação que armazenam, preservam, divulgam e dão acesso à produção intelectual de comunidades científicas. São normalmente dotados de capacidades de pesquisa não só pela informação referente aos documentos (título, autor, etc)¹, mas também pela informação contida no próprio documento.

Repositórios institucionais (RI) são locais *online* para a armazenar e preservar - em forma digital - as produções intelectuais de uma instituição, particularmente instituições onde se faça investigação. Os dois principais objectivos de possuir um repositório institucional são:

- permitir acesso aberto à produção intelectual da pesquisa institucional através do auto-arquivo;
- armazenar e preservar outros bens digitais institucionais, incluindo literatura não publicada (ex., relatórios técnicos), que de outro modo poderia ser facilmente perdida ou dificilmente acedida.

1 Que passaremos a designar como metadados.

A origem da noção de “repositório institucional” é dupla:

Os RI estão parcialmente ligados à noção de interoperabilidade digital que, por sua vez, está ligada à *Open Archives Initiative* (OAI) e ao protocolo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH). Os RI estão também parcialmente ligados à noção de biblioteca digital – colecionando, armazenando, classificando, catalogando, preservando e fornecendo acesso a conteúdo digital, analogamente ao que é realizado por uma biblioteca convencional.

Como facilmente se percebe, um RI fará uso de um repositório digital para garantir o cumprimento do seu propósito.

O DSpace é um repositório digital, usado por muitas instituições como parte integrante do seu RI. Desenvolvido em parceria pelo *MIT Libraries* e pela *Hewlett-Packard Labs*, a plataforma DSpace serve uma grande variedade de necessidades de arquivo digital.

O DSpace aceita todas as formas de material digital incluindo texto, imagens, vídeos e ficheiros áudio Os seus possíveis conteúdos incluem os seguintes:

- Artigos
- Relatórios Técnicos
- Ensaios de trabalho (*working papers*)
- Ensaios de Conferência (*conference papers*)
- E-teses
- Datasets: estatísticos, geoespaciais, matlab, etc.
- Images: visual, científicas, etc.
- Ficheiros Áudio
- Ficheiros Vídeo
- *Learning objects*²
- Reformatted digital library collections

Usando o serviço *Global Handle Registry* (GHR), todos os documentos inseridos no DSpace podem ficar referenciados com um identificador a nível global, isto é, um identificador único que permite a referência unívoca de um documento, garantindo acesso imediato ao mesmo a partir de qualquer parte do mundo.

2 Implantação no LNEC: o início

Tendo como premissa a vantagem de possuir um repositório que centralize a maioria da documentação científica institucional e os acessos à mesma, iniciaram-se no ano de 2006 os esforços para conseguir montar uma plataforma de repositório digital do Laboratório. Pela experiência positiva das Universidades do Minho e de Évora na adopção da plataforma DSpace, iniciaram-se os trabalhos de implantação deste *software* no LNEC.

Instância piloto e levantamento de especificações

A fim de perceber quão adequado às necessidades do Laboratório o DSpace era,

² Quaisquer recursos digitais que possam ser usados para ajudar a educação;

tal como se apresentava, foi colocada em funcionamento uma instância de teste, bem como estabelecido um grupo de utilizadores piloto.

Foi inicialmente necessário identificar a melhor forma de organizar a estrutura global sob a qual assentaria a distribuição de documentos, fazendo uso, claro está, das unidades de organização disponibilizadas, nomeadamente **comunidades e colecções**.

A organização seleccionada obedece sensivelmente ao organigrama da instituição (ver figura 1), criando assim uma estrutura encimada pelos diversos departamentos. Directamente abaixo, figuram os seus núcleos constituintes e, dentro destes, foi criada uma organização por tipo de documento, facilitando a distribuição e consulta dos mesmos.

Depois de vencidas as dificuldades iniciais inerentes à aprendizagem da mecânica da plataforma - não esquecer que havia a necessidade de avaliar não só a componente de utilizador final, que apenas consulta informação, mas também submissão de documentos e mesmo certas tarefas de administração - o grupo piloto começou eficazmente a identificar pormenores que podiam ser melhorados e até novas funcionalidades.



Figura 1: Organização Departamental

Modificações

Nesta secção serão apresentadas algumas das alterações realizadas à plataforma DSpace, de modo a que esta se adaptasse mais convenientemente às necessidades do Laboratório.

Segurança

Uma das principais preocupações, desde o início, foi a segurança e confidencialidade a ser necessariamente assegurada para alguns documentos.

Muitos dos documentos produzidos no Laboratório são confidenciais, pelas mais diversas razões. Muitas vezes, a simples apresentação do título pode ser reveladora, pelo que se sentiu a necessidade de estender os mecanismos de segurança existentes. Estas alterações, tanto a nível de motor de pesquisa como a nível de motor de *browsing*, inibem a apresentação ou listagem de quaisquer documentos que tenham sido submetidos em colecções confidenciais a utilizadores que não tenham autorização explícita para aceder a essas mesmas

colecções. A listagem de colecções assinaladas como confidenciais também é sujeita autorizações, o que na prática se traduz em que, quando um utilizador não autenticado ou que não possua autorização explícita aceda à comunidade correspondente a um núcleo, não veja as suas colecções confidenciais.



Figura 2: Página de núcleo, visão sem permissões. Note-se a alteração da secção “Entradas Recentes” e que não aparecem todas as colecções do núcleo.



Figura 3: Página de núcleo, visão com permissões totais. Note-se que aparecem colecções confidenciais e itens confidenciais na listagem “Entradas recentes”.

Depósito de documentos

A principal queixa por parte dos utilizadores que tinham como responsabilidade introduzir documentação no DSpace era o facto de não saberem quais os campos de metadados a preencher, já que estes eram sempre os mesmos, apesar do tipo de documento em questão. Isto levou a que se analisasse o problema e implementasse uma orientação diferente do processo de submissão, nomeadamente orientada ao tipo de documento.

