

REPOX: Uma infra-estrutura XML para a PORBASE

Nuno Freire

Biblioteca Nacional
Campo grande 83
1741-081 Lisboa

E-mail: nuno.freire@bn.pt

Hugo Manguinhas, José Borbinha

INESC ID
Rua Alves Redol 9, Apartado 13069
1000-029 Lisboa

E-mail: hugomanguinhas@hotmail.com, jlb@ist.utl.pt

RESUMO

Este artigo descreve o conceito inovador de Repositório de Metadados, definido como uma infra-estrutura aberta para armazenar, preservar e gerir colecções de registos de metadados. Esta infra-estrutura é desenhada de acordo com os requisitos do modelo de referência ISO 14721:2003 Open Archival Information System (OAIS), para sistemas de informação abertos para arquivos com este objectivo. Um Repositório de Metadados pode ser utilizado para gerir transparentemente grandes volumes de metadados independentemente do seu esquema ou formato, sendo utilizado por outros serviços que, por sua vez oferecem serviços sobre esses mesmos dados. As suas principais funções são a submissão, armazenamento, preservação a longo prazo e recuperação da informação. O armazenamento de toda a informação é efectuado no sistema de ficheiros local, em estruturas XML, garantindo assim o máximo de independência em relação ao ambiente computacional. A relevância e viabilidade deste conceito são demonstradas através do sistema operacional REPOX instalado na Biblioteca Nacional, onde se gerem colecções de metadados bibliográficos e de autoridade do catálogo colectivo PORBASE, e metadados descritivos e de autoridade do sistema de informação do Arquivo da Cultura Portuguesa Contemporânea. O sistema REPOX suporta já um número considerável de serviços na Biblioteca Nacional, com destaque para os serviços de OAI-PMH e Z39.50 da Biblioteca Nacional Digital (BND) e da PORBASE, o serviço URN.PORBASE.ORG, a interface para o serviço Google Scholar, assim como outros serviços internos de suporte à operação e controlo de qualidade na PORBASE e de publicação dos metadados da BND. O REPOX permitiu o desenvolvimento flexível de todos estes serviços de um modo independente das características dos restantes sistemas de informação que suportam os processos de criação e manutenção dos metadados em si.

PALAVRAS-CHAVE: Metadados, Repositórios de metadados, PORBASE.

INTRODUÇÃO

Este artigo descreve o conceito inovador de Repositório de Metadados, definido como uma infra-estrutura aberta para armazenar, preservar e gerir colecções de registos de metadados, independentemente do seu esquema ou formato.

Esta infra-estrutura é desenhada de acordo com os requisitos do modelo de referência ISO 14721:2003 Open Archival Information System (OAIS), orientada

aos sistemas de informação abertos para arquivos com este objectivo.

Este artigo começa por discutir detalhadamente o conceito de Repositório de Metadados, seguindo-se com a descrição do sistema operacional REPOX instalado na Biblioteca Nacional (BN), onde se gere colecções de metadados bibliográficos e de autoridade do catálogo colectivo PORBASE, metadados descritivos e de autoridade do sistema de informação do Arquivo da Cultura Portuguesa Contemporânea, e metadados descritivos da Biblioteca Nacional Digital. Serão também descritos alguns dos serviços que são suportados pelo sistema REPOX na Biblioteca Nacional.

O CONTEITO DE REPOSITÓRIO DE METADADOS

O conceito de repositório de metadados surge da aplicação de metodologias de desenho de sistemas de informação, denominadas arquitecturas empresarias, ao ambiente das bibliotecas, arquivos, museus, centros de documentação, etc. Aliadas a estas metodologias surgiu recentemente o conceito de arquitecturas orientadas a serviços, onde a palavra chave é flexibilidade! Seguindo este modelo, o desenho e desenvolvimento de sistemas de informação baseia-se numa visão do mundo em que serviços são montado e reutilizados para dar uma resposta rápida a novas tarefas e necessidades das organizações. Esta flexibilidade tem guiado a forma como os novos sistemas de informação estão a ser construídos actualmente.

Numa organização com uma arquitectura empresarial bem desenhada, a separação entre a informação e o sistema de informação deve ser clara. A reengenharia das arquitecturas de informação de acordo com este requisito tem sido uma prioridade nas organizações durante as duas últimas décadas. No entanto, esta questão não tem sido reconhecida de forma geral na área das bibliotecas, arquivos, etc. Aqui, num grande número de casos, os processos das organizações, as estruturas de informação e as aplicações têm sido misturadas num mesmo cesto de requisitos e soluções, dificultando assim a separação entre a informação e o sistema de informação.

É neste sentido que definimos um Repositório de Metadados como sendo um serviço em rede para preservação e recuperação de registos de metadados com os seguintes requisitos:

- Registo de esquemas: Este inclui funções para o registo de qualquer esquema de metadados que possa ser expresso em XML.

