

Agregação de Conteúdos para a The European Library e EUROPEANA

Gilberto Pedrosa, João Edmundo, José Borbinha

INESC-ID

Rua Alves Redol, nº 9
1000-029 Lisboa – Portugal

Tel: +351213100300

E-mail: {gilberto.pedrosa, joao.edmundo, jlb}@ist.utl.pt

Nuno Freire

The European Library, National Library of the
Netherlands

Prins Willem-Alexanderhof 5
2595 BE Den Haag – Nederland

Tel: +31703140310

E-mail: nuno.freire@ist.utl.pt

RESUMO

Nesta comunicação descreve-se a plataforma REPOX, com destaque para as suas possibilidades para uso por qualquer entidade que, num cenário de interoperabilidade em bibliotecas ou repositórios digitais, pretenda oferecer serviços de “data provider”, “service provider” ou agregador. A plataforma foi toda desenvolvida em linguagem JAVA, e é disponibilizada em código aberto (sendo por isso de uso livre). Está ainda desenhada e desenvolvida para poder funcionar em servidores Windows ou Linux. Descrevem-se ainda os cenários mais típicos de uso desta ferramenta, fornecendo-se detalhes relevantes para entidades potencialmente interessadas no seu uso. São descritas ainda, para referência, as formas de uso da ferramenta nas infraestruturas The European Library e Europeana, assim como noutras iniciativas como os projetos EuropeanaLocal e EuDML.

PALAVRAS-CHAVE: Interoperabilidade, Registos Bibliográficos, OAI-PMH, XML, “Data Providers”, “Service Providers”, Agregadores, TEL, Europeana

INTRODUÇÃO

Nesta comunicação é feita uma descrição da plataforma REPOX (<http://rebox.ist.utl.pt/>) e do seu uso em vários casos de interoperabilidade em bibliotecas digitais.

Os objetivos de desenvolvimento desta ferramenta são os de disponibilizar tecnologia fácil de instalar e usar para instituições que queiram recolher registos bibliográficos, queiram disponibilizar um agregador para gerir fontes de dados ou ainda oferecer simplesmente os seus registos bibliográficos. Todos os registos recolhidos são armazenados internamente em formatos XML, sendo suportados serviços de transformação de metadados que permitem a tradução de praticamente qualquer formato original para qualquer formato desejado.

REPOX suporta canais de recolha e disponibilização de registos por HTTP, Z39.50, FTP, OAI-PMH (<http://www.openarchives.org/pmh/>), ou outros canais relevantes. Inclui assim servidores e clientes, acrescidos

de funções próprias para gestão de processos comuns em fornecedores de dados (“data providers”) e agregadores em redes de bibliotecas digitais.

Esta plataforma foi inicialmente desenvolvida pela BNP – Biblioteca Nacional de Portugal, a pensar em necessidades internas e da rede PORBASE (FREIRE). Evoluiu entretanto para uma plataforma genérica, adaptada durante vários projectos financiados por programas comunitários associados às iniciativas The European Library e Europeana.

Para satisfazer as necessidades concretas dos serviços The European Library e Europeana, o trabalho realizado mais recentemente na plataforma REPOX centrou-se na vertente de agregação, tendo em vista a sua utilização como agregador central de metadados por essas entidades, o que é neste momento já uma realidade (isto é, os metadados neste momento recolhidos pelos serviços The European Library e Europeana são todos geridos por instalações REPOX).

Numa perspectiva inovadora, que se pretende ser já efectiva para o portal da The European Library no final de 2012, mas ainda experimental para o portal Europeana, foram desenvolvidos para o sistema REPOX serviços de recolha de conteúdos para indexação. Os principais objetivos passaram por tornar o sistema escalável, para que cada instalação possa gerir centenas de fontes de dados com vários milhões de registos. Em consequência foram melhoradas muitas das suas funcionalidades de gestão de recolhas, a reforçada a sua tolerância a erros (tais como quebras nas ligações pelas Internet, registos mal formatados, etc.).

O documento prossegue com uma descrição geral da plataforma e dos tipos de cenários onde a mesma pode ser usada.

CONTEXTO DE MOTIVAÇÃO PRINCIPAL – INICIATIVAS THE EUROPEAN LIBRARY E EUROPEANA

O portal Europeana (<http://www.europeana.eu/> – Figura 1) fornece um serviço de pesquisa de recursos culturais, digitalizados ou nascidos digitais, existentes por toda a Europa, tais como livros, pinturas, filmes, objetos de museus, registos de arquivos digitalizado, etc.

Atualmente possibilita a pesquisa de mais de 20 milhões de objetos de mais de um milhar de instituições distribuídas por quase todos os países do Conselho da Europa.

A Biblioteca Europeia (The European Library) é um serviço de pesquisa bibliográfica que agrega catálogos de 48 bibliotecas nacionais da Europa, em 35 línguas (<http://search.theeuropeanlibrary.org/> – Figura 2). Os recursos referenciados podem ser apenas físicos, digitalizados, ou nascidos digitais (livros, posters, mapas, registos de som, vídeos, etc.). Esta rede tem sido recentemente alargada a novos fornecedores que não apenas bibliotecas nacionais, nomeadamente aos membros das redes DRIVER – Networking European Scientific Repositories e CERL – Consortium of European Research Libraries.

Parte da tecnologia utilizada nos serviços de suporte à Europeana e The European Library é partilhada entre os dois serviços, em consequência de uma raiz tecnológica e organizacional comum (a iniciativa Europeana teve a sua génese na estrutura criada inicialmente para o portal TEL), e de projectos de desenvolvimento em que as duas iniciativas têm cooperado. Nesses desenvolvimentos têm estado ainda envolvidas regularmente outras entidades, nomeadamente unidades académicas e de investigação, como tem sido o caso em Portugal do IST – Instituto Superior Técnico (<http://www.ist.utl.pt>), através do Departamento de Engenharia Informática, e do INESC-ID (<http://www.inesc-id.pt>), através do Grupo de Sistemas de Informação.