Como já foi referido, o processo de submissão de documentos era sempre igual, isto é, requisitava sempre os mesmos metadados independentemente de estes serem adequados ou não ao tipo de documento em questão. De facto, o tipo de documento mais não era do que um desses metadados.

Na implementação realizada, o âmbito foi alterado de modo a que o tipo de documento a ser inserido ditasse quais os campos pedidos pelo sistema, facilitando em larga escala o trabalho de quem deposita os documentos. Inerentemente a esta modificação, foi necessário criar ferramentas de gestão de tipos de documentos, que facilitasse o controlo de quais os campos que cada tipo requeria, bem como a criação de novos tipos.

Ferramentas de importação

Embora não directamente uma alteração à plataforma, realizaram-se também esforços no sentido de conseguir importar de forma automática documentação existente noutras plataformas, espalhadas pelo Laboratório. Exemplo disto foi a importação dos dados de uma aplicação previamente desenvolvida para albergar as referências bibliográficas das publicações do Departamento de Hidráulica e Ambiente (DHA).

O DSpace requer que, para cada documento criado no sistema, seja fornecido pelo menos um ficheiro. Esse ficheiro é tipicamente a versão digital correspondente ao documento criado. No entanto, os dados provenientes da aplicação referida eram apenas metadados, não contendo a versão digital dos documentos que referenciavam. Embora este facto vá um pouco contra o intuito principal da plataforma, optou-se por introduzir um documento “falso” temporário, que será, quando possível, substituído pelo documento real.

Edição de documentos

Com o âmbito de permitir que grupos designados de utilizadores possam realizar alterações a documentos, a fim de substituir, por exemplo, ficheiros de um documento ou mesmo adicionar novas versões do mesmo, foi desenvolvido o mecanismo de edição de documentos. Este mecanismo permite associar a uma colecção um grupo de utilizadores que terá permissão para realizar alterações nos documentos nela depositadas, sem possuir quaisquer outras permissões sobre a colecção.

Apresentação

Na sequência da importação dos dados da aplicação de referências bibliográficas

do DHA, criaram-se os mecanismos necessários a que se pudessem ver as listagens resultantes de pesquisas sob a forma de referências bibliográficas. Isto permitiu que os utilizadores que anteriormente consultavam a aplicação continuassem a obter a mesma formatação de informação, embora num sistema totalmente diferente.

Outras extensões

Para além das implementações realizadas no Laboratório foram adoptadas algumas extensões já existentes.

Uma dessas extensões, desenvolvida pela Universidade do Minho, permite comentar documentos, ficando esses comentários visíveis para a comunidade. É possível responder aos comentários quer de forma autenticada quer de forma anónima. Esta extensão visa proporcionar uma troca de opinião sobre a documentação mais directa, de tal forma que esta fique no próprio sistema.

Outra extensão utilizada permite delegar algumas responsabilidades de administração a outros utilizadores. Isto é extremamente útil para, por exemplo, permitir que directores de núcleo possam criar novas colecções ou comunidades, bem como gerir outros aspectos mais locais sem necessidade de recorrer ao administrador.

Responsáveis pela introdução de documentos

Na fase inicial, a tarefa de introdução de documentos foi distribuída pelos diversos núcleos que compunham o grupo piloto. Este modo de procedimento permitiu identificar problemas como a falta de standardização no preenchimento dos campos quando realizado por uma grande heterogeneidade de pessoas.

3 Implantação no LNEC: o presente

Estão disponíveis, neste momento, cerca de 4900 documentos no repositório do Laboratório, divididos praticamente entre todos os núcleos e sectores do mesmo. Este valor representa cerca de 12.3Gb de informação, com grande tendência a aumentar, especialmente se se optar por adoptar o DSpace como acervo de imagens (fotografias, mapas, etc), como se pretende.

O modo distribuído de introdução de documentos deu lugar a um modo mais centralizado. Actualmente, a grande maioria dos documentos são submetidos ao sistema num ponto comum à sua criação física, nomeadamente o ponto de impressão. A secção de Artes Gráficas do Laboratório ficou assim com a tarefa de, depois de imprimir um documento, submeter a sua versão digital ao DSpace, garantindo assim um maior nível de uniformização no preenchimento dos metadados. No entanto, alguns documentos produzidos no Laboratório não são impressos como meio de publicação, como é o caso de, por exemplo, artigos submetidos a conferências ou revistas. Neste caso, mantêm-se autorizações em vigor para que cada núcleo possa introduzir estes documentos.

Continuam também presentemente os esforços para introduzir e organizar o máximo possível de documentação mais antiga do Laboratório.

4 Implantação no LNEC: o futuro

De futuro será necessário fazer um controlo da qualidade dos dados presentes na plataforma, isto é, assegurar uma uniformização dos metadados já introduzidos. Exemplo disto são os autores: actualmente figuram inúmeros duplicados do mesmo autor, variando apenas por pontuação ou espaçamentos. É nossa opinião de que uma melhor qualidade deste tipo de dados assegurará uma utilização muito mais eficaz do sistema, quer a nível de confiança na sua informação quer a nível de disseminação de material.

Desde o início do projecto de implantação do repositório digital, o grande objectivo tem sido centralizar neste a totalidade – ou a grande maioria – dos documentos utilizados por outras plataformas que, de uma forma ou de outra, lidem com documentos. Exemplos destas plataformas são as inerentes à gestão de expediente, biblioteca e livraria. Esse objectivo ainda está a ser perseguido, ainda que por etapas. É também objectivo que esta mudança aconteça de forma completamente transparente para os utilizadores dos sistemas.