



UNIVERSIDADE DO PORTO

Reitoria - IRIC

Universidade Digital

Análise da integração dos objectos digitais do repositório da FEUP no repositório da UP



Elaborado por:

Bruno Miguel Oliveira Pombal

Porto
Novembro de 2008

RESUMO

Começa-se este documento com uma definição de repositório digital. Apresenta-se de uma forma abreviada o repositório da FEUP (Faculdade de Engenharia da Universidade do Porto), o repositório da UP (Universidade do Porto) e o projecto que visa a constituição de uma estrutura organizacional e tecnológica, para implementar uma camada de dados pan-europeia, que permita o uso avançado de recursos de conteúdos na área da investigação no ensino superior – DRIVER.

Em seguida, faz-se uma análise crítica e construtiva, ao processo de importação dos metadados do repositório da FEUP para o repositório da UP. Esta análise culminou na identificação de necessidades de afinação e melhoramento nos metadados (do lado do repositório da FEUP) e nos procedimentos e metadados (do lado do repositório da UP). Foi ainda criada uma tabela de mapeamento do conteúdo do campo <dc:type> dos metadados do repositório da FEUP, nos metadados do repositório da UP e no DRIVER. São também apresentados diversos *links* que permitem concretizar de forma mais otimizada, a importação de metadados do repositório da FEUP para o da UP.

Por último, são tecidas breves considerações acerca das colecções de acesso condicionado no repositório da UP e sobre o módulo de publicações do SIGARRA, referindo a forma deste, poder ajudar a aumentar a qualidade e a quantidade de informação disponibilizada no repositório da UP.

ÍNDICE

1. Introdução	1
2. Repositórios: o caso da FEUP, da UP e o projecto DRIVER	1
2.1. Repositório da FEUP	1
2.2. Repositório da UP	4
2.3. DRIVER	5
3. Importação de metadados do DigiTool da FEUP para o DSpace da UP	8
3.1. Análise aos metadados do DigiTool	9
3.2. Análise do processo de importação de documentos e metadados para o DSpace.....	12
4. Conclusões	22
5. Referências bibliográficas.....	23
6. Bibliografia.....	23

1. INTRODUÇÃO

Neste documento faz-se uma análise da integração dos objectos digitais e metadados do repositório DigiTool¹ da FEUP (Faculdade de Engenharia da Universidade do Porto) no repositório DSpace² da UP (Universidade do Porto) segundo as directivas do DRIVER³ (Digital Repository Infrastructure Vision for European Research).

2. REPOSITÓRIOS: O CASO DA FEUP, DA UP E O PROJECTO DRIVER

Os repositórios digitais, definem-se como “coleções de informação digital, que podem ser construídas de diferentes formas e com diferentes propósitos. Podem ser colaborativos e com um controlo suave dos conteúdos e da autoridade dos documentos, tal como as dirigidas para o público em geral (a Wikipedia é um exemplo). Mas podem, também, ter um alto nível de controlo e ser concebidas para promover a literacia e uma aprendizagem responsável, dirigidos a públicos específicos de utilizadores, como, por exemplo, os estudantes.” [1]

2.1. REPOSITÓRIO DA FEUP

A Biblioteca da FEUP gere um repositório⁴ que engloba os conteúdos da biblioteca digital, do repositório institucional, do museu digital e do arquivo electrónico da FEUP. Actualmente, estão disponíveis, em formato electrónico, quase 12000 documentos dos quais se salienta:

¹ <http://www.exlibrisgroup.com/category/DigiToolOverview> [consulta: 12 de Novembro de 2008]

² <http://www.dspace.org/> [consulta: 12 de Novembro de 2008]

³ <http://www.driver-repository.eu/> [consulta: 12 de Novembro de 2008]

⁴ <http://digitoool.fe.up.pt/> [consulta: 11 de Novembro de 2008]

- Todas as teses de doutoramento e dissertações de mestrado apresentadas na FEUP;
- Todas as teses de doutoramento e dissertações de mestrado apresentadas por professores e investigadores da FEUP noutras instituições nacionais ou estrangeiras que existam na Biblioteca;
- Documentos registados no SiFEUP, na área das bibliografias dos docentes e investigadores;
- Trabalhos de alunos (filmes, animações, *websites*);
- Livros, capítulos de livros, normas, patentes, artigos, comunicações;
- Fotografias dos objectos dos núcleos museológicos;
- Documentos administrativos geridos pelo Arquivo;
- Materiais pedagógicos;
- Notícias e recortes de imprensa.

No âmbito do repositório, integra-se o projecto de validação da bibliografia dos docentes e investigadores, que inclui a verificação dos metadados das publicações registadas no SiFEUP, a disponibilização dos documentos em formato electrónico e a inserção da menção ao registo nas principais bases de dados internacionais (EiCompendex, Scopus, Web of Science), caso este registo se verifique. A disponibilização do texto integral dos documentos realiza-se em conformidade com o disposto nos direitos de autor, ou seja, maioritariamente via *pre-prints*.

O repositório da FEUP está registado nos *websites* internacionais de divulgação de repositórios institucionais, de modo a aumentar a sua visibilidade e partilha, nomeadamente:

- The Directory of Open Access Repositories - <http://www.opendoar.org> ;
- ROAR - <http://roar.eprints.org> ;
- Open Archives Initiative - <http://www.openarchives.org> ;
- DRIVER - <http://www.driver-community.eu> .

Os conteúdos de âmbito institucional do Repositório da FEUP, podem ser automaticamente integrados no repositório da UP por iniciativa deste.

Na figura 1, apresenta-se simplificada, o modo de obtenção, tratamento e depósito de teses, dissertações e publicações do SiFEUP em formato digital no DigiTool.