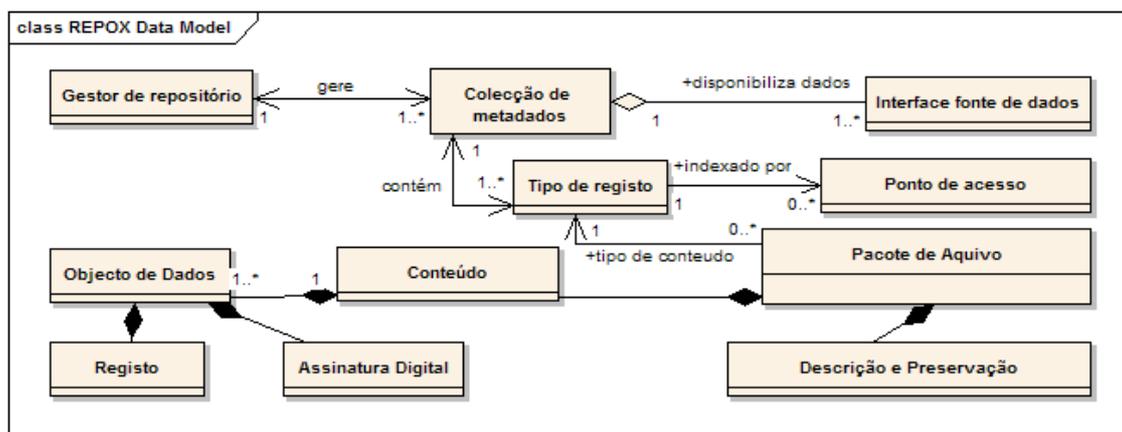


Ilustração 1 – O modelo de dados do REPOX

- **Submissão:** Este inclui funções para a submissão de conjuntos de registos de metadados pertencendo a um esquema previamente registado, assumindo que esses registos podem ser representados em XML. Este inclui também os serviços para sincronização com fontes de dados externas de forma independente da tecnologia que os cria ou exporta.
- **Armazenamento Ilimitado:** Este significa que a gestão do espaço de armazenamento, incluindo a sua capacidade deve ser gerida pelo repositório de metadados, e em nenhuma ocasião deve ser uma preocupação dos serviços externos.
- **Recuperação:** Este inclui todas as funções necessárias para a recuperação dos registos pelos seus identificadores (assegurando-se assim ligações persistentes).

Também podemos incluir na definição de Repositório de Metadados, o seguinte requisito opcional:

- **Persistência:** Este implica a capacidade de assegurar a persistência dos registos de metadados (preservação a longo prazo) e dos seus identificadores. O que significa que uma vez que um registo seja submetido deverá ser garantida sempre a sua recuperação e acesso. Inclui assim a gestão de versões dos registos bem como a sua identificação independente. Na prática este requisito indica que os registos nunca são actualizados ou apagados, apenas são acrescentadas novas versões.

Este repositório pode ser visto como um novo serviço primário com relevância para as arquitecturas dos sistemas de informação de bibliotecas, arquivos, centros de documentação e museus.

O MODELO DO REPOX

REPOX é uma implementação do conceito de Repositório de Metadados. A forma como os registos são codificados e geridos aborda os requisitos de preservação da OAIS[1], bem como os de [2] e [3], uma vez que oferece uma plataforma aberta sem a utilização de tecnologias proprietárias para a gestão dos processos

de preservação. É também robusto, flexível e oferece mecanismos para a auto descrição e validação dos metadados.

O Modelo de dados definido para o REPOX encontra-se na ilustração 1. Este prevê que o repositório possa gerir várias colecções de metadados. Estas devem ter pelo menos uma fonte de dados que estará associada a uma interface de fonte de dados. Estas interfaces são componentes de software responsáveis por obter os registos, verificar e codificá-los de acordo com o seu esquema XML, e entregá-los ao REPOX. A recolha de registos pode ocorrer com qualquer periodicidade, definida na configuração da colecção (no caso da BN, a periodicidade mais comum é diária).

Cada colecção está associada a um ou mais tipos de registos. Estes representam as entidades que são recolhidas da fonte de dados representadas em registos XML e geridas pelo REPOX. Os requisitos mínimos para um registo é apenas que este tenha um identificador unívoco dentro de uma colecção/tipo de registo.

Outros pontos de acesso podem permitir a recuperação de registos para além do acesso por identificador. Para esse propósito, podem ser associados pontos de acesso aos tipos de registo. Este recolhem, de cada registo, informação relevante para construir os respectivos índices de pesquisa (estes podem ser mantidos numa base de dados relacional por questões de eficiência uma vez que não fazem parte do modelo fundamental de preservação).

Cada registo gerido pelo REPOX é encapsulado num pacote de arquivo, definido tal como no modelo OAIS. É codificado em XML e armazenado num ficheiro, assegurando-se assim que o seu armazenamento se mantém independente de qualquer sistema de armazenamento. Estes pacotes de arquivo contêm toda a informação de origem e historial do registo que inclui todas as versões anteriores.