Em termos simples, as infraestruturas dos portais Europeana e The European Library estão desenhadas para executar processos contínuos de recolha de metadados a partir de entidades fornecedoras, tais como bibliotecas nacionais no caso do TEL, e no caso do serviço Europeana ainda de arquivos, museus, videotecas ou fonotecas, outras bibliotecas além das nacionais, e outras instituições culturais.

A política do portal The European Library passa por recolher os metadados diretamente dos seus fornecedores originais, enquanto a política do portal Europeana é recolher os metadados a partir de entidades intermediárias, designadas de agregadores.

O objetivo do conceito de agregador é que estes venham a ser a interface das entidades que venham a ser criadas para representar fornecedores de regiões ou áreas geográficas, ou mesmo fornecedores sectoriais. Por exemplo, no caso das bibliotecas nacionais o serviço The European Library funciona já como agregador de metadados dessas bibliotecas para a Europeana. Há mesmo planos para que na realidade o serviço The European Library venha a ser o agregador de topo para a Europeana de todas as bibliotecas europeias, as quais se podem associar a agregadores de em múltiplos níveis inferiores.

Tecnicamente, um agregador pode ser ainda constituído não para fornecer directamente o serviço da Europeana, mas para fornecer outro agregador num nível superior, como será neste caso o The European Library, que

poderá ser assim o topo de uma árvore de agregadores (neste caso, das bibliotecas europeias).