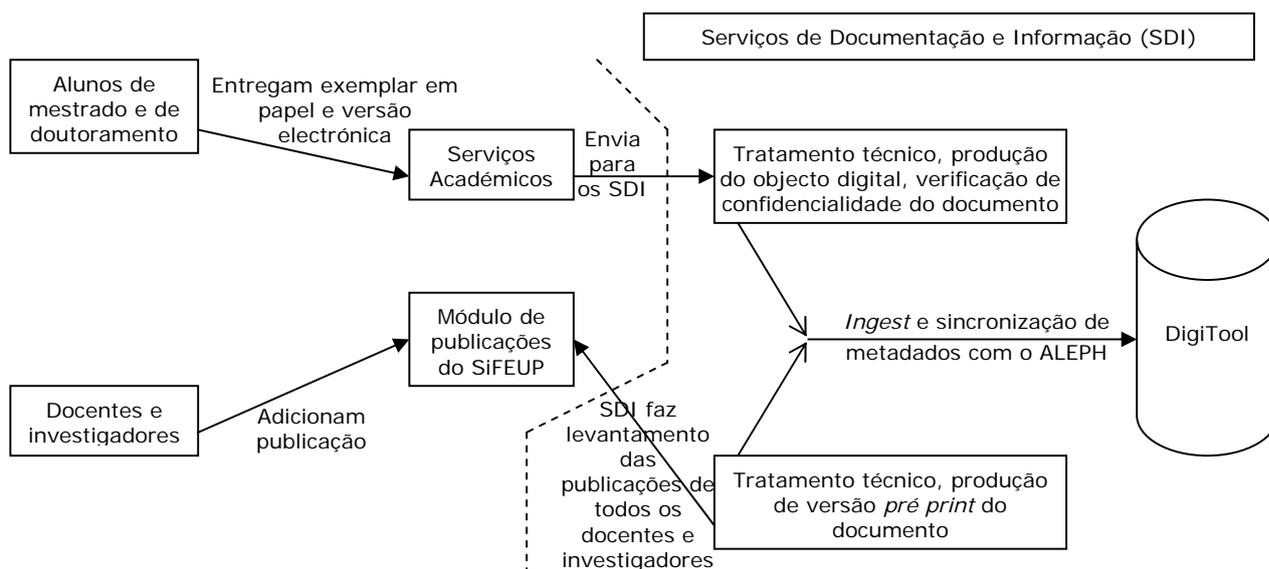


Figura 1: modo de obtenção, tratamento e depósito de teses, dissertações e publicações do SiFEUP em formato digital no DigiTool

Como se pode ver pela figura 1, os Serviços Académicos enviam as teses e dissertações apresentadas na FEUP para os SDI (Serviços de Documentação e Informação). Em seguida, os SDI, fazem o tratamento técnico e produzem a versão electrónica destes documentos e verificam a sua confidencialidade. Por último, efectuam a tarefa de *ingest* do objecto digital no DigiTool e sincronizam os seus metadados com os do ALEPH.

No caso das publicações do SiFEUP, são os SDI que tomam a iniciativa de fazer o seu levantamento. Para cada publicação, tentam obter uma versão digital da mesma, que seja passível de disponibilização em *open access*, habitualmente um *pre-print*. Seguidamente, fazem o respectivo tratamento técnico e o *ingest* do objecto digital no DigiTool e sincronizam os seus metadados com o ALEPH.

2.2. REPOSITÓRIO DA UP

O repositório da UP pretende alojar a produção científica da sua comunidade académica. Os seus conteúdos incluem desde publicações a teses e dissertações ou outros trabalhos de carácter científico. Neste repositório, é disponibilizado o acesso as publicações em texto integral das diferentes UO (Unidades Orgânicas) da UP. Os metadados das publicações acima referidas, são automaticamente transferidos para o Repositório da UP a partir da informação registada no módulo de publicações do sistema de informação SIGARRA de cada UO. Para além das publicações de carácter científico, estão previstas outras colecções, que visam apoiar a aprendizagem *online* de alunos com necessidades especiais, das quais se destaca a colecção BAES⁵ (Biblioteca Aberta do Ensino Superior) onde existem, entre outros, livros em formato áudio e braile. Apenas os estudantes do ensino superior com necessidades educativas especiais, podem aceder a colecções específicas que lhes são dirigidas, designadas pelo título genérico de colecções ALFA. Para tal, têm que se autenticar no repositório, com as mesmas credenciais que utilizam para o acesso à rede *wireless* (e-U/eduroam) das suas instituições.

Na figura 2, apresenta-se simplificada, o modo de obtenção, a partir do SIGARRA, dos metadados das publicações de alunos, professores e investigadores da UP, para o DSpace. Este processo é aplicado a todas as UO.

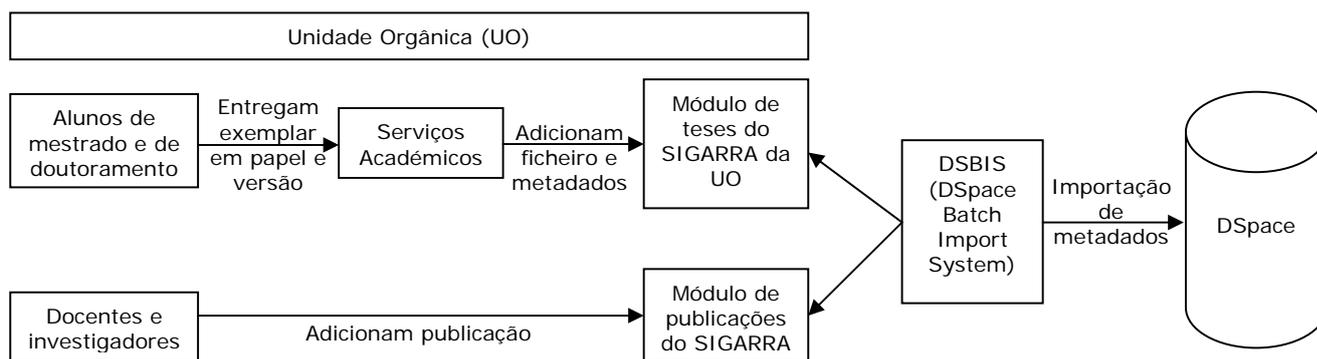


Figura 2: modo de obtenção dos documentos e metadados, a partir do SIGARRA, das publicações de alunos, professores e investigadores da UP para o DSpace

⁵ <http://repositorio.up.pt/jspui/index.html#baes> [consulta: 11 de Novembro de 2008]

Como se pode ver pela figura 2, os Serviços Académicos de cada UO registam as teses e dissertações no módulo de teses do SIGARRA. As publicações dos docentes e investigadores são inseridas pelos próprios, no módulo de publicações do SIGARRA. Para cada UO é feito um DSBI (DSpace Batch Import) através do DSBS (DSBI System) que tem como função, importar para o DSpace os metadados de todas as teses, dissertações e publicações, existentes no SIGARRA de cada UO.

2.3. DRIVER

O DRIVER é um projecto dinamizado por um consórcio financiado pela União Europeia, que visa a constituição de uma estrutura organizacional e tecnológica, para implementar uma camada de dados pan-europeia, que permita o uso avançado de recursos de conteúdos na área da investigação no ensino superior. O DRIVER desenvolve uma infra-estrutura de serviços e uma infra- estrutura de dados. Ambas estão concebidas para instrumentar os recursos e serviços existentes na rede de repositórios.