A autenticidade é um requisito importante do nosso problema. O caso de garantir a autenticidade a longo prazo de registos electrónicos foi abordada em [4] e [5]. No REPOX os registos são armazenados como ficheiros XML simples, no sistema de ficheiros local, o que não é suficiente para garantir o nível de autenticidade necessário. Por esta razão o REPOX utiliza assinaturas digitais para os registos (segundo a recomendação do W3C para sintaxe e processamento de assinaturas

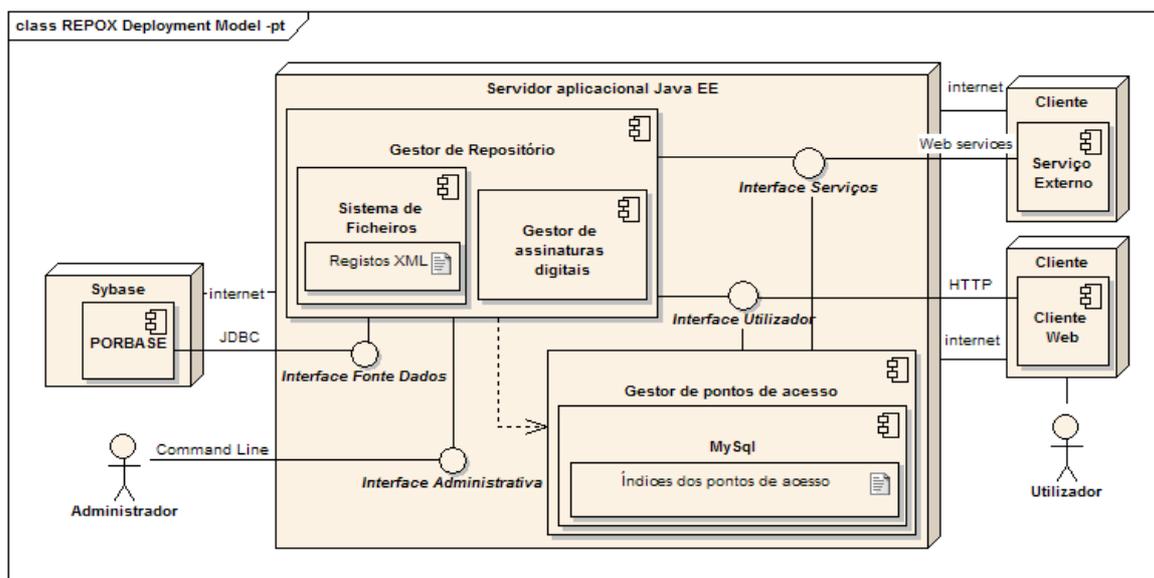


Ilustração 2 – Instalação REPOX na Biblioteca Nacional

digitais em XML¹) e implementa processos de verificação de autenticidade dos registos.

A ARQUITECTURA DO REPOX

O componente principal da infra-estrutura REPOX é o gestor do repositório. A ilustração 2 mostra um exemplo da instalação deste componente como uma aplicação Java EE² implementando o modelo já descrito na secção anterior.

O gestor do repositório recolhe os registos das fontes de dados por via das respectivas interfaces. Os registos são assinados digitalmente pelo gestor de assinaturas digitais, encapsulados em pacotes de arquivo e são arquivados no sistema de ficheiros. Após este processo, o gestor de pontos de acesso cria e armazena os índices de pesquisa numa base de dados relacional MySQL.

O sistema pode ser consultado por utilizadores através de uma interface para pesquisa, navegação e acesso aos registos individualmente, mostrando toda a informação disponível no repositório. Ferramentas de linha de comando permitem a um administrador do sistema criar, actualizar ou eliminar pontos de acesso; exportar registos; despoletar manualmente a execução de recolhas de metadados das fontes de dados externas, etc. Um conjunto semelhante de serviços são disponibilizados através de web services, para serem utilizados por máquinas, em outros serviços externos que utilizam o repositório como fonte de metadados. Exemplos de serviços externos são descritos, para o caso da Biblioteca Nacional, nas secções seguintes.

O REPOSITÓRIO DA BN: COLECÇÕES

A infra-estrutura REPOX encontra-se em funcionamento BN. Entre as colecções de metadados aí geridas importa destacar três: a PORBASE, a Biblioteca Nacional Digital e o Arquivo de Cultura Portuguesa Contemporânea, esta última ainda em fase

experimental.

A PORBASE – Base de Dados Bibliográfica Nacional³, é um catálogo colectivo em linha das bibliotecas portuguesas, sendo actualmente a maior base de dados bibliográficos do país, reflectindo as colecções de cerca de 170 bibliotecas Portuguesas, incluindo a BN, a entidade que tem a responsabilidade da sua coordenação.

O sistema integrado de gestão de bibliotecas que suporta a PORBASE armazena os dados num sistema de gestão de base de dados relacional. É também sobre este motor de bases de dados que assenta o catálogo em linha que disponibiliza a PORBASE para consulta através da Internet.