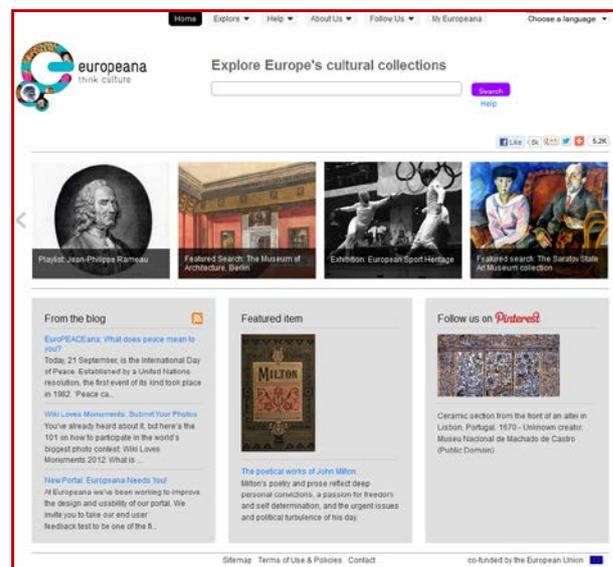


Figura 1: O portal EUROPEANA

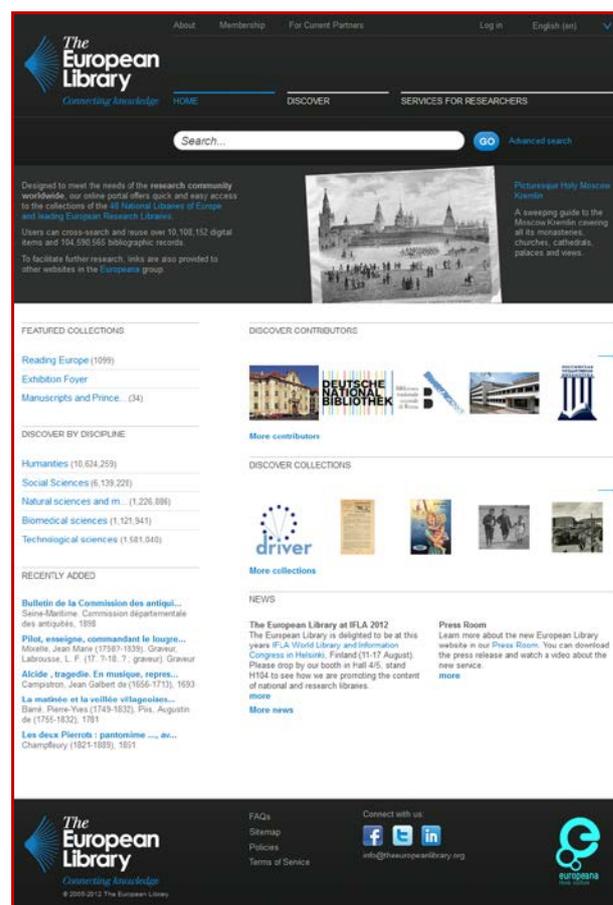


Figura 2: O portal The European Library

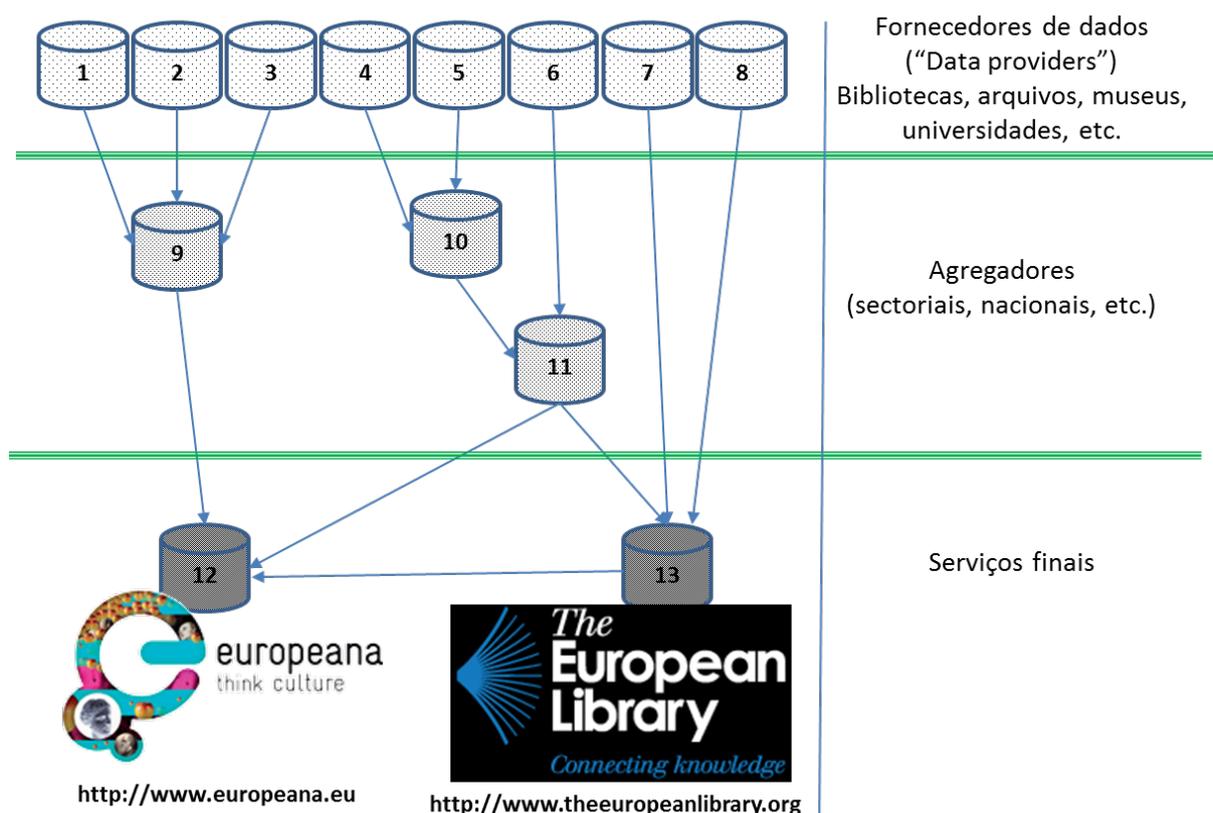


Figura 3: Arquitectura de recolha de dados onde a plataforma REPOX pode ser utilizada

CONCEITOS FUNDAMENTAIS

Os cenários de interoperabilidade já detetados como mais comuns em contextos de bibliotecas digitais levaram à consolidação e aceitação pela comunidade do dos seguintes conceitos básicos:

- Fornecedor de Dados (“Data Provider”) – entidade disposta a fornecer ou compartilhar um ou mais conjuntos de registos (coleções), como por exemplo uma biblioteca, um arquivo, etc.
- Fonte de dados (“Data Source”) – fonte de um conjunto de registos, que normalmente é um servidor OAI-PMH, mas para REPOX pode ser um caminho num sistema de arquivos local, ou qualquer outro serviço de transporte de dados, como o FTP, HTTP-GET, ODBC (acesso à base de dados on-line), ou mesmo Z39.50.
- Fornecedor de Serviços (Service Provider) – entidade que recolhe conjuntos de dados de um ou mais fornecedores de dados, para uma utilização pretendida.
- Agregador (“Aggregator”) – entidade que é ao mesmo tempo é fornecedor de serviços e fornecedor de dados. Como fornecedor de serviços, agrega fontes de dados que posteriormente irão ser fornecidas a outros fornecedores de serviços (como por exemplo a estrutura The European Library faz com as bibliotecas para a Europeana).

Antes dos metadados dos recursos culturais chegarem aos portais da The European Library e Europeana, eles podem passar por várias organizações intervenientes, tal como se mostra na Figura 3. Nessa figura, as entidades numeradas de 1 a 8 pretendem representar fornecedores de dados, cada um podendo oferecer uma ou mais fontes de dados. As entidades 9 a 11 pretendem ilustrar fornecedores de serviços que recolhem os dados das entidades anteriores. Como esses dados não se destinam a ser “consumidos” localmente, mas se destinam a ser de novo disponibilizados a outros fornecedores de serviços, o serviço que estas entidades 9 a 11 prestam é assim um serviço de agregador. A entidade 11 em especial é ilustrada a recolher dados de fornecedores originais e do agregador 10. A entidade 12 ilustra o papel da The European Library, que é ao mesmo tempo:

- Um fornecedor de serviços, que recolhe dados de outras entidades (aqui ilustradas como 7, 8, 11) e os utiliza num serviço próprio (o portal da The European Library).
- Um agregador de dados, neste caso para a Europeana.

As entidades que organizam e mantêm coleções de interesse cultural (bibliotecas, arquivos, museus, etc) são denominadas neste contexto por ‘fornecedores de dados’. Embora no caso da Europeana, os metadados se refiram apenas a objectos disponibilizados em linha, no caso da The European Library, estas coleções podem ser apenas bibliográficas, não necessitando que sejam objectos digitais disponíveis em linha.

Embora em certos casos os fornecedores de dados transfiram os seus metadados directamente para a

Europeana, o caso mais frequente é que eles passem primeiro por uma ou mais organizações intermédias, os denominados agregadores. Estes podem ser de várias naturezas, como projectos temporários de agregação de conteúdos, ou serviços sustentáveis sem termo.

Note-se também que a maioria dos agregadores são também serviços finais, que podem funcionar a nível regional, nacional, ou sectorial, e disponibilizam as colecções agregadas nos seus próprios portais. Um exemplo a nível nacional é o Registo Nacional de Objectos Digitais, ou RNOD (<http://rnod.bnportugal.pt>), que agrega colecções de entidades portuguesas, visando a coordenação e difusão desses recursos, a nível nacional e internacional.

Os agregadores podem também variar em relação aos serviços que disponibilizam aos fornecedores de dados ao nível da verificação dos metadados de acordo com os requisitos técnicos da Europeana, melhoria dos metadados, transformação de metadados, e disponibilização de acordo com os requisitos da Europeana, ou seja pelo protocolo OAI-PMH.

A iniciativa The European Library é um dos principais agregadores da Europeana, operando a nível europeu no sector das bibliotecas. É também um serviço final disponibilizando no seu portal, para além das colecções digitais agregadas para a Europeana, outras colecções bibliográficas.

A plataforma REPOX pode ser utilizada por qualquer um dos tipos de intervenientes acima referidos. Os fornecedores de dados podem utilizar a plataforma REPOX para disponibilizar as suas colecções de metadados por OAI-PMH, incluindo a sua transformação para diferentes formatos de metadados.

Os agregadores e serviços finais podem utilizar a plataforma REPOX para gerir os seus fornecedores de dados, que podem ser também agregadores, e todo o processo de recolha, manipulação e disponibilização das colecções de metadados. No caso da The European Library e da Europeana, a plataforma permite suportar estas funções em grande escala, suportando centenas de fontes de dados com centenas de milhões de registos.