Na fase actual, este projecto centra-se nos recursos textuais que podem ser recolhidos através do protocolo OAI-PMH em *open access*. Por recursos textuais entende-se artigos científicos, teses de doutoramento, documentos de trabalho, livros electrónicos e resultados similares de actividades de investigação científica. Por *open access*, entende-se acesso sem qualquer forma de pagamento, licenciamento, controlo de acesso com *password*, controlo de acesso mediante IP, etc.

Para facilitar a integração dos diversos repositórios com o DRIVER foram criadas directrizes. As directrizes, são um instrumento para mapear ou traduzir os metadados utilizados no repositório para os metadados em Dublin Core, tal como são recolhidos pelo DRIVER. Estas directrizes focam essencialmente três questões: colecções, metadados e implementação do protocolo OAI-PMH.

No que respeita às colecções do repositório, é obrigatório utilizar “sets” (conjuntos) que definam as colecções com texto integral. Se todos os recursos do repositório forem textuais, ou seja incluírem os metadados e o texto integral e, se todos os recursos forem acessíveis em *open access*, o uso de *sets* é opcional. Os *sets* são um componente normalizado do protocolo OAI-PMH e são utilizados para apontar ou filtrar partes específicas de um repositório. Se o repositório contiver também itens não textuais, ou itens de acesso pago e/ou registos exclusivamente com metadados, pode-se utilizar o mecanismo de *sets* para filtrar esses itens quando se disponibilizam por OAI-PMH de modo a poderem ser importados pelo DRIVER.

Os recursos textuais possuem vários requisitos obrigatórios e recomendados:

- Obrigatórios:

- i. O repositório contém recursos digitais textuais;
- ii. Estão em formatos amplamente utilizados e difundidos (PDF, TXT, RTF, DOC, TeX, etc.);
- iii. Estão em *open access*, disponíveis directamente do repositório para qualquer utilizador sem restrições como autorizações ou pagamento;
- iv. São descritos por registos de metadados;
- v. Os recursos textuais e de metadados estão ligados entre si de tal modo, que um utilizador final possa aceder ao recurso textual através do identificador (normalmente um URL) no registo de metadados;
- vi. O URL de um recurso inscrito no registo de metadados está permanentemente acessível e nunca se altera ou se atribui a outro recurso;
- vii. Um identificador único identifica o registo de metadados e o recurso textual (não há apontadores para sistemas externos, como um sistema nacional de bibliotecas ou uma editora).

- Recomendados:

- viii. Verificação transparente da integridade de um recurso textual;

- ix. Medidas de controlo de qualidade (do conteúdo científico) dos recursos textuais expostos para limitá-los a, por exemplo, os recursos textuais incluídos no relatório científico anual;
- x. O URL de um recurso inscrito no registo de metadados baseia-se num esquema de identificadores persistentes como: DOIs, URNs, ARKs.

No que respeita aos metadados, foram definidas alguns requisitos obrigatórios e recomendados para solucionar as dificuldades semânticas que surgem de diferentes interpretações do Dublin Core, nomeadamente:

- Obrigatórios:

- xi. Os metadados são estruturados como DC (Dublin Core) não qualificado segundo a norma ISO 15836:2003;
- xii. Os elementos individuais de DC devem ser usados de acordo com o Documento *Exposing textual resources with OAI-PMH*⁶ - anexo 1.

- Recomendados:

- xiii. Os metadados são estruturados de acordo com esquemas mais completos como Dublin Core qualificado ou MODS;
- xiv. A escolha do idioma dos metadados fica ao critério do fornecedor de conteúdos. O idioma recomendado é o inglês;
- xv. O idioma recomendado para o resumo (a inclusão de resumo é opcional) do artigo é o inglês.

No que respeita ao protocolo OAI-PMH, foram também definidos alguns requisitos obrigatórios e outros recomendados para solucionar os problemas que surgem nas diferentes implementações no repositório local, nomeadamente:

- Obrigatórios:

⁶ http://www.driver-support.eu/documents/DRIVER_guidelines_annexes_1%200.pdf [consulta: 11 de Novembro de 2008]

- xvi. O repositório deve estar em conformidade com o OAI e de acordo com as especificações do documento *Exposing textual resources with OAI-PMH* - anexo 2;
 - xvii. Deve existir um identificador de repositório e deve utilizar-se o esquema de identificador OAI de acordo com o documento *Exposing textual resources with OAI-PMH* - anexo 2;
 - xviii. Se (e apenas se) o repositório contém outros recursos para além dos que são obrigatórios nos recursos textuais, deve ser definido um *set* OAI que identifique a colecção de recursos textuais digitais com acesso directo de acordo com o documento *Exposing textual resources with OAI-PMH* - anexo 2.
- Recomendado:
- xix. Disposições para a alteração do URL Base de acordo com o documento *Exposing textual resources with OAI-PMH* - anexo 2;
 - xx. Resposta ao *Identify* completa, incluindo o uso opcional da declaração *Description* de acordo com o documento *Exposing textual resources with OAI-PMH* - anexo 2;
 - xxi. Uso do DIDL XML-container de acordo com o documento *Exposing textual resources with OAI-PMH* - anexo 3.

3. IMPORTAÇÃO DE METADADOS DO DIGITOOL DA FEUP PARA O DSPACE DA UP

Como já foi visto anteriormente, o repositório da UP importa documentos e metadados do SIGARRA das diversas UO, de modo a integrar os seus conteúdos no repositório da UP. No contexto deste relatório, importa estudar o caso particular da FEUP.

Tendo a FEUP um repositório próprio, de recursos textuais e de metadados, teve a necessidade de adequar os metadados e as permissões dos conteúdos que disponibiliza às normas e leis existentes. Os seus metadados estão a ser afinados de modo a poderem ser disponibilizados no DRIVER. Dessa forma, na sua maioria, os metadados disponíveis no DigiTool da FEUP, estão compatíveis com as necessidades do

repositório da UP. Contudo, para que todo o processo de importação de metadados do DigiTool para o repositório da UP decorra de forma automática, há um conjunto de regras que têm que ser verificadas e afinadas quer do lado da FEUP, quer do lado da UP.

Em seguida, identifica-se e, sempre que possível, propõe-se soluções para alguns problemas que deveriam ser resolvidos, de modo a haver uma importação automática para o repositório da UP dos metadados disponibilizados pelo OAI do DigiTool e de forma a haver compatibilidade com o DRIVER.