A PORBASE é uma importante fonte de informação cujo potencial tem sido subaproveitado, problema que a BN tem encarado desenvolvendo novos serviços baseados no repositório de metadados REPOX. A segurança e preservação do dados da PORBASE são também questões importantes abordadas pela BN.

A PORBASE é mantida por um sistema de informação proprietário desenhado para suportar os processos tradicionais de catalogação e pesquisa. O esquema de dados deste sistema está optimizado para os requisitos tradicionais de suporte a estas actividades. Esse esquema nem sempre é compatível com os requisitos dos novos serviços que se pretendem desenvolver. Para além disso a possível utilização do mesmo servidor que é usado para catalogação e pesquisa pelo público iria sobrecarregar esse para tarefas que podem muitas vezes ser executadas noutro servidor. Finalmente, a replicação dos dados, codificados num esquema aberto e armazenados numa infra-estrutura objectiva, é uma resposta às questões de segurança e preservação.

A ilustração 3 mostra a instanciação de duas colecções de dados da BN no modelo de dados do REPOX. Para a colecção de dados da PORBASE, dois tipos de registos são geridos: UNIMARC⁴ bibliográfico e UNIMARC

¹ <http://www.w3.org/TR/xmlsig-core/#sec-ignatureAlg>

² <http://java.sun.com/javase/index.jsp>

³ <http://www.porbase.org>

⁴ <http://www.unimarc.info>

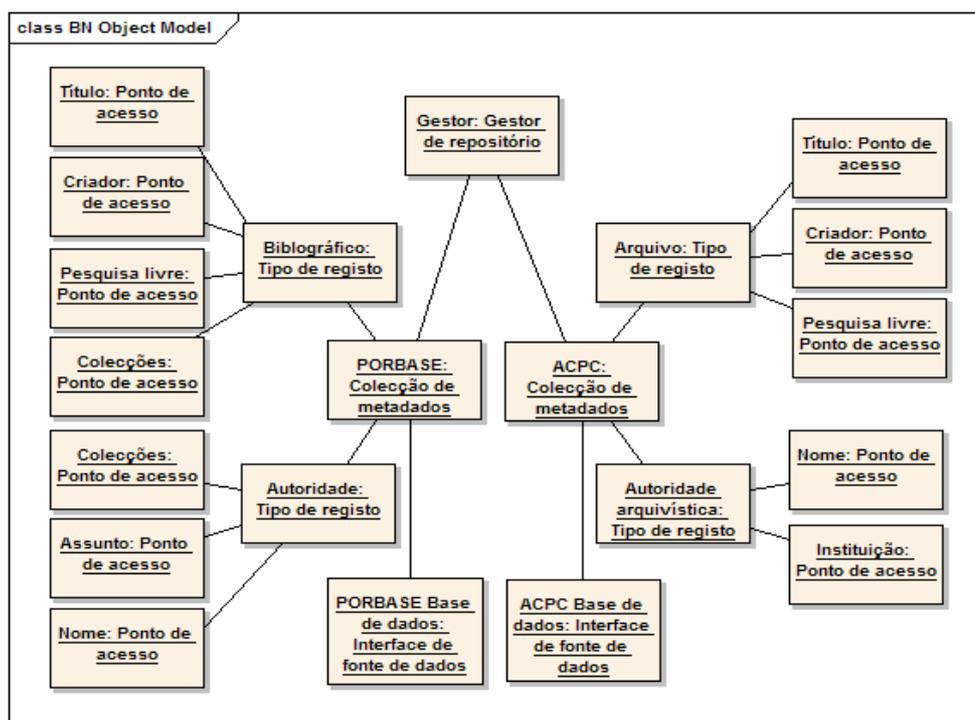


Ilustração 3 – Duas colecções de metadados da Biblioteca Nacional

autoridades.

Para os registos bibliográficos, existem pontos de acesso que permitem a recuperação de registos por título, autores, identificadores (ISBN, ISSN, cota, número de depósito legal, etc.), bem como para outros casos específicos que são requisitados por serviços que utilizam o repositório. Para os registos de autoridade, existem pontos de acesso para nomes de pessoas, nomes de instituições, assuntos, datas, etc.

Os dados do sistema de informação da PORBASE são mantidos num sistema de gestão de bases de dados SyBase. A interface de fonte de dados do REPOX para esta base de dados extrai os registos bibliográficos e de autoridade através duma ligação directa. Os registos estruturados em UNIMARC, são recolhidos diariamente e codificados em XML segundo o esquema MarcXchange⁵. Estas duas colecções contam com mais de três milhões de registos (incluindo tanto registos bibliográficos como de autoridade), totalizando cerca de 28 *gigabytes* codificados em XML.