REPOX – GESTOR DE AGREGAÇÃO E INTEROPERABILIDADE DE DADOS

A ferramenta utilizada atualmente nas infraestruturas Europeana e TEL para a recolha dos metadados é a plataforma REPOX. Esta plataforma foi desenvolvida inicialmente pelo IST e INESC-ID para a Biblioteca Nacional e para o projecto DIGMAP (<http://portal.digmap.eu/> – Figura 6).

Graças aos desenvolvimentos dos últimos dois anos, a plataforma REPOX é neste momento um componente de infraestrutura sofisticado e complexo. Trata-se de uma ferramenta em código aberto e uso livre (Figura 6) para publicação ou recolha de metadados e conteúdos, suportando vários protocolos, entre eles o OAI-PMH.

Na sua versão actual, a plataforma REPOX está vocacionada especialmente para metadados bibliográficos. No entanto o seu suporte a processos de transformação dos esquemas de metadados permite que

possa ser usada em qualquer contexto onde os esquemas dos metadados utilizados sejam conhecidos

CONTEXTO DE DESENVOLVIMENTO – PROJECTO EUROPEANA LIBRARIES

Neste momento decorre uma iniciativa concreta, o projeto Europeana Libraries (<http://www.europeana-libraries.eu/> – Figura 3), que pretende adicionar ao portal das The European Library um número largo de bibliotecas universitárias do consórcio LIBER (representando vários países europeus).

Para além de ter como objectivo alargar deste modo o número de fornecedores de metadados para os portais TEL e Europeana, este projecto tem ainda o objectivo inovador de, na sequência dos processos de recolha dos metadados, criar as soluções técnicas para se recolher também todos os conteúdos referenciados por esses metadados e que existam em formato digital.



Figura 4: O projeto Europeana Libraries

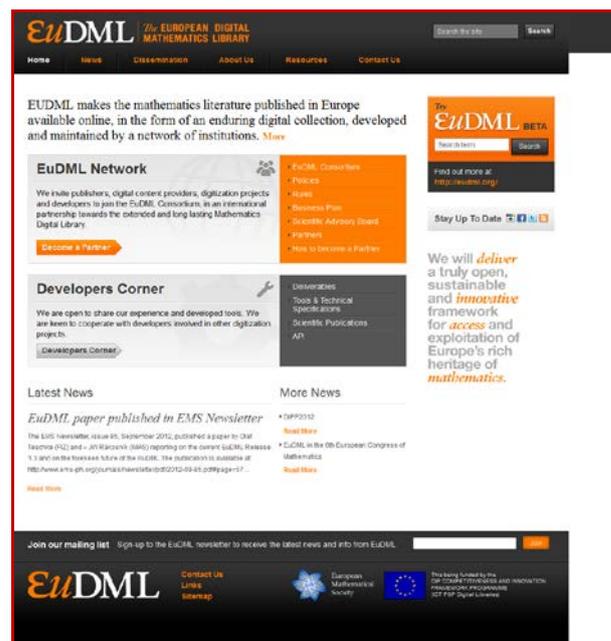


Figura 5: O projeto EuDML



REPOX
Data Aggregation and Interoperability Manager



INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa



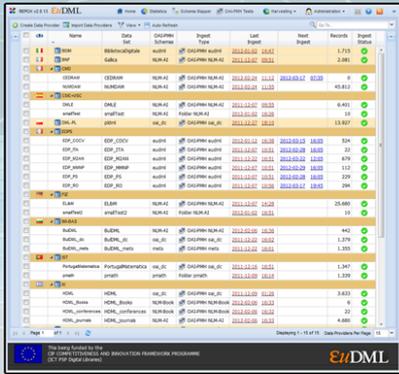
Home - **Home**

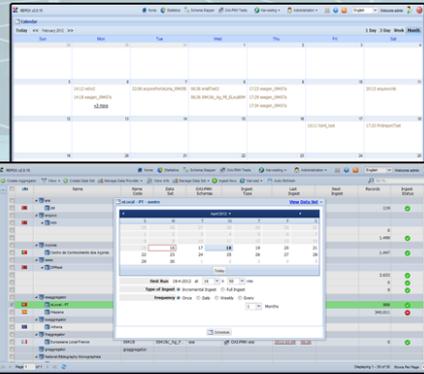
- Documentation
- Data formats
- Download
- Known Issues
- Contacts
- REPOX installations

REPOX v2

Now with a new look and new features!







Download Now!

REPOX - Data Aggregation and Interoperability Manager

REPOX is a framework to manage metadata spaces. It comprises several channels to import metadata from data providers, services to transform metadata between schemas according to user's specified rules, and services to expose the results to the exterior. This tailored version of REPOX aims to provide to all the TEL and European partners a simple solution to import, convert and expose their bibliographic data via OAI-PMH, by the following means:

- Cross platform
It is developed in Java, so it can be deployed in any operating system that has an available Java virtual machine.
- Easy deployment
It is available with an easy installer, which includes all the required software.
- Support for several metadata formats and encodings
It supports UNIMARC and MARC21 schemas, and encodings in ISO 2709 (including several variants), MarcXchange or MARCXML. During the course of the TELplus project, support will be added for other possible encodings required by the partners.
- Metadata crosswalks
It offers crosswalks for converting UNIMARC and MARC21 records to simple Dublin Core as also to TEL-AP (TEL Application Profile). A simple user interface makes it possible to customize these crosswalks, and create new ones for other formats.








Figura 6: Portal de disponibilização e acesso livre à plataforma REPOX

O objectivo destas recolhas não é o de substituir o acesso a esses conteúdos nos repositórios originais (o acesso aos conteúdos nos repositórios originais deverá ser sempre o modelo de funcionamento por omissão) mas o de extrair desses conteúdos o seu texto, através de várias técnicas possíveis, para o indexar e assim melhor os serviços de pesquisa e navegação.

CONTEXTO DE DESENVOLVIMENTO – PROJECTO EuDML

O projeto EuDML – European Digital Mathematics Library (<http://project.eudml.org/>- Figura 5), iniciou-se a 1 de Fevereiro de 2010, com o objectivo de desenvolver a infra-estrutura para a Biblioteca Digital Europeia de Matemática, promovida por vários editores e sociedades europeias de Matemática. A coordenação

geral deste projecto cabe ao IST, e a coordenação científica à Universidade de Grenoble (França). Tecnicamente, o projecto está ainda alinhado com os requisitos da The European Library e da Europeana.