3.1. ANÁLISE AOS METADADOS DO DIGITool

O repositório da FEUP, publica num só *set* em OAI-PMH todos os seus objectos digitais passíveis de serem disponibilizados em *open access*.

No caso das publicações, o repositório da FEUP possui cerca de 2000 registos correspondentes às publicações dos seus docentes e investigadores contudo, de momento, apenas 328 estão a ser publicados em OAI-PMH. Isto prende-se com questões legais, visto que apenas estes 328 documentos estão já disponíveis em formato *pre-print*. Espera-se que, a curto prazo, todas as publicações estejam disponíveis em OAI-PMH.

Em relação às teses e dissertações, existe um problema que tem que ser resolvido de forma a ser respeitado o requisito iii. As teses e/ou dissertações confidenciais, têm restrições de acesso. No entanto, também estão a ser publicadas em OAI-PMH pois, não há forma de as distinguir das restantes. Uma possível solução para esse problema seria utilizar o campo "rights" do Dublin Core para distinguir os documentos confidenciais dos não confidenciais. Desta forma, apenas seriam disponibilizados por OAI-PMH, as teses e dissertações sem restrições de acesso. Diga-se que, em relação às teses e/ou dissertações, não há forma de distinguir as que são apresentadas na FEUP, por pessoas que são docentes e/ou investigadores (enquanto alunos) das que são apresentadas por pessoas que são apenas alunos. Esta distinção não está feita nos metadados pois, na altura da apresentação de uma tese e/ou dissertação todas as pessoas são alunos. Contudo, há forma de

distinguir as teses e/ou dissertações que não foram apresentadas na FEUP, por docentes/investigadores da FEUP. Nestas situações, nos metadados não existe o campo "`<dc:contributor>Universidade do Porto. Faculdade de Engenharia</dc:contributor>`".

Os metadados existentes do DigiTool são obtidos directa e automaticamente, através de sincronização com o ALEPH. O campo `<dc:creator>` é preenchido com o conteúdo do campo 700 do ALEPH. Contudo, quando o número de autores de um documento é superior a 3, no ALEPH o campo 700 fica por preencher e todos os autores são mapeados nos metadados do DigiTool como "contributors". Como o campo `<dc:creator>` é obrigatório, é necessário encontrar uma solução para este problema. Uma possível solução pode passar por utilizar uma folha de estilo. Desta forma, seria possível mapear correctamente e de forma automática, os dados recolhidos do ALEPH no DigiTool.

Segundo as orientações do DRIVER, a primeira ocorrência do campo `<dc:type>` deverá conter o tipo de documento. No entanto, como se pode ver na figura 3, nos metadados do DigiTool, por vezes esta condição não é verificada. Este problema também pode ser resolvido com o recurso a uma folha de estilo, de modo a mapear correctamente os dados que estão no ALEPH nos metadados do DigiTool.