Ambos os tipos de registos são organizados em colecções por um ponto de acesso designado “colecção”. Exemplos de colecções para os registos bibliográficos são as que são construídas para cada biblioteca cooperante da PORBASE (uma colecção por biblioteca). Outro exemplo são as colecções criadas para o TEL – The European Library⁶, utilizando os critérios definidos para esse propósito. Outras colecções existem com propósitos variados, baseadas assuntos, locais, datas, línguas, etc. Para esta colecção em particular, o sistema actual mantém um total de 33 pontos de acesso e gere 21 colecções de registos bibliográficos e 4 de registos de autoridade.

Uma outra colecção de metadados gerida pela infraestrutura REPOX da BN, ainda em fase experimental, diz respeito ao Arquivo de Cultura Contemporânea Portuguesa⁷ (ACPC). O ACPC reúne a colecção de espólios literários da BN tendo-se iniciado em 1981 com seis espólios - Almeida Garrett, Eça de Queirós, Jaime Batalha Reis, Camilo Pessanha, Fernando Pessoa e António Pedro. A estes se têm vindo juntar, ao longo do tempo, outros nomes representativos do universo literário e histórico-cultural português.

O ACPC utiliza um sistema de gestão de arquivo que utiliza uma base de dados Microsoft SQL Server para armazenamento de dados. Assim, tal como acontece na colecção da PORBASE, os registos de metadados são recolhidos diariamente através de uma ligação directa à base de dados. Os registos de arquivo encontram-se num formato que segue as regras ISAD(G)⁸, e também registos de autoridade de acordo com as ISAAR(CPF)⁹. Após a sua recolha, os registos de arquivo são codificados em EAD¹⁰ (versão de 2002), e os registos de autoridade em EAC¹¹. Formatos estes em que os registos são armazenados no repositório de metadados. Uma outra colecção diz respeito à BND - Biblioteca Nacional Digital. Esta colecção agrupa registos relativos a obras da BND de dois tipos: registos bibliográficos UNIMARC da PORBASE (estes são replicados a partir da colecção da PORBASE gerida pelo REPOX); e registos que por alguma razão não seguiram as regras Portuguesas de catalogação, e que

⁵ <http://www.bs.dk/marcxchange/>

⁶ <http://www.theeuropeanlibrary.org/>

⁷ <http://acpc.bn.pt>

⁸ http://www.ica.org/biblio/cds/isad_g_2e.pdf

⁹ http://www.ica.org/biblio/isaar_eng.html

¹⁰ <http://www.loc.gov/ead/>

¹¹ http://eprints.rclis.org/archive/00000316/01/pitti_eng.pdf

existem na BND em vários esquemas, embora sendo principalmente Dublin Core

O REPOSITÓRIO DA BN: SERVIÇOS EXTERNOS

Um elevado número de serviços na Biblioteca Nacional estão já em produção estável utilizando a infra-estrutura REPOX. Alguns dos serviços aqui descritos já existiam antes da entrada em funcionamento desta infra-estrutura, no entanto, outros serviços apenas foram tornados viáveis apenas com a entrada em funcionamento da infra-estrutura REPOX.

- URN.PORBASE.ORG** - Acesso à PORBASE por Identificadores Univocos (<http://urn.porbase.org>): este serviço disponibiliza os registos da PORBASE para as bibliotecas portuguesas em vários formatos através de uma interface por http. Na ilustração 4 encontra-se um exemplo da interface para utilizadores deste serviço. O serviço URN tira partido dos índices existentes no REPOX para permitir que os registos sejam obtidos através de vários identificadores (ISBN, ISSN, número de depósito legal, etc.). Os registos são disponibilizados em vários esquemas (UNIMARC, Dublin Core) e em vários formatos de codificação (XML, HTML, texto, etc.). Estes formatos são gerados em tempo real através de transformações XSLT a partir do registo armazenado no REPOX. Este serviço é acedido principalmente por sistemas de gestão de bibliotecas, especialmente de bibliotecas portuguesas, que importam os registos directamente da PORBASE para as bases de dados locais, facilitando assim o trabalho de catalogação aí levado a cabo.

Registos criados/alterados nos últimos 90 dias pelos utilizadores da área "Área de Aquisições"

Total de registos sem erros: 1391
Total de registos com erros: 265 (075 erros)

Nome	Utilizador	Registos sem erros	Registos com erros	# erros
Biblioteca Nacional de Lisboa	anonimo	0	0	0
Biblioteca Nacional de Coimbra	anonimo	308	20	90
Biblioteca Nacional de Évora	anonimo	266	0	15
Biblioteca Nacional de Faro	anonimo	0	0	0
Biblioteca Nacional de Vila Rica	anonimo	287	16	61

Relatório de erros do registo NCB:1156479

Data: 16-03-2006 01:31
Hora: (actualizar)

Erros:

Campo(s)	Descrição
801	Ausência do campo 801
200	2º indicador do campo 700 inválido