Aproveitando o contexto deste projecto, pretende-se ainda vir a promover, em colaboração com a Sociedade Portuguesa de Matemática e outras entidades do meio, uma iniciativa nacional para a PtDML – Biblioteca Nacional de Matemática.

Neste projecto a plataforma REPOX é mais uma vez a tecnologia utilizada para recolher os dados dos fornecedores, provenientes de vários países europeus. Neste caso a tecnologia REPOX é utilizada como fornecedor de serviços, já que é usada num serviço final. Esse serviço utiliza os dados recolhidos para oferecer um portal de pesquisa a bibliografia de

Matemática existente na Europa em acesso livre, que no momento da escrita desta comunicação já ultrapassou os 200.000 títulos.

No final da fase de projecto, o serviço permanente será mantido pelo FIZ Karlsruhe como um complemento ao Zentralblatt MATH (<http://www.zentralblatt-math.org/>), um serviço já clássico na comunidade internacional.

No contexto deste projecto a plataforma REPOX foi enriquecida com capacidades para recolha de conteúdos de texto para além dos metadados. Assim, em cada registo de metadados recolhido é detetado o endereço da publicação descrita, sendo esta depois recolhida e posta à disposição do serviço de pesquisa EuDML. A publicação é então depois processada para recolha do seu texto integral (incluindo fórmulas matemáticas), o que permitirá enriquecer o serviço de pesquisa. Como a publicação permanece propriedade da entidade que fornece os dados, que apenas autoriza o serviço EuDML

a utilizar o texto para pesquisa, e não para acesso, quando um utilizador no portal EuDML pretende aceder a esse conteúdo a hiperligação que é oferecida é a original, do fornecedor de dados.

PROCESSOS REPOX – PROCESSOS DE RECOLHA

Na plataforma REPOX os processos de recolha de dados podem ser convenientemente programados e monitorizados, de acordo com várias opções. É suportada a recolha e publicação de dados estruturados segundo a norma ISO 2709 ou qualquer esquema XML de dados, com suporte nativo embutido para esquemas usados com frequência, como MarcXchange, Dublin Core e os perfis de metadados da Europeana, TEL e EuDML.

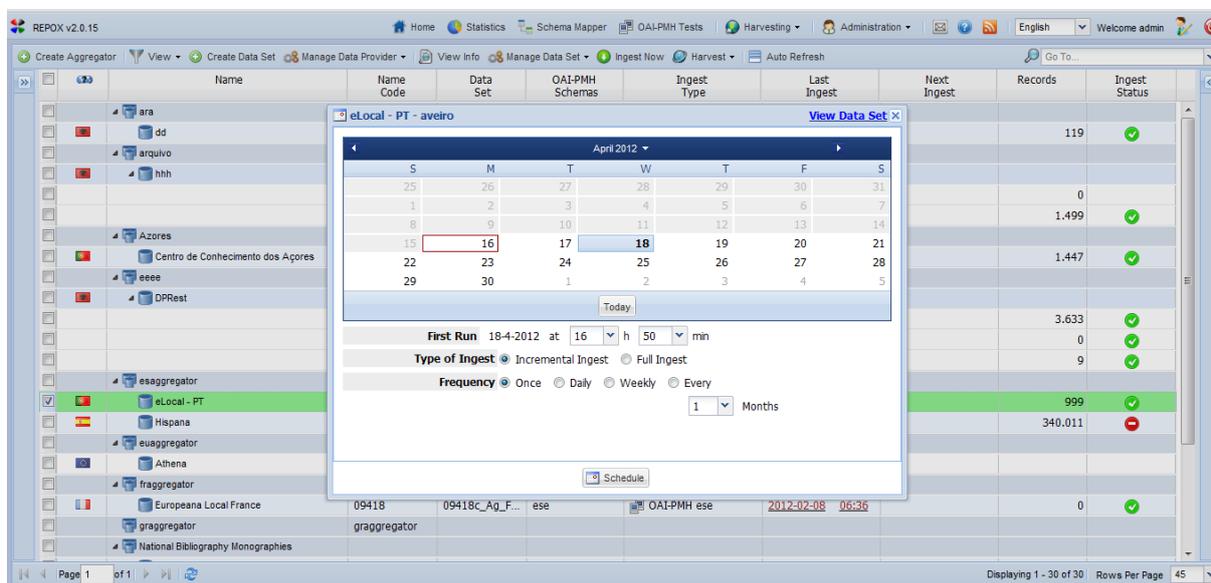


Figura 7: Calendarização de recolhas (exemplo do projeto Europeana Local)