Existe uma outra questão, relacionada com o campo `<dc:format>`. Embora este campo seja apenas recomendado, também poderá ser relevante. Segundo as especificações do DRIVER, o campo `<dc:format>` deverá conter o MIME *type* do documento, por exemplo, "`<dc:format>application/pdf</dc:format>`" ou "`<dc:format>video/quicktime</dc:format>`" e, como também pode ser visto na figura 3, no DigiTool esse campo costuma ser preenchido com outro tipo de informação.

```

<?xml version="1.0" encoding="UTF-8" ?>
- <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2008-11-12T16:12:07Z</responseDate>
  <request verb="GetRecord" identifier="oai:digitool.fe.up.pt:34901"
    metadataPrefix="oai_dc">http://digitool.fe.up.pt/OAI-PUB</request>
- <GetRecord>
- <record>
- <header>
  <identifier>oai:digitool.fe.up.pt:34901</identifier>
  <timestamp>2008-10-15T01:17:45Z</timestamp>
  <setSpec>feup</setSpec>
</header>
- <metadata>
- <record xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>Optimização de configurações das redes eléctricas de parques
    eólicos</dc:title>
  <dc:creator>Resende, Fernanda de Oliveira</dc:creator>
  <dc:subject>Engenharia electrónica</dc:subject>
  <dc:subject>Sistemas de energia</dc:subject>
  <dc:subject>Parques eólicos</dc:subject>
  <dc:subject>Universidade do Porto</dc:subject>
  <dc:subject>621.3(043)</dc:subject>
  <dc:description>Tese de mestr.. Engenharia Electrotécnica e de Computadores.
    Faculdade de Engenharia. Universidade do Porto. 1999</dc:description>
  <dc:publisher>[s. n.]</dc:publisher>
  <dc:contributor>Lopes, João Abel Peças</dc:contributor>
  <dc:contributor>Universidade do Porto. Faculdade de Engenharia</dc:contributor>
  <dc:date>1999</dc:date>
  <dc:type>text</dc:type>
  <dc:type>masterthesis</dc:type>
  <dc:format>XX, 176 f., 30 cm</dc:format>
  <dc:language>por</dc:language>
  <dc:relation>FEUP > SDI > RiFEUP</dc:relation>
  <dc:identifier>http://eos.fe.up.pt:1801/webclient/DeliveryManager?
    pid=34901&custom_att_2=simple_viewer</dc:identifier>
  </record>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

Figura 3: metadados descritivos do registo com o PID 34901 no DigiTool (consultado em 2008-11-12)

3.2. ANÁLISE DO PROCESSO DE IMPORTAÇÃO DE DOCUMENTOS E METADADOS PARA O DSPACE

Como já foi referido anteriormente, o DSpace importa os documentos e os metadados, que estão em acesso público (ver figura 4), a partir do SIGARRA de cada UO, através do mecanismo de DSBI, armazenando-os na sua base de dados e dando-lhes um identificador único.

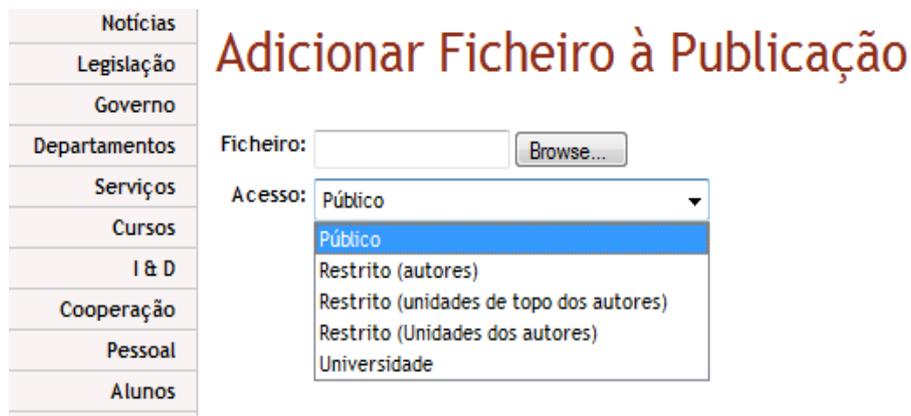


Figura 4: definição do tipo de acesso de uma publicação no SIGARRA

Actualmente, a política que está implementada no repositório da UP é de que, a responsabilidade pelos documentos disponibilizados e pelos respectivos metadados, é de quem os insere no SIGARRA.

É importante notar que, o *workflow* de importação que está montado no repositório da UP acarreta a integração de uma grande quantidade de documentos e de metadados, que provêm de origens muito diferentes. É também de realçar que, os documentos e metadados são inseridos no SIGARRA por mais de um milhar de utilizadores diferentes. Assim, de modo a assegurar a qualidade da informação disponibilizada no repositório, torna-se vital existir uma validação dos documentos e dos metadados, antes da sua disponibilização ao público no repositório da UP.

Na análise que efectuei ao repositório da UP, pude constatar que, não é feita qualquer tipo de validação dos documentos e metadados que são importados a

partir do SIGARRA. Isto origina uma série de problemas de inconsistência de informação, de duplicação de informação e até mesmo questões legais, nomeadamente:

- Existe a possibilidade do ficheiro não corresponder aos metadados. Como exemplo apresenta-se o registo disponível no repositório da UP com o *link* <http://repositorio.up.pt/jspui/handle/10216/404> , que aponta para um registo que se encontra na colecção "Livro" da comunidade da FEUP. Pode constatar-se que, os metadados dizem respeito ao livro, mas o recurso textual que está disponível, diz respeito a uma errata do livro. Existe ainda a possibilidade do ficheiro associado a uma publicação (em acesso público) ter sido trocado por lapso. Neste caso, o recurso textual e os metadados não irão coincidir. Este problema origina inconsistências no repositório da UP pois, uma publicação inserida no módulo de publicações do SIGARRA nestas condições é automaticamente importada para o repositório da UP.
- No caso de uma publicação do SIGARRA, que já tenha sido importada para o repositório da UP e que, entretanto, tenha sido apagada do SIGARRA porque, por exemplo, o seu "tipo" estava errado (campo não re-editável pelo utilizador), e que tenha sido de novo publicada, vai haver duplicação de informação no repositório da UP. Isto acontece porque a informação das publicações apagadas não chega ao repositório da UP;
- Existe a possibilidade de ser o pessoal administrativo a ficar incumbido pelos docentes/investigadores de adicionar as publicações no SIGARRA. O pessoal administrativo, habitualmente, não possui informação sobre quais são as publicações que podem ser disponibilizadas em *open access* ou se devem colocar versões *pre-print* ou *pos-print* em função da revista. Isto pode originar problemas legais pois, uma publicação inserida no módulo de publicações do SIGARRA, com um ficheiro associado em acesso público é automaticamente importada para o repositório da UP.

O facto do DSBI só importar as publicações do SIGARRA que têm ficheiro associado em acesso público, revela uma atitude passiva do repositório da UP. Esta atitude

faz com que a maioria dos metadados existentes no SIGARRA não sejam importados para o repositório da UP, condicionando assim o seu crescimento.

De modo a solucionar estes e outros problemas, proponho que a ligação que é feita ao SiFEUP pelo DSBIS passe a ser feita ao DigiTool da FEUP, como ilustrado na figura 5. No DigiTool há a garantia que não há registos duplicados e que todas as publicações existentes no SiFEUP (com ou sem ficheiro) existem em *open access* em conformidade com as políticas de todas as revistas⁷.

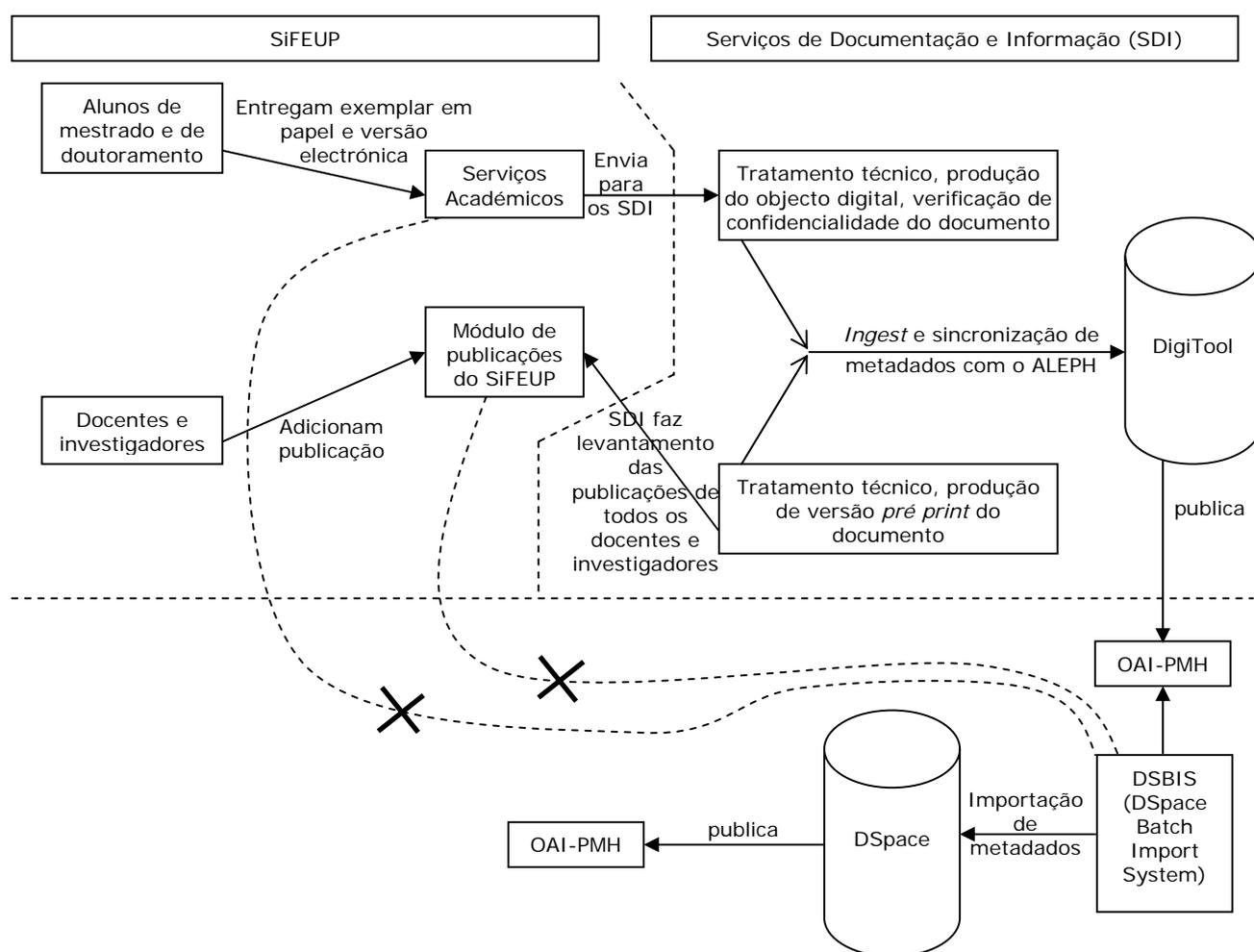


Figura 5: modo de obtenção dos documentos e dos metadados, a partir do DigiTool da FEUP, das publicações de alunos, professores e investigadores da UP para o DSpace

⁷ As políticas das revistas podem ser consultadas em <http://www.sherpa.ac.uk/romeo.php>

Para a importação de documentos e de metadados do DigiTool da FEUP com o recurso ao DSBIS, torna-se essencial possuir um mapeamento do campo <dc:type> do DigiTool para o campo com o mesmo nome no DSpace. Na tabela 1 apresenta-se esse mapeamento e, acrescentou-se ainda uma coluna com o mapeamento segundo as directivas do DRIVER II.

Tabela 1: Mapeamento dos campos <dc:type> do DigiTool da FEUP nos códigos do SIGARRA e do DRIVER

Designação	Código no campo <dc:type> nos metadados do DigiTool da FEUP	Código do módulo de publicações do SIGARRA	Código do <dc:type> segundo as directivas do DRIVER II
Dissertação de mestrado	masterthesis	Tese.M	masterThesis
Tese de doutoramento	phdthesis	Tese.D	doctoralThesis
Actas de conferência nacional	nationalconfproc	Acta	conferenceItem
Artigo em actas de conferência nacional	nationalconfpaper	Art.ActaNac	conferenceItem
Artigo em actas de conferência internacional	internationalconfpaper	Art.ActaInt	conferenceItem
Apresentação em conferência nacional	nationalconfpresentation	Não existe	conferenceItem

Apresentação em conferência internacional	internationalconfpresentation	Não existe	conferenceItem
Poster em conferência internacional	internationalconfposter	Não existe	conferenceItem
Artigo em revista internacional	internationaljnlpaper	Art.Rev.Int	article
Artigo em revista nacional	nationaljnlpaper	Art.Rev.Nac	article
Capítulo de livro	bookchapter	Cap.Liv	bookPart
Livro	book	Livro	book
Relatório	report	Relat	report
Publicação didáctica	didacticalpub	relat.PD	lecture
Relatório técnico	technicalrep	relat.RT	report
Imagens do museu	image	Não existe	other

Também poderá ser útil aos responsáveis pelo repositório da UP, conhecer uma série de instruções, baseadas em URLs, para recolher metadados e recursos textuais a partir do DigiTool, nomeadamente:

- Verificar as bases e os respectivos sets:
 - URL: <http://digitool.fe.up.pt/OAI-PUB?verb=ListSets>
 - OUTPUT:

```

<?xml version="1.0" encoding="UTF-8" ?>
- <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2008-11-13T15:57:50Z</responseDate>
  <request verb="ListSets">http://digitool.fe.up.pt/OAI-PUB</request>
- <ListSets>
  - <set>
    <setSpec>dtlaleph</setSpec>
    <setName />
  </set>
  - <set>
    <setSpec>feup</setSpec>
    <setName>FEUP</setName>
  </set>
</ListSets>
</OAI-PMH>

```

- Obter informações sobre o repositório:

- URL: <http://digitool.fe.up.pt/OAI-PUB?verb=Identify>
- OUTPUT:

```

<?xml version="1.0" encoding="UTF-8" ?>
- <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2008-11-13T16:01:49Z</responseDate>
  <request>http://digitool.fe.up.pt/OAI-PUB</request>
- <Identify>
  <repositoryName>FEUP Institutional Repository</repositoryName>
  <baseURL>http://digitool.fe.up.pt/OAI-PUB</baseURL>
  <protocolVersion>2.0</protocolVersion>
  <adminEmail>lmcosta@fe.up.pt</adminEmail>
  <earliestDatestamp>2008-07-22T14:39:38Z</earliestDatestamp>
  <deletedRecord>transient</deletedRecord>
  <granularity>YYYY-MM-DDThh:mm:ssZ</granularity>
- <description>
  - <oai-identifier xmlns="http://www.openarchives.org/OAI/2.0/oai-identifier"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai-identifier
      http://www.openarchives.org/OAI/2.0/oai-identifier.xsd">
    <scheme>oai</scheme>
    <repositoryIdentifier>digitool.fe.up.pt</repositoryIdentifier>
    <delimiter>:</delimiter>
    <sampleIdentifier>oai:digitool.fe.up.pt:29672</sampleIdentifier>
  </oai-identifier>
  </description>
</Identify>
</OAI-PMH>

```

- Recuperar um registo. Por exemplo, o registo com o PID 33165 pode ser recuperado da seguinte forma:
 - URL: http://digitool.fe.up.pt/OAI-PUB?verb=GetRecord&identifier=oai:digitool.fe.up.pt:33165&metadataPrefix=oai_dc
 - OUTPUT:

```

<?xml version="1.0" encoding="UTF-8" ?>
- <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
  http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2008-11-13T16:04:46Z</responseDate>
  <request verb="GetRecord" identifier="oai:digitool.fe.up.pt:33165"
    metadataPrefix="oai_dc">http://digitool.fe.up.pt/OAI-PUB</request>
  - <GetRecord>
  - <record>
  - <header>
    <identifier>oai:digitool.fe.up.pt:33165</identifier>
    <datestamp>2008-11-07T02:17:02Z</datestamp>
    <setSpec>feup</setSpec>
  </header>
  - <metadata>
  - <record xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns:dcterms="http://purl.org/dc/terms/"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
    <dc:title>Fracture analysis of forks of heavy duty lift truck</dc:title>
    <dc:subject>Soldadura</dc:subject>
    <dc:subject>Metalografia</dc:subject>
    <dc:subject>Análise de falhas</dc:subject>
    <dc:subject>Fadiga dos materiais</dc:subject>
    <dc:subject>620.18</dc:subject>
    <dc:subject>621.791</dc:subject>
    <dc:contributor>Figueiredo, M. V.</dc:contributor>
    <dc:contributor>Oliveira, F. M. F.</dc:contributor>
    <dc:contributor>Gonçalves, João Paulo Moreira</dc:contributor>
    <dc:contributor>Castro, Paulo Manuel Salgado Tavares de</dc:contributor>
    <dc:contributor>Fernandes, António Augusto</dc:contributor>
    <dc:type>internationaljnlpaper</dc:type>
    <dc:identifier>(SIFEUP)4911</dc:identifier>
    <dc:identifier>(SCI)169967500002</dc:identifier>
    <dc:identifier>(Compendex)1386648893</dc:identifier>
    <dc:identifier>(Scopus)2-s2.0-0035480336</dc:identifier>
    <dc:language>eng</dc:language>
    <dc:relation>Engineering Failure Analysis. - Oxford, 2001. - p. 411-421, Vol. 8, n.º 5. -
    ISSN 1350-6307</dc:relation>
    <dc:relation>FEUP > SDI > RiFEUP</dc:relation>
    <dc:identifier>http://eos.fe.up.pt:1801/webclient/DeliveryManager?
    pid=33165&custom_att_2=simple_viewer</dc:identifier>
  </record>
  </metadata>
  </record>
  </GetRecord>
</OAI-PMH>

```

- Recuperar tudo o que está num *set* e obter os identificadores de registo:
 - URL: http://digitool.fe.up.pt/OAI-PUB?verb=ListIdentifiers&metadataPrefix=oai_dc
 - OUTPUT:

```
<?xml version="1.0" encoding="UTF-8" ?>
- <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2008-11-13T16:08:04Z</responseDate>
  <request verb="ListIdentifiers" metadataPrefix="oai_dc">http://digitool.fe.up.pt/OAI-PUB</request>
  - <ListIdentifiers>
    - <header>
      <identifier>oai:digitool.fe.up.pt:13933</identifier>
      <timestamp>2008-07-22T14:39:38Z</timestamp>
      <setSpec>feup</setSpec>
    </header>
    - <header>
      <identifier>oai:digitool.fe.up.pt:13935</identifier>
      <timestamp>2008-07-22T14:39:38Z</timestamp>
      <setSpec>feup</setSpec>
    </header>
    - <header>
      <identifier>oai:digitool.fe.up.pt:13939</identifier>
      <timestamp>2008-07-22T14:39:38Z</timestamp>
      <setSpec>feup</setSpec>
```

- Recuperar os metadados que estão num *set*:
 - URL: http://digitool.fe.up.pt/OAI-PUB?verb=ListRecords&metadataPrefix=oai_dc
 - OUTPUT:

```
<?xml version="1.0" encoding="UTF-8" ?>
- <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2008-11-13T16:07:48Z</responseDate>
  <request verb="ListRecords" metadataPrefix="oai_dc">http://digitool.fe.up.pt/OAI-PUB</request>
  - <ListRecords>
    - <record>
      - <header>
        <identifier>oai:digitool.fe.up.pt:13933</identifier>
        <timestamp>2008-07-22T14:39:38Z</timestamp>
        <setSpec>feup</setSpec>
      </header>
      - <metadata>
        - <record xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
          xmlns:dcterms="http://purl.org/dc/terms/"
          xmlns:dc="http://purl.org/dc/elements/1.1/">
          <dc:title>Challenge Desafio Unico</dc:title>
          <dc:creator>SIC Noticias</dc:creator>
          <dc:description>O Challenge Desafio Unico correu à chuva na pista da Boavista
            aumentando o espectáculo dado pelos velinhos Fiat Uno.</dc:description>
          <dc:date>07/21/2007</dc:date>
          <dc:type>newsclipping</dc:type>
          <dc:format>WMV</dc:format>
          <dc:relation>FEUP > SICC > DCI</dc:relation>
          <dc:identifier>http://eos.fe.up.pt:1801/webclient/DeliveryManager?
            pid=13933&custom_att_2=simple_viewer</dc:identifier>
        </record>
      </metadata>
    </record>
    - <record>
      - <header>
```

- Recuperar os registos desejados por "timestamp":

- URL:

com data e hora:

http://digitool.fe.up.pt/OAI-PUB?verb=ListRecords&from=2008-11-10T00:00:01Z&until=2008-11-13T23:59:59Z&set=feup&metadataPrefix=oai_dc

ou apenas com "data":

http://digitool.fe.up.pt/OAI-PUB?verb=ListRecords&from=2008-11-10&until=2008-11-13&set=feup&metadataPrefix=oai_dc

Ou ainda só com o parâmetro "from":

http://digitool.fe.up.pt/OAI-PUB?verb=ListRecords&from=2008-11-10&set=feup&metadataPrefix=oai_dc

- OUTPUT:

```
<?xml version="1.0" encoding="UTF-8" ?>
- <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2008-11-13T16:20:01Z</responseDate>
  <request verb="ListRecords" from="2008-11-10" until="2008-11-13" metadataPrefix="oai_dc"
    set="feup">http://digitool.fe.up.pt/OAI-PUB</request>
- <ListRecords>
- <record>
- <header>
  <identifier>oai:digitool.fe.up.pt:33179</identifier>
  <timestamp>2008-11-12T02:16:56Z</timestamp>
  <setSpec>feup</setSpec>
</header>
- <metadata>
- <record xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:title>Increasing the efficiency of composite single-shear lap joints using
    bonded inserts</dc:title>
  <dc:subject>Engenharia mecânica</dc:subject>
  <dc:contributor>Camanho, Pedro Manuel Ponces Rodrigues de
    Castro</dc:contributor>
  <dc:contributor>Tavares, C. M. L.</dc:contributor>
  <dc:contributor>Oliveira, R. de</dc:contributor>
  <dc:contributor>Marques, António Torres</dc:contributor>
  <dc:contributor>Ferreira, António Joaquim Mendes</dc:contributor>
  <dc:type>internationaljnlpaper</dc:type>
  <dc:identifier>(SIFEUP)14220</dc:identifier>
  <dc:identifier>(SCI)229063600002</dc:identifier>
  <dc:identifier>(Compendex)5189072594</dc:identifier>
  <dc:identifier>(Scopus)2-s2.0-17444401739</dc:identifier>
  <dc:language>eng</dc:language>
  <dc:relation>Composite : Part B : Engineering. 1359-8368. 2005. Vol. 36, n.º 5. p.
    372-383</dc:relation>
  <dc:relation>FEUP > SDI > RiFEUP</dc:relation>
  <dc:identifier>http://eos.fe.up.pt:1801/webclient/DeliveryManager?
    pid=33179&custom_att_2=simple_viewer</dc:identifier>
</record>
</metadata>
</record>
- <record>
- <header>
  <identifier>oai:digitool.fe.up.pt:32825</identifier>
  <timestamp>2008-11-12T02:16:56Z</timestamp>
  <setSpec>feup</setSpec>
```

Nas UO que, ao contrário da FEUP, apenas possuem o SIGARRA como repositório de documentos e de metadados, pode ser usada a funcionalidade de validação do módulo das publicações⁸. Esta funcionalidade, desde que utilizada por alguém com o *know-how* necessário para fazer validações, incluindo a validação da disponibilização dos documentos em *open access* em conformidade com as políticas das revistas, pode ser útil para deixar de haver inconsistência entre os metadados e os documentos no SIGARRA e, conseqüentemente, aumentar a qualidade do repositório da UP.

Contudo, coloca-se ainda um outro problema. O problema da maioria das publicações no SIGARRA não terem o ficheiro associado. A funcionalidade de exportar para o ALEPH, existente no módulo das publicações após ter sido concluída a validação, poderá ser útil para minimizar este problema. Com estes registos no ALEPH, os serviços de biblioteca das diferentes UO, poderiam disponibilizar o ficheiro que estivesse em conformidade com as políticas das revistas respectivas. Em seguida bastaria “suspender a validação”, adicionar o ficheiro à publicação, e validar novamente a publicação. Desta forma, poderia ser aumentada a quantidade de informação disponibilizada pelo repositório da UP. Obviamente que, se a UO não possuir o ALEPH, todo este processo será mais difícil.

No repositório da UP existem conteúdos que têm o acesso condicionado – as colecções ALFA. Estes conteúdos, se forem publicados em OAI-PMH no mesmo *set* dos restantes conteúdos que estão em *open access*, inviabilizam a integração do repositório da UP no DRIVER. Uma forma simples de resolver este problema, sem ter que se criar vários *sets*, pode passar pela inclusão do campo <dc:rights> nos metadados destes conteúdos. Assim, seria possível filtrá-los, ou seja, não seriam incluídos no *set* do OAI.

⁸ <http://moodle.up.pt/course/view.php?id=69&username=guest&password=quest#5> - ponto 4 [consulta: 13 de Novembro de 2008]

4. CONCLUSÕES

Após a conclusão deste estudo, constatou-se que é possível a integração dos metadados do DigiTool da FEUP no DSpace da UP. Para tal, basta que sejam utilizados os mapeamentos correctos para que a informação disponibilizada pelo DigiTool seja correctamente importada e mapeada para o DSpace.

No caso da Unidade Orgânica FEUP, chegou-se à conclusão que seria vantajoso fazer o DSBI ao DigiTool e cortar a sua ligação ao SiFEUP. Actualmente o repositório da UP com o DSBI ao SiFEUP tem 627 documentos na comunidade FEUP – Faculdade de Engenharia. Se o DSBI fosse correctamente efectuado ao DigiTool, o repositório da UP passaria a ter 1306 dissertações de mestrado, 679 teses de doutoramento, 43 trabalhos académicos e 329 artigos em revistas científicas internacionais na comunidade FEUP – Faculdade de Engenharia, totalizando assim 2357 documentos. A estes objectos poderiam ainda ser somados 1016 objectos do acervo digital do museu da FEUP.

Chegou-se à conclusão que a falta de validação no módulo de publicações do SIGARRA e/ou no repositório da UP leva a inconsistências na informação disponibilizada e pode acarretar problemas legais para os docentes/investigadores ou até para a Universidade do Porto.

Concluiu-se também que a forma como o módulo de publicações do SIGARRA está pensado, juntamente com a forma como o DSBI interage com o SIGARRA, origina pouca importação de informação para o repositório da UP. Torna-se imperativo que, os serviços de documentação e informação das diversas Unidades Orgânicas, comessem a assumir um papel activo na questão das publicações, de forma a possibilitarem a colocação da Universidade do Porto no primeiro lugar dos rankings de publicações científicas em Portugal.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Martins, Ana Bela; Rodrigues, Eloy; Nunes, Manuela Barreto – *Repositórios de informação e ambientes de aprendizagem: Criação de espaços virtuais para a promoção da literacia e da responsabilidade social*. Newsletter RBE n.º 3. [em linha]. [s.l.: RBE, 2008].

<http://www.rbe.min-edu.pt/news/newsletter3/repositorios.pdf> . [consulta: 13 de Novembro de 2008].

6. BIBLIOGRAFIA

- <http://www.driver-repository.eu/PublicDocuments> [consulta: 10 de Novembro de 2008]
- <http://validator.driver.research-infrastructures.eu/validatorWeb/> [consulta: 10 de Novembro de 2008]
- <http://www.openarchives.org/> [consulta: 10 de Novembro de 2008]
- <http://www.driver-repository.eu/> [consulta: 10 de Novembro de 2008]
- <http://dublincore.org/> [consulta: 10 de Novembro de 2008]
- <http://moodle.up.pt/course/view.php?id=69&username=guest&password=guest#5> - ponto 4 [consulta: 13 de Novembro de 2008]
- Feijen, Martin, et al.; *Serviços de Documentação da Universidade do Minho, trad. - Exposição de recursos textuais com o protocolo OAI-PMH: Directrizes para fornecedores de conteúdos*. [em linha]. [s.l.]: DRIVER, 2008.
http://www.driver-support.eu/documents/DRIVER_guidelines_1%2000_pt_PT.pdf. [consulta: 10 de Novembro de 2008].

- Feijen, Martin - *Conselhos para Implementação das Directrizes DRIVER*. [em linha]. [s.l.]: DRIVER, 2008.
http://www.driver-support.eu/documents/Advice_for_implementation_of_the_DRIVER_guidelines_pt_PT.pdf. [consulta: 10 de Novembro de 2008].
- Graça, Pedro - *Publicação de conteúdos usando o "Publishing Based OAI Data Provider" do DigiTool*. [Relatório Técnico]. Porto: FEUP, 2008.
- Ramos, Rui - *Sugestões para futuros desenvolvimentos do módulo de publicações*. [Relatório Técnico]. Porto: UP, 2008.