Ilustração 5 – Serviço Qualicat

Acesso à PORBASE por Identificadores Univocos

Acesso a Registos Bibliográficos e de Autoridade

Passo 1: Valor do Identificador: 1
Passo 2: Espaço do identificador: Registos Bibliográficos: Identificador de Registo, Cota, Nº de Depósito Legal, ISBN, ISSN, PURL; Registos de Autoridade: Identificador de Registo
Passo 3: Esquema: UNIMARC, Dublin Core, Dublin Core Qualificado, MODS, UNIMARC Autoridades, CAC
Passo 4: Forma: Texto (normal), XML (código), MARCXML (código), MarcXChange (código), ISBD (texto), ISO 2709 (código), ISO 2709 (texto), RDF (código)
Passo 5: Procurar

URL: <http://urn.porbase.org/urn/unimarc/marcxchange?id=1>

```
<record format="Unimarc" type="bibliographic">
<leader>01463cam 2200397 450 </leader>
<controlfield tag="001">323613</controlfield>
<controlfield tag="005">20030117160300.0</controlfield>
- <datafield ind1=" " ind2=" " tag="095">
<subfield code="a">PTBN00339700</subfield>
</datafield>
- <datafield ind1=" " ind2=" " tag="100">
<subfield code="a">198804 26d1572 k y0pera0103 bac</subfield>
</datafield>
- <datafield ind1="0" ind2=" " tag="101">
<subfield code="a">por</subfield>
</datafield>
- <datafield ind1=" " ind2=" " tag="102">
<subfield code="a">PT</subfield>
<subfield code="b">Lisboa</subfield>
</datafield>
```

Ilustração 4 – O serviço URN.PORBASE.ORG

- Qualicat:** Este é um serviço interno da BN que visa assegurar o controlo de qualidade diário do registos bibliográficos da BN, de acordo com os requisitos do formato UNIMARC, as regras portuguesas de catalogação e também algumas regras específicas da PORBASE. Este serviço disponibiliza relatórios quantitativos e qualitativos dos registos trabalhados pelos utilizadores como se mostra na ilustração 5.
- OAI-PMH¹² Service** (<http://oai.bn.pt>): Este serviço disponibiliza os registos da PORBASE em projectos de cooperação, permitindo que cópias de registos existentes num outro local sejam mantidas actualizadas. É assim por exemplo através deste serviço que o portal da TEL – *The European*

The European Library

PESQUISA COLEÇÕES TESOUROS BIBLIOTECAS SOBRE

Está a pesquisar em:

- lista de colecções por omissão
- mapas & atlas, cartografia
- fotografias, cartazes e imagens
- retratos
- literatura infantil
- obras digitalizadas
- jornais e periódicos
- manuscritos
- colecções de música
- religião
- artigos científicos
- teses e dissertações
- selecção personalizada

selecione as suas próprias colecções:

- navegar em todas as colecções
- navegar nas colecções por assunto
- pesquisar colecções pela descrição

A Biblioteca Europeia destina-se a pesquisar as existências das bibliotecas nacionais europeias