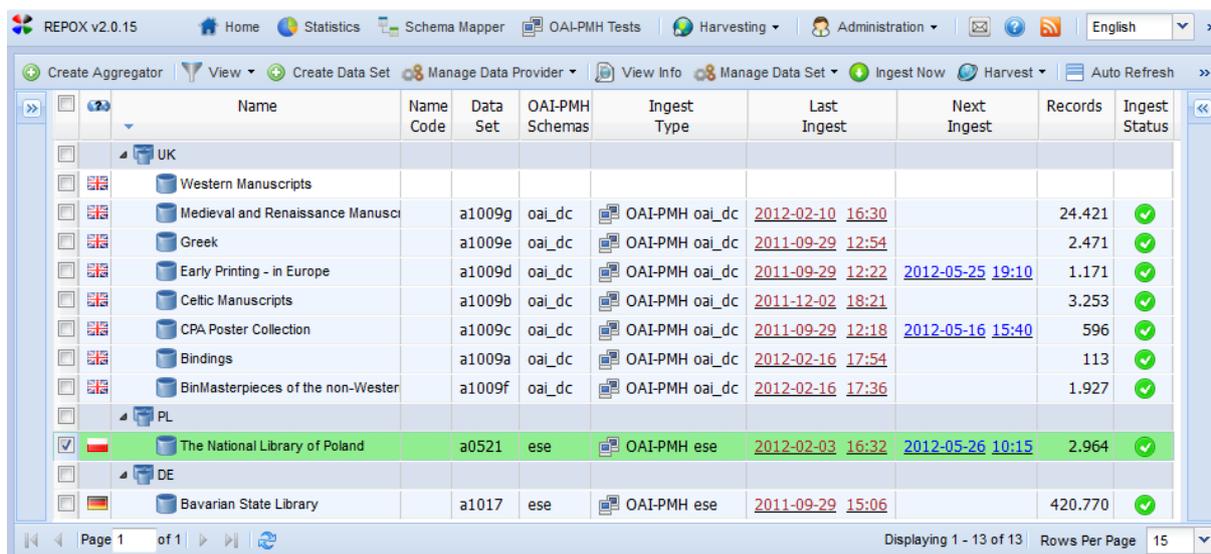


Figura 8: Detalhe de um agregador sectorial/nacional (exemplo do serviço TEL – projeto Europeana Libraries)

Para otimizar a sincronização dos conjuntos de dados e apoiar processos complexos do lado do fornecedor de serviços ou do agregador, o REPOX pode lidar com identificadores de registro de duas formas: dados extraídos do registro, usando uma expressão XPath, ou identificadores locais criado pelo REPOX (nem todos os fornecedores de dados expõem os seus dados com identificadores persistentes, logo o REPOX fornece suporte para lidar com isso implementando técnicas que tentam gerar identificadores persistentes com base na estrutura e conteúdo do registro, mas robusto o suficiente para ser usado nas suas várias versões).

As tarefas de recolha podem ser executadas imediatamente ou agendadas para ocorrerem regularmente, como por exemplo uma vez por dia, semana ou mês (Figura 7).

Para além, do uso tradicional de um servidor OAI-PMH A ferramenta suporta várias opções para os processos de recolha que se detectaram como sendo relevantes para agregadores: a partir de uma pasta local que contém os registos; por HTTP ou FTP; e até mesmo através de um processo integrado para recolha de servidores Z39.50.

PROCESSOS REPOX – SUPORTE PARA FORNECEDOR DE SERVIÇOS E PARA AGREGADOR

Quando usado como um agregador (Figura 8), a plataforma REPOX pode gerir todos os fornecedores de dados (“Data Providers”) e respetivas fontes de dados (“Data Sets”), bem como as suas informações mais relevantes: localização geográfica (país – Figura 9), nome, esquemas disponíveis dos dados, método de ingestão dos dados, últimos dados ingeridos, ingestões programadas, estatísticas e estado do sistema (na Figura 10 ilustra-se um cenário de recolha de conteúdos, já referido na descrição do projecto EuDML, que pode ser desencadeado após a recolha dos metadados), etc.

A plataforma oferece ainda opções para gerir processos de consolidação de dados e transformação, e para republicação de dados.

Para os fornecedores de serviços, a plataforma REPOX pode ser usada para recolher e gerir múltiplas fontes de dados. Além de recolher os registos provenientes de fontes externas, outros serviços adicionais estão disponíveis: estatísticas (para criação de perfis simples dos conjuntos de dados ingeridos), técnicas básicas de “business intelligence” (para criação de perfis de dados personalizados), exportação de dados (onde também o serviço de transformação de dados pode ser usado) e motor de busca (um sistema REPOX pode ser instalado em conjunto com um motor de busca como por exemplo o Solr – <http://lucene.apache.org/solr/> – oferecendo desta forma um serviço de busca e recuperação do conjuntos de dados recolhidos).

Toda a gestão do sistema é feita “on-line”. A página inicial da plataforma REPOX torna possível a gestão de todos os fornecedores de dados e fontes de dados, de acordo com os seus atributos mais relevantes: país, nome, esquemas disponíveis, método de ingestão, últimos ingerir, programadas ingerir e número de registos (Fig. 8 e 9). Nesses cenários, é possível ver os relatórios de recolha, agendar as recolhas de metadados para um tempo específico (diário, semanal ou mensal) e até mesmo a exportação dos registos recolhidos para o sistema de ficheiros dentro de um arquivo ZIP – por omissão os registos são armazenados na base de dados do REPOX, logo uma cópia de segurança pode ser útil.

É também possível aplicar uma variedade de filtros sobre a grelha principal, construídos através de um conjunto de atributos pertencentes aos fornecedores de dados e fontes de dados (Figura 10). Desta forma a procura de novas entidades as quais desconhecemos o nome torna-se mais célere. Para quando sabemos identificar o que procuramos é possível usar simplesmente a barra de procura.

