

# RETRIEVO 2

## **AGREGAÇÃO DE CONTEÚDOS E PESQUISA FEDERADA**

CARACTERÍSTICAS E REQUISITOS  
TÉCNICOS

### **SOBRE O DOCUMENTO**

Identificador	WP181108		
Aprovado por	Luís Miguel Ferros	<b>Aprovado em</b>	2018-06-19
Classificação	Público		
Distribuição	N/A		

### **REVISÕES**

#	Data	Autor	Alterações
1	2018-05-11	Luís Miguel Ferros, Miguel Ferreira	Documento inicial
2	2018-06-19	Luís Miguel Ferros	Revisão final

## **SUMÁRIO EXECUTIVO**

O Retrieve é um software que permite implementar portais de pesquisa e recuperação de informação sobre qualquer fonte de informação existente numa organização. Por exemplo, este poderá ser usado para pesquisar simultaneamente no catálogo da biblioteca, no sistema de gestão documental e no sítio Web da instituição através de uma única interface de pesquisa.

Este documento tem como objetivo descrever as principais características, funcionalidades e vantagens do software Retrieve, bem como os requisitos técnicos necessários à sua correta instalação.

**PONTO DE  
ACESSO ÚNICO  
A TODOS OS  
RECURSOS DE  
INFORMAÇÃO  
DA SUA  
ORGANIZAÇÃO**

## **RETRIEVO**

O Retrievio é um software que permite implementar portais de pesquisa e recuperação de informação sobre qualquer fonte de informação existente numa organização. Por exemplo, este poderá ser usado para pesquisar simultaneamente no catálogo da biblioteca, no sistema de gestão documental e no sítio Web da instituição através de uma única interface de pesquisa.

O Retrievio não tem como objetivo substituir as fontes de informação existentes, mas sim fornecer um ponto de acesso único e privilegiado à informação que se encontra distribuída pelos diferentes quadrantes da organização.

Após localizar a informação no Retrievio, o utilizador será redirecionado para o sistema original que contém a informação para esta ser apresentada em contexto e de forma completa.

Para além de permitir a pesquisa centralizada de informação, este software atua também como fornecedor de informação para outros sistemas, podendo ser integrado com agregadores internacionais de informação como a Europeia, o Portal Europeu de Arquivos, o Portal OpenAire ou serviços de pesquisa federada como a b-on.

### Informação acessível a partir de um **único ponto de acesso**

O Retrieve permite localizar informação que se encontra dispersa por vários sistemas de informação, mesmo que à partida estes sejam incompatíveis.

A pesquisa é sempre realizada através de uma interface gráfica comum, evitando a necessidade do utilizador ter de aceder a cada um dos sistemas individualmente.

### Aumento da **visibilidade** da instituição

Através do Retrieve, uma determinada instituição poderá disponibilizar na Internet informação de vários sistemas internos através de um único portal.

Além de consolidar num único local a informação de vários sistemas, fornece aos seus utilizadores mecanismos que facilitam a pesquisa e recuperação de informação.

### **Uniformização** de sistemas heterogéneos

O Retrieve é compatível com vários protocolos de comunicação normalizados (e.g. OAI-PMH, Z39.50, SRU), bem como serviços SOAP, conectores SQL e um conjunto alargado de gateways de acesso a bases de dados de publicações científicas (e.g. EBSCO, ABI, ProQuest, etc.). Independentemente dos protocolos implementados ou do sistema que suporta a informação, o Retrieve será sempre capaz de a indexar e apresentar convenientemente ao utilizador final.

### Ativos de **informação sob controlo**

O Retrieve disponibiliza um conjunto de estatísticas e relatórios de gestão que refletem o estado de operação do sistema.

Permite, assim, monitorizar cada fonte de informação através de estatísticas como distribuição do número de registos por sistema recolhido, registos e fontes de informação mais consultados, crescimento mensal, relatórios de agregação, estatísticas de acesso, entre outros.

## **Qualidade da informação assegurada**

Para garantir a qualidade da informação recolhida, o Retrieve dispõe de um motor de verificação de conformidade que lhe permitirá validar a informação recolhida de acordo com os critérios estabelecidos. Os critérios de qualidade são totalmente configuráveis pelo administrador do sistema e adaptáveis a qualquer situação ou contexto organizacional.

Assim, o portal apenas expõe os registos de informação que apresentam um nível de qualidade mínimo estabelecido.