Ilustração 6 – O portal de A Biblioteca Europeia

¹²<http://www.openarchives.org/>

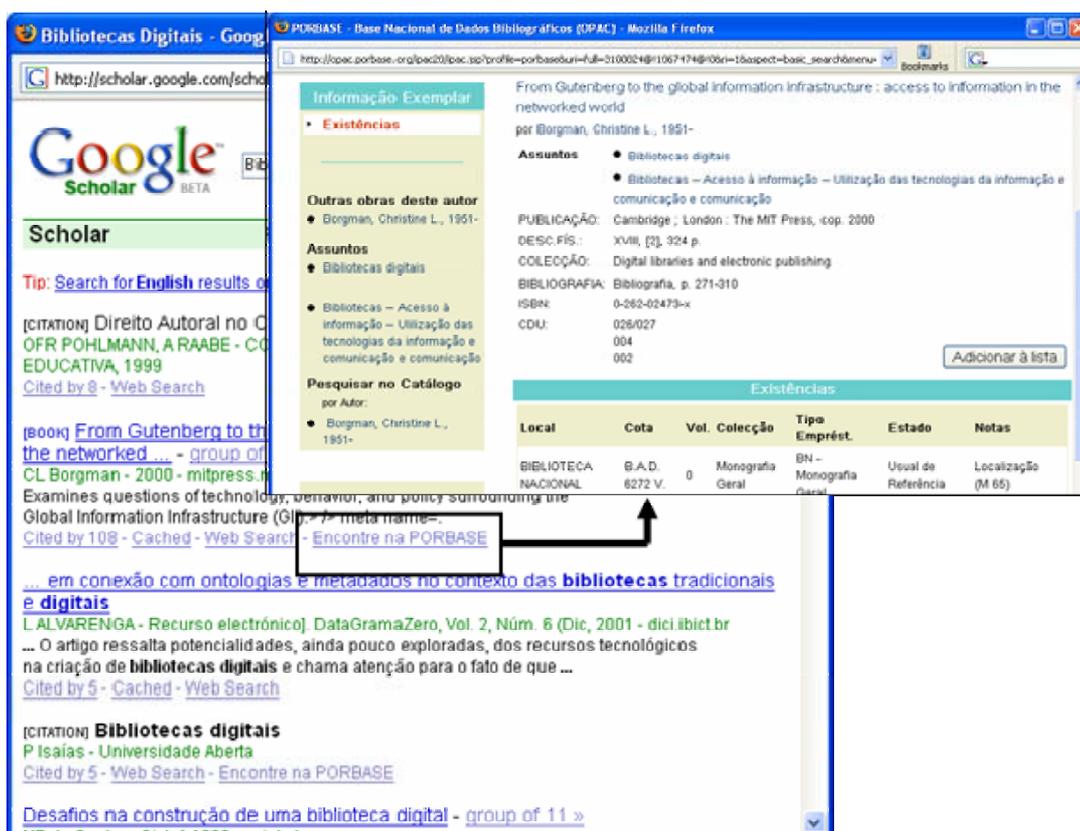


Ilustração 7 – Acesso à PORBASE pelo Google Scholar

Library¹³, recolhe regularmente os registos da PORBASE, e os disponibiliza para pesquisa e acesso (ilustração 6)

O servidor OAI-PMH obtém os registos através do REPOX, organizados por colecções que são geridas pelos pontos de acesso de colecção. Este serviço tem também a capacidade de converter os registos entre vários formatos disponibilizando os registos bibliográficos UNIMARC da PORBASE noutros formatos. No caso das colecções para a TEL, os registos do são transformados para o formato de dados da TEL. Outro exemplo, é o portal de manuscritos do CERL¹⁴ que obtém os registos no formato MODS¹⁵.

- **Google Scholar**¹⁶: A PORBASE foi uma das bases de dados bibliográficas pioneiras a partilhar os seus registos bibliográficos com a Google Scholar para indexação. O mecanismo de partilha dos registos foi construído com base na infra-estrutura REPOX. Os utilizadores portugueses do Google Scholar podem agora encontrar ligações para o OPAC da PORBASE quando as suas pesquisa encontrem resultados que existem numa das bibliotecas

cooperantes da PORBASE, como se mostra na ilustração 7.

- **IRIS**: Este serviço suporta a actualização na PORBASE das bases de dados bibliográficas das bibliotecas cooperantes. Implementa um processo de trabalho e oferece funcionalidades de controlo de qualidade, correcção automática de erros nos registos, detecção de duplicados, relatórios, e integração automática na PORBASE de registos bibliográficos e de existências. O ecrã principal do processo de integração e um resumo de um relatório de pré-processamento de uma base de dados mostra-se na ilustração 9.

Relatório resumido de: 2005-11-25
Ver [relatório detalhado](#)

Utilizador	Bibliográficos Criados	Bibliográficos Alterados	Bibliográficos Apagados	Autoridades Criadas	Autoridades Alteradas	Autoridade Apagadas
aalves	0	28	0	10	4	0
acarol	0	13	0	14	14	0
apires	0	18	0	11	7	0
asantos	33	1	0	0	0	0
barbara	0	44	0	15	3	0
bpmp02	0	15	0	3	0	0
bpmp03	2	0	0	0	0	0
bpmp04	3	0	0	1	0	0
bpmp05	0	0	0	0	0	0
bpmp06	8	5	0	1	0	0
caetano	0	8	0	10	12	0
dfontes	0	9	0	2	0	0
ecalapa	0	14	0	10	0	0
fatima	22	0	0	0	0	0
freitas	5	1	0	0	0	0
goreti	4	8	0	1	0	0
icosta	0	10	0	4	3	0
ines	0	73	0	0	1	0
ipuqa	0	6	0	13	8	0

Ilustração 8 – Relatório de actividade da PORBASE

¹³<http://www.theeuropeanlibrary.org>

¹⁴<http://www.cerl.org/Manuscripts/CERL%20Manuscripts%20Portal.htm>

¹⁵<http://www.loc.gov/standards/mods/>

¹⁶ <http://scholar.google.com>

Lotes de Registos Pendentes (Processo: Lotes Parciais)

[Submeter novo lote](#)Ver lotes do processo

Lote	Data	Nº registos	Pré-processamento	Detecção de Duplicados	Integração de Novos Registos	Integração de Novos Exemplares
BPEV (1 lote)						
bpev001_BC.iso	22-11-2005	21189	[relatório]	[relatório]	[concluído]	
CCRCEN (1 lote)						
ccrcen_BC.iso	22-11-2005	11504	[relatório]	[relatório]	[concluído]	
CENCAL (1 lote)						
cencal_BC.iso	24-10-2005	3811	[relatório]	[relatório]	[concluído]	
CMPEN (1 lote)						
cmpen003.iso	22-11-2005	3673	[relatório]	[relatório]	[concluído]	
ESTB (1 lote)						
estb_BC.iso	24-10-2005	1348	[relatório]	[relatório]	[concluído]	
MTPJB (2 lotes)						
mtpjb003a.iso	25-11-2005	3000	[relatório]	[relatório]	[concluído]	
mtpjb003b.iso	25-11-2005	4800	[relatório]	[relatório]	[concluído]	
UALLC (1 lote)						
uallc001_BC.iso	18-11-2005	18897	[relatório]	[relatório]	[concluído]	
UNLCHC (1 lote)						
unlchc002_BC.iso	22-11-2005	12769	[relatório]	[relatório]	[concluído]	