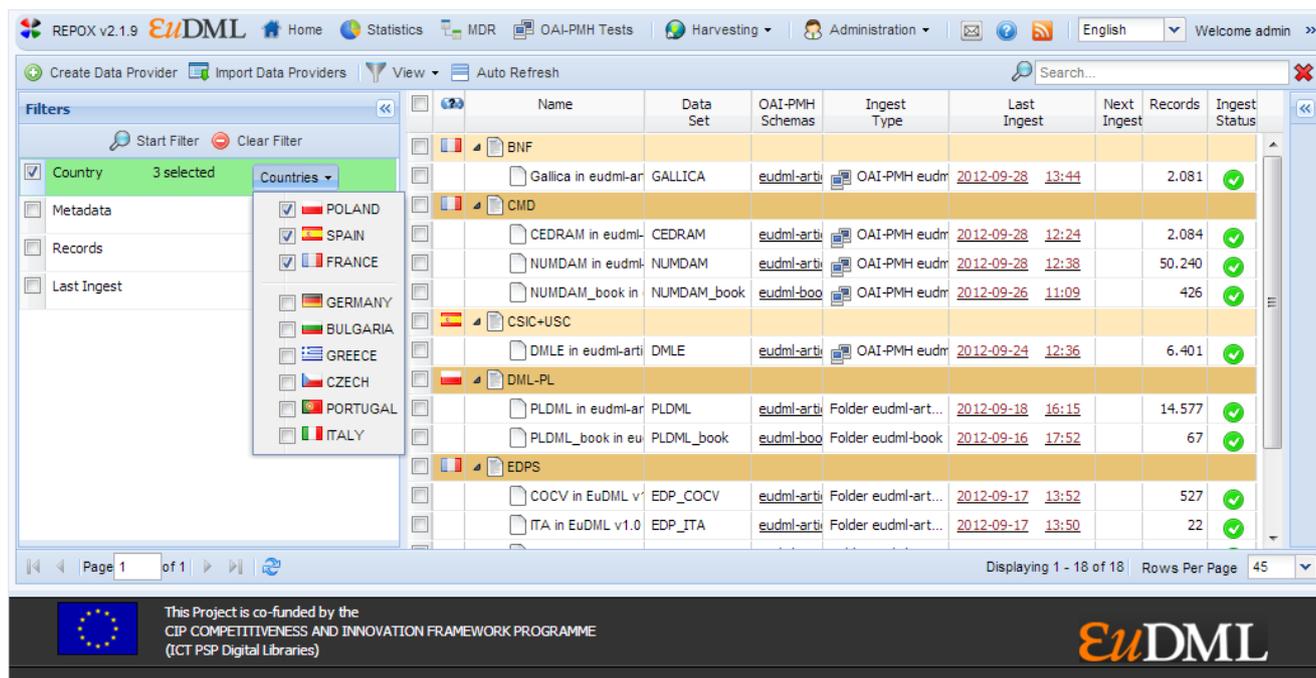


Figura 9: Filtragem da visualização de fornecedores (exemplo do serviço EuDML)

Full-Text Harvester v0.9.6.13 Saved Harvests Harvest Full-Text Harvest Result Administration

Saved Harvests

Refresh Empty Harvest Edit Harvest Remove Harvest

ID	Extract Txt?	Start Harvest	End Harvest	Records	Harvest Status
DML_CZ_Mono...	<input type="checkbox"/>	2012-09-16 21:02:30	2012-09-16 21:21:30	1.782	⊖
CEDRAM	<input type="checkbox"/>	2012-09-19 09:30:11	2012-09-19 10:06:09	1.000	⊖
PLDML_book	<input type="checkbox"/>	2012-09-17 14:03:46	2012-09-17 14:20:50	67	⊖
GDZ_Band	<input type="checkbox"/>	2012-09-19 11:55:17		747	15%
GDZ_RusDML	<input checked="" type="checkbox"/>	2012-09-19 11:55:30		8.283	3%
GDZ_Monographs	<input type="checkbox"/>	2012-09-19 11:55:43		1.549	23%
GDZ_Mathemati...	<input type="checkbox"/>	2012-09-19 11:55:52		702	1%
PMath	<input type="checkbox"/>	2012-09-17 23:14:22	2012-09-18 03:59:40	1.347	✓
HDML_conferen...	<input type="checkbox"/>	2012-09-17 11:25:37	2012-09-17 13:13:15	932	✓
NUMDAM_book	<input type="checkbox"/>	2012-09-16 20:46:09	2012-09-16 22:18:53	426	✓
HDML_journals	<input type="checkbox"/>	2012-09-17 11:28:38	2012-09-17 14:57:29	2.340	✓
BUMI	<input type="checkbox"/>	2012-09-17 11:07:43	2012-09-17 12:01:10	1.868	✓
PLDML	<input type="checkbox"/>	2012-09-18 16:27:51	2012-09-19 03:53:21	14.577	✓
DML_CZ_Proce...	<input type="checkbox"/>	2012-09-19 09:38:22	2012-09-19 10:54:57	2.030	✓
BuDML	<input type="checkbox"/>	2012-09-17 12:35:54	2012-09-17 13:14:27	591	✓
DML_CZ_Serial	<input type="checkbox"/>	2012-09-16 21:19:40	2012-09-16 22:56:43	2.057	✓

Page 1 of 1 Displaying 1 - 16 of 16

Harvest Details: GDZ_RusDML

⊖ Stop Harvest

Last Ingest Information

Start Harvest: 2012-09-19 11:55:30

End Harvest:

Duration: 04:09:57

Records: 16.748

Harvest Status: RUNNING

File Number: 213

Size: 22.3 MB

Failed Records: 415

General Information

Id: GDZ_RusDML

OaiUrl: http://bd2.inesc-id.pt:8080/repoX2Eudml/OAIHandler

OaiSet: GDZ_RusDML

Metadata Prefix: eudml-article

Xpath: /article:article/article:front/article:article-meta/article:ext-link

Harvested Records: 8.283 of 16.748

Total File Number: 213

Total Harvest Size: 25.5 MB

Figura 10: Recolha de conteúdos (exemplo do serviço EuDML)