## ARQUITETURA E MÓDULOS APLICACIONAIS

O software Retrievio é composto por 3 módulos aplicativos, os quais estão representados na figura seguinte.

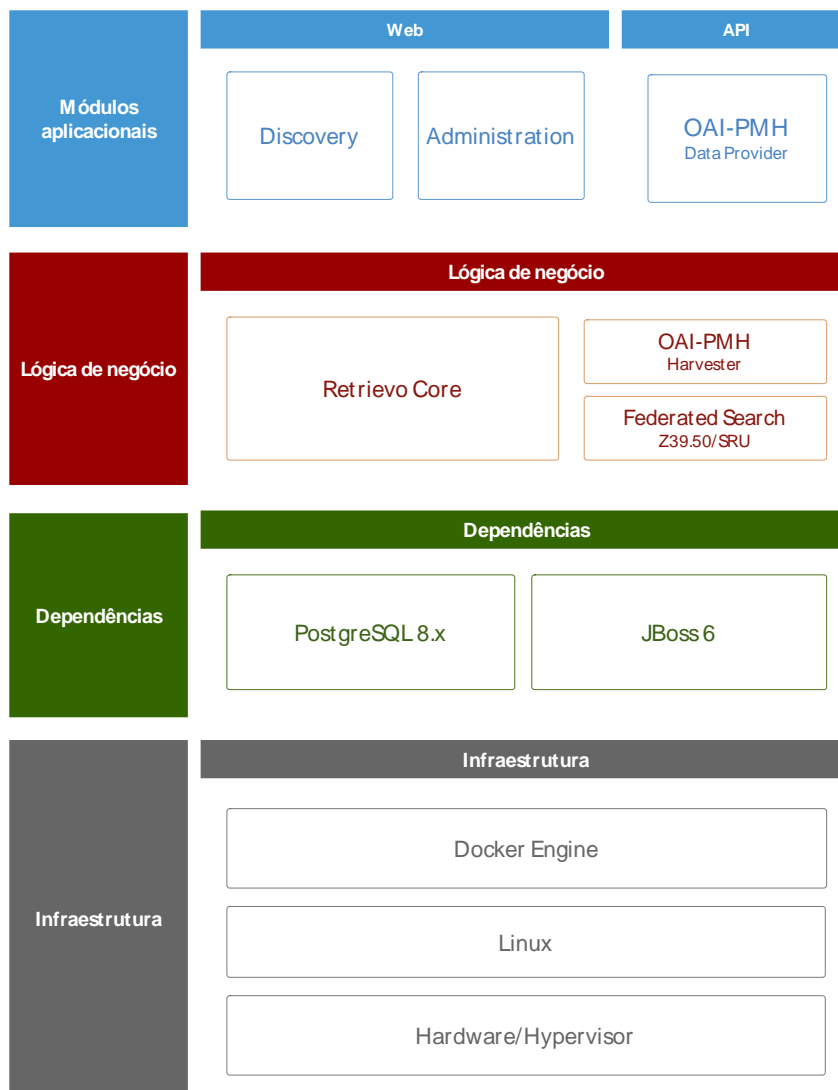


Figura 1 - Arquitetura do Retrievio.

## DISCOVERY

O módulo Discovery fornece ao utilizador um ponto de acesso único de pesquisa para recuperação de informação originária de diferentes fontes de informação. O principal objetivo é facilitar o acesso à informação, uma vez que fornece, sobre uma única interface, capacidades de descoberta e recuperação de informação proveniente de fontes de dados heterogéneas.

O utilizador poderá realizar pesquisas transversais em todos os registos recolhidos (pesquisa simples) ou realizar pesquisas mais específicas em determinados campos de informação, sobre determinadas fontes de informação, em conjugação com outros filtros (pesquisa avançada). Pode, por exemplo, pesquisar sobre todos os campos de todos os registos de todos os repositórios, ou simplesmente pesquisar no Título dos repositórios A e B.

Os resultados da pesquisa são consolidados e indexados por título, autor, assunto, datas, ou outros atributos previamente configurados.

Por norma, os resultados de pesquisa apresentam apenas um subconjunto da metainformação do registo original, a qual deverá ser suficiente para identificar a informação que se procura. Para auxiliar a identificação dessa informação, também é recolhida e apresentada, quando existente, uma miniatura do documento associado.