© Biblioteca Nacional, Lisboa, 2004

PORBASE
Base Nacional de Dados Bibliográficos

Relatório do Pré-processamento

Lote
bpev001_BC.iso (formato PORBASE5)

Relatórios
[Relatório de conteúdo](#)
[Relatório de erros](#)

Operações de Correção
Lote resultante: 21189 registos

Filtro analíticos
Lote de analíticos: 0 registos
Lote sem analíticos: 21189 registos

Filtro Bloco 4XX
Lote com bloco 4XX: 0 registos
Lote sem bloco 4XX: 21189 registos

Ilustração 9 – IRIS: Integração de bases bibliográficas dos cooperantes da PORBASE

- **Relatórios de actividade:** Este serviço cria relatórios sobre a actividade de catalogação na PORBASE. Os utilizadores do sistema são notificados por email de todas as operações (criação, alteração, remoção) levadas a cabo nas últimas 24 horas, sobre os registos bibliográficos, autoridades e exemplares. Pode-se ver uma fracção de um relatório na ilustração 8. Estes dados são obtidos a partir do REPOX após a sua sincronização diária com a PORBASE.

CONCLUSÃO

O sistema REPOX entrou em funcionamento operacional no dia 1 de Outubro de 2005, seguindo-se a implementação de novos serviços bem como a reengenharia dos serviços já existentes. De imediato melhorou consideravelmente a disponibilidade e o desempenho do acesso aos dados pelos variados serviços da Biblioteca Nacional e das bibliotecas cooperantes da PORBASE. O sistema de gestão de bibliotecas da PORBASE ficou também livre de processar os acessos aos registos por parte de outras bibliotecas e de vários serviços que o sobrecarregavam já de modo bastante significativo.

Um outro factor de extraordinária importância é o facto de desta forma o REPOX representar ainda uma cópia de segurança dos dados da PORBASE, expressa em XML. Neste momento tal representa já um total de cerca de 3 milhões de registos, que assim estão

preservados, independentes de qualquer software ou hardware específico.

Estão em fase de análise de requisitos e desenho, serviços de estatísticas mais avançados da PORBASE (e do REPOX em geral), bem como o estudo de soluções de *data warehouse* para suportar serviços gerais de *data mining* e especificamente de suporte à decisão.

O REPOX começa a ser aplicado noutros contextos para além do da Biblioteca Nacional. Na arquitectura, do projecto DIGMAP¹⁷, está prevista a utilização de uma infra-estrutura REPOX como repositório de metadados para as colecções de cartografia das bibliotecas parceiras no projecto, bem como repositório de metadados geográficos sobre os quais se planeiam serviços para infra-estruturas de dados espaciais e sistemas de informação geográfica.

Muitos dos serviços construídos para a Biblioteca Nacional têm aplicabilidade imediata noutras bibliotecas. Como exemplos, temos as bibliotecas que se estão a juntar à Biblioteca Europeia e as bibliotecas parceiras do projecto DIGMAP. Para estas existe interesse na utilização do REPOX juntamente com os serviços de OAI-PMH e Z39.50, estando já a decorrer testes com a Biblioteca Real da Bélgica para a criação de um repositório de metadados para suportar serviços de OAI-PMH, Z39.50 e OpenURL para a sua base bibliográfica. Por esta razão, está a ser criado um pacote de software juntando REPOX com estes serviços para fácil utilização noutras bibliotecas.

¹⁷ <http://www.digmap.eu/>

NOTAS

- [1] Consultative Committee for Space Data Systems. OAIS - Reference Model for an Open Archival Information System (2002)
- [2] Electronic Resource Preservation and Access Network Urbino Workshop: XML for Digital Preservation. ERPANET (Oct. 2002) Retrieved Jan. 27, 2007 from <http://eprints.erpanet.org/archive/00000002/01/UrbinoWorkshopReport.pdf>
- [3] Boudrez, F.: XML and electronic record-keeping (2002) Retrieved Jan. 27, 2007 from http://www.expertisecentrumdavid.be/davidproject/teksten/XML_erecordkeeping.pdf
- [4] Gladney H.M.: Trustworthy 100-Year Digital Objects: Evidence After Every Witness Is Dead. ACM Transactions on Information Systems, Vol. 22, No. 3 (2004) 406–436.
- [5] Waugh, A., Wilkinson, R., Hills, B., Dell'oro, J.: Preserving Digital Information Forever. Proceedings of the fifth ACM conference on Digital libraries (2000): 175-184