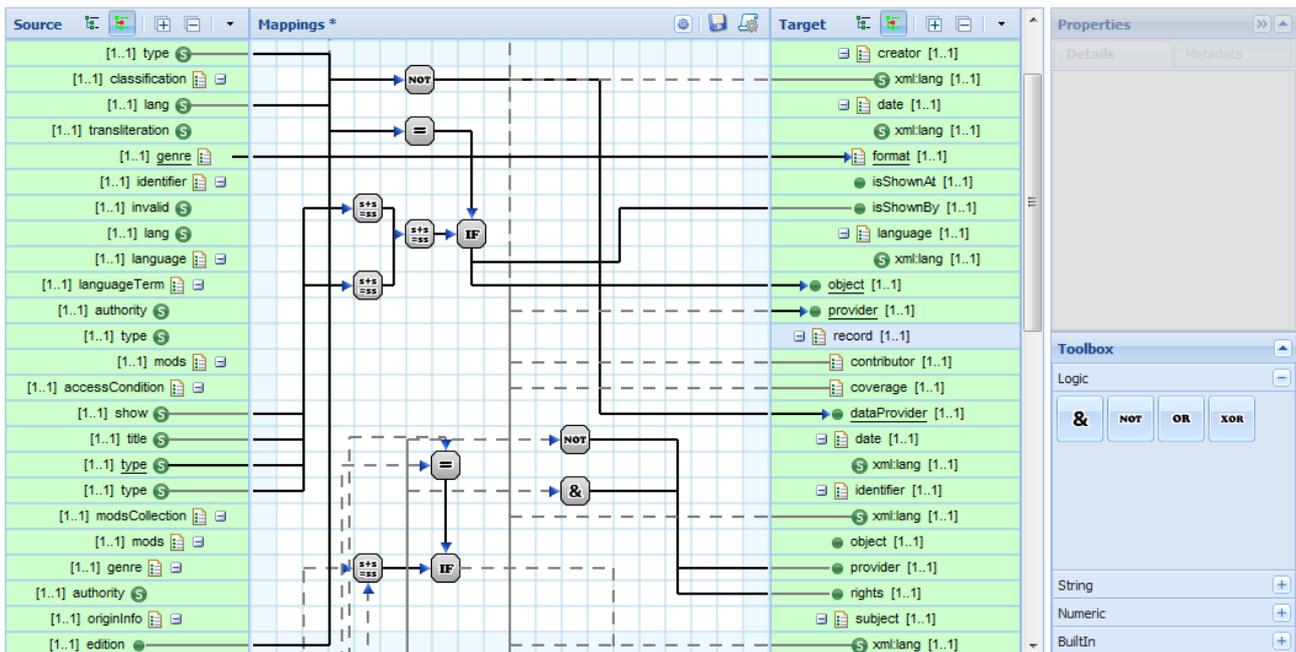


Figura 11: Módulo de suporte ao mapeamento entre diferentes esquemas de metadados

Description	Source Format	Destination Format	Schema	Metadata Namespace	Used	Download
OAI to ESE	oai_dc	ese	http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd	http://www.europeana.eu/schemas/ese/	0	
OAI Arquivo Municipal de Lisboa 2 ...	oai_dc	ese	http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd	http://www.europeana.eu/schemas/ese/	0	
oai_dc to tel	oai_dc	tel	http://krait.kb.nl/coop/tel/handbook/terterms.html	http://krait.kb.nl/coop/tel/handbook/terterms.html	0	
oaiFpc2ese	oai_dc	ese	http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd	http://www.europeana.eu/schemas/ese/	0	
cmAveiroTo ESE	aveiroxml	ese	http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd	http://www.europeana.eu/schemas/ese/	0	
museumxml2ese	museumxml	ese	http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd	http://www.europeana.eu/schemas/ese/	0	
MARC21 BGUC to ESE	MarcXchange	ese	http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd	http://www.europeana.eu/schemas/ese/	0	
MARC21 to oai_dc	MarcXchange	oai_dc	http://www.openarchives.org/OAI/2.0/oai_dc.xsd	http://www.openarchives.org/OAI/2.0/	0	
MARC21 to tel	MarcXchange	tel	http://krait.kb.nl/coop/tel/handbook/terterms.html	http://krait.kb.nl/coop/tel/handbook/terterms.html	0	
MARC21 to Qualified Dublin Core	MarcXchange	qdc			0	
Unimarc To ESE	MarcXchange	ese	http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd	http://www.europeana.eu/schemas/ese/	0	
UNIMARC BibliotecaDigitalDoAlente...	MarcXchange	ese	http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd	http://www.europeana.eu/schemas/ese/	0	
UNIMARC CasaFfessoa 2 ESE	MarcXchange	ese	http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd	http://www.europeana.eu/schemas/ese/	0	
UNIMARC Figueiro dos Vinhos 2 ESE	MarcXchange	ese	http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd	http://www.europeana.eu/schemas/ese/	0	
UNIMARC Hemeroteca 2 ESE	MarcXchange	ese	http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd	http://www.europeana.eu/schemas/ese/	0	
UNIMARC MTCComunidade 2 ESE	MarcXchange	ese	http://www.europeana.eu/schemas/ese/ESE-V3.4.xsd	http://www.europeana.eu/schemas/ese/	0	

Figura 12: Esquemas e mapeamentos disponíveis no serviço EuDML

PROCESSOS REPOX – TRANSFORMAÇÃO DE DADOS

A plataforma REPOX permite a gestão tanto dos esquemas que definem os registos das fontes de dados no seu repositório, como os mapeamentos dos dados do seu formato original para outro, para posteriormente disponibilizar ambos por OAI-PMH (Figura 11).

Estes mapeamentos podem ser criados através de uma interface visual prática que simplifica a criação da correspondência entre os atributos de dois esquemas diferentes (Figura 12).

CONCLUSÕES

A plataforma REPOX começou por ser idealizada para a Biblioteca Nacional Digital e para a PORBASE, mas o seu verdadeiro potencial acabou por ser concretizado no projecto TELplus e noutros que se lhe seguiram.

Esta é a contribuição nacional mais relevante para os serviços TEL e Europeia, constituindo a espinha dorsal da infra-estrutura OAI-PMH dessas redes. Esta tecnologia é no entanto usada num largo número de outras entidades, especialmente bibliotecas da rede TEL e outras da rede Europeia.

Em terminologia da “Open Archives Initiative”, esta ferramenta pode ser assim usada tanto por “data providers” como por “services providers”. Em terminologia dos domínios TEL e Europeia, pode-se dizer que a plataforma pode ser ainda usada para serviços agregadores.

A plataforma REPOX está também especialmente preparada para colecções de metadados seguindo o recente Europeia Data Model (EDM) que sucede ao Europeia Semantic Elements (ESE), visando melhorar a preservação da riqueza dos dados (que muitas vezes existe nos formatos originais como os formatos MARC, EAD, etc.) a quando da troca com a Europeia, sem

prejudicar a interoperabilidade. A plataforma tem assim suporte especializado para recolher colecções de metadados já codificados em EDM, transformar metadados de esquemas originais para EDM, e disponibilizar metadados em EDM (por OAI-PMH ou vários outros meios).

REFERÊNCIAS

FREIRE, Nuno; MAGUINHAS, Hugo; BORBINHA, José – Metadata Spaces: The Concept and a Case with REPOX – S. Sugimoto et al. (Eds.). ICADL 2006, LNCS 4312, p. 293-302, 2006. Springer Berlin / Heidelberg 2006, p. 293-302. ISBN 978-3-540-49375-4

AGRADECIMENTOS

Este trabalho foi financiado por fundos nacionais através da FCT – Fundação para a Ciência e a Tecnologia, no contexto do projeto plurianual PEst-OE/EEI/LA0021/2011.