O utilizador, após localizar a informação que pretende, poderá consultar o registo original, sendo encaminhado para a fonte de dados que forneceu esse registo, no qual será apresentada a informação completa (metainformação e ficheiros associados), posicionado no contexto certo (e.g. árvore de classificação) e a partir do qual poderão ser desencadeados serviços adicionais (e.g. empréstimo de documentos).

O design gráfico deste módulo pode ser adaptado para ir ao encontro da identidade visual do cliente.

### Respostas num piscar de olhos

O aumento do número de registos na instância central do Retrieve, resultado da constante incorporação de novas fontes de informação, obrigou a que fossem criadas novas soluções de engenharia capazes de suportar eficazmente grandes volumes de informação.

Com capacidade para pesquisar em dezenas de milhões de registos sem qualquer diminuição do desempenho, o Retrieve é o produto certo para lidar com a recuperação de informação em ambientes que reúnem elevados volumes de registos de informação.

## ADMINISTRATION

O Retrievio é acompanhado de um módulo de administração que permite ao utilizador configurar e gerir todo o sistema. Entre as operações disponíveis, é possível listar, modificar e adicionar novas fontes de informação para agregação e disponibilização nas interfaces de pesquisa, bem como configurar os seus parâmetros de agregação.

Permite também consultar estatísticas de agregação, relatórios de problemas ocorridos durante a agregação, estatísticas de crescimento do número de registos por fonte de dados, bem como estatísticas de consulta, entre outros indicadores.

### Compatível com qualquer plataforma

O Retrievio é uma solução 100% Web. Para utilizar o sistema, apenas necessita de um browser. O software está centralizado num servidor, não havendo a necessidade de instalação de aplicações nos postos de trabalho dos utilizadores. O acesso ao sistema poderá ser feito em qualquer parte e a partir de qualquer dispositivo, *desktop* ou *tablet*.

## OAI-PMH DATA PROVIDER

O Retrievio recolhe e concentra informação de várias fontes de informação. Esta informação é validada, normalizada e devidamente indexada para suportar as pesquisas realizadas através do Módulo Discovery.

Uma vez reunida informação anteriormente dispersa, o Retrievio pode fornecer informação a outros portais agregadores de conteúdos (e.g. Europeana, Portal Europeu de Arquivos, OpenAire, etc.), recorrendo ao módulo OAI-PMH Data Provider.

Este módulo permite, assim, a criação de um canal único de fornecimento de informação a outros portais, evitando a necessidade de estabelecer adesões a título individual para cada uma das fontes de dados associadas ao Retrievio.

Este módulo disponibiliza a informação através do protocolo OAI-PMH ou através de ficheiros para envio através de mecanismos assíncronos (e.g. FTP). Os formatos de metadados suportados são o DC (Dublin Core) e o apeEAD (formato de dados promovido pelo Portal Europeu de Arquivos).

### Levar a informação ainda mais longe

O Retrievio poderá fornecer a sua informação a outros portais agregadores de conteúdos recorrendo ao seu módulo OAI-PMH Data Provider. O Retrievio funcionará, assim, como o canal único de fornecimento de informação, facilitando o processo de adesão a plataformas internacionais de agregação de conteúdos (e.g. Europeana, Portal Europeu de Arquivos).



## ACESSIBILIDADE DE CONTEÚDOS WEB

As Diretrizes de Acessibilidade para Conteúdo Web (WCAG) 2.0 são um conjunto de recomendações emanadas pela W3C que visam tornar o conteúdo Web mais acessível. O cumprimento destas diretrizes faz com que o conteúdo publicado na Web se torne acessível a pessoas com incapacidade, nomeadamente cegueira e baixa visão, surdez e baixa audição, dificuldades de aprendizagem, limitações cognitivas, limitações de movimentos, incapacidade de fala, fotossensibilidade, entre outros. Seguir estas diretrizes permite também que o conteúdo Web se torne mais usável por utilizadores em geral e por dispositivos móveis como *smartphones*, *tablets* ou relógios de pulso.

Dada a importância desta temática, foi criada legislação que visa promover a adoção destas diretrizes por todo o Estado português. O Regulamento Nacional de Interoperabilidade Digital (RCM n.º 91/2012 de 8 de novembro) esclarece a obrigatoriedade do cumprimento das diretrizes WCAG 2.0 nível AA por todos os sítios Web do Estado que disponibilizem serviços online<sup>1</sup>. O artigo 2º da lei n.º 36/2011 determina que a lei se aplica a:

- Órgãos de soberania;
- Serviços da administração pública central, incluindo institutos públicos e serviços desconcentrados do Estado;
- Serviços da administração pública regional;
- Sector empresarial do Estado.

A KEEP SOLUTIONS é solidária com esta iniciativa e assegura que todos os seus produtos estão em plena conformidade com o nível AA+ das Diretrizes de Acessibilidade para Conteúdo Web (WCAG) 2.0.

### Ideal para redes de repositórios

O Retrievio é perfeito para suportar redes de repositórios de informação num contexto local, nacional ou internacional.

No âmbito local pode ser utilizado, por exemplo, para concentrar e dar acesso a informação de carácter cultural e/ou patrimonial produzida por instituições de um município.

No âmbito nacional ou internacional pode ser utilizado para concentrar informação de diferentes entidades. Exemplos deste tipo de arquitetura são o Repositório Científico de Acesso Aberto de Portugal e o Portal Português de Arquivos.

---

<sup>1</sup> Ver também Lei n.º 36/2011, de 21 de junho. Estabelece a adoção de normas abertas nos sistemas informáticos do Estado (PDF, 188KB); RCM n.º 91/2012 de 8 de novembro. Regulamento Nacional de Interoperabilidade Digital (PDF, 16,1MB); DGAEP. (Fevereiro de 2013). Estrutura orgânica da AP Portuguesa pós-PREMAC (PDF, 1,8MB); DGTF. (30/09/2012). Definição de Sector Empresarial do Estado (PDF, 102KB).

## REQUISITOS TÉCNICOS

O Retrievio pode ser instalado em qualquer servidor Linux com características de *hardware* convencionais.

Dado que toda a aplicação funciona em ambiente Web, os postos de trabalho dos utilizadores apenas necessitam de um browser para aceder à aplicação (e.g., Internet Explorer, Google Chrome, Mozilla Firefox ou outro).

A tabela seguinte ilustra as características técnicas de um servidor capaz de suportar uma instância do Retrievio com 8 milhões de registos.

### SERVIDOR

RAM	8 GB 16 GB recomendados
CPU	Intel Quad-Core ou superior
Disco	100 GB Depende do número de registos bibliográficos, imagens associadas e da sua taxa de crescimento
Sistema operativo	Ubuntu 16.04 LTS ou compatível Sem custos de licenciamento
Software	Docker engine Sem custos de licenciamento
Rede	100 Mbit/s ou superior 1 Gbit/s é recomendável em contexto de incorporação de grandes volumes de representações digitais

### POSTO DE TRABALHO

RAM	4 GB
CPU	Intel Dual-Core ou superior
Monitor	1280x768 pixéis ou superior
Sistema operativo	Windows/Linux/macOS
Software	Web browser
Rede	100 Mbit/s ou superior 1 Gbit/s é recomendável em contexto de incorporação de grandes volumes de representações digitais

 [www.keep.pt](http://www.keep.pt)  
 +351 253 066 735  
 [info@keep.pt](mailto:info@keep.pt)  
 [sales@keep.pt](mailto:sales@keep.pt)  
 KEEP SOLUTIONS, LDA.  
Rua Rosalvo de Almeida, n° 5,  
4710-429 Braga  
Portugal

## KEEP SOLUTIONS

A KEEP SOLUTIONS é uma empresa que tem como missão o fornecimento de soluções avançadas para gestão e preservação de informação.

A nossa abordagem assenta no fornecimento de sistemas de informação e na prestação de serviços com vista a permitir aos nossos clientes realizar uma gestão mais eficiente dos seus ativos de informação.

A empresa iniciou a sua atividade em 2008, tendo adquirido o estatuto de spin-off académica da Universidade do Minho, por se tratar de uma iniciativa empresarial com fortes laços de cooperação com centros de investigação e departamentos desta instituição.

Os nossos clientes encontram-se principalmente no setor público, nos domínios cultural, educacional, patrimonial e científico, mais concretamente nas áreas de arquivo, biblioteca e museu.

Apostamos no desenvolvimento contínuo de soluções inovadoras. Para isso, permanecemos ativos na produção de conhecimento científico, participando ativamente em projetos de I&D em cooperação com instituições nacionais e internacionais.