



Repositório de Objectos Digitais Autênticos

# RODA

Repositório de Objectos Digitais Autênticos

---

Descrição do sistema

Direcção-Geral de Arquivos

2009

**Ficha técnica**

Título: Descrição do sistema

Id.

Classificação: 010.01.03 DSIPE\000078

Versão: V 0.5

Produtor: Direcção-Geral de Arquivos

Idioma: Português

Descritores: preservação digital, património arquivístico, RODA, políticas

Autor: Miguel Ferreira

Colaborador: Cecília Henriques

Colaborador: Francisco Barbedo

Colaborador: Glória Santos

Colaborador: José Carlos Ramalho

Colaborador: Luís Faria

Colaborador: Luís Corujo

Colaborador: Rui Castro

Data/Hora: 2009-3-3

Tipo de recurso: Documento de arquivo (projectos internos)

Formato de dados: PDF

Estatuto de utilização: Acesso público

© DGARQ, 2009

## Sumário

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>4</b>
<b>2</b>	<b>ARQUITECTURA E INFRA-ESTRUTURA TECNOLÓGICA.....</b>	<b>5</b>
2.1	Arquitectura geral do RODA .....	5
2.2	Infra-estrutura de hardware .....	7
2.3	Sistemas operativos .....	8
2.4	Infra-estrutura de rede .....	9
<b>3</b>	<b>SEGURANÇA .....</b>	<b>9</b>
3.1	Segurança de nível físico .....	9
3.2	Segurança da rede .....	10
3.3	Segurança dos dados .....	10
3.4	Segurança de nível social .....	11
<b>4</b>	<b>PROCESSOS.....</b>	<b>12</b>
4.1	Ingestão .....	12
4.2	Preservação .....	14
4.3	Gestão.....	19
4.4	Disseminação .....	29
<b>5</b>	<b>ORGANIZAÇÃO E ESTRUTURA DA INFORMAÇÃO .....</b>	<b>34</b>
5.1	Metainformação .....	34
5.2	Caracterização dos SIP .....	35
5.3	Caracterização dos AIP.....	36
5.4	Caracterização dos DIP .....	38
5.5	Política de identificadores.....	39
<b>6</b>	<b>REFERÊNCIAS .....</b>	<b>39</b>

# 1 Introdução

O Repositório de Objectos Digitais Autênticos (RODA) é composto por um conjunto alargado de serviços que trabalham em conjunto para armazenar, preservar e dar acesso a todo o tipo de informação digital. Trata-se, efectivamente, de uma arquitectura orientada ao serviço (SOA) que procura, sobretudo, garantir o acesso continuado à informação custodiada sem que haja prejuízo da performance e/ou da segurança dos dados.

O RODA armazena e preserva informação que poderá ser de extrema sensibilidade. Devido a este motivo, é sensato assumir que este será alvo de ataques por parte de indivíduos estranhos à organização, ataques esses que poderão ser efectuados através da rede externa (Internet), interna (i.e. Intranet) e/ou através de acesso físico às máquinas que garantem o acesso à informação recorrendo a técnicas de engenharia social ou outras. A infra-estrutura que suporta o RODA deverá ser capaz de contornar todas essas agressões oferecendo um nível elevado de segurança.

O desempenho do repositório é fundamental para garantir uma boa experiência, tanto para os profissionais de arquivo que irão gerir e alimentar o sistema numa base diária, como para os seus clientes, quer estes sejam produtores ou consumidores de informação.

A tolerância a falhas e a alta-disponibilidade são também factores importantes que foram considerados durante a fase projecção do sistema, no entanto, no âmbito deste projecto, considerou-se que estes seriam de menor importância dado o contexto organizativo em que o sistema se enquadra.

Sob uma perspectiva de total transparência, este documento descreve a arquitectura interna do RODA, a infra-estrutura tecnológica que o suporta, os mecanismos de segurança implementados, os processos internos de funcionamento e a forma como a informação é organizada e armazenada no seu interior.

Este documento não dispensa a consulta dos seguintes documentos anexos:

- RODA: Política de preservação digital
- RODA: Manual de procedimentos de disseminação
- RODA: Manual de procedimentos de ingestão
- RODA: Manual de procedimentos de administração

## 2 Arquitectura e infra-estrutura tecnológica

### 2.1 Arquitectura geral do RODA

O RODA está organizado em várias unidades funcionais de acordo com o modelo de referência Open Archival Information System [2]. A Figura 1 apresenta os macro componentes que constituem a arquitectura do RODA.

Na base da figura podemos encontrar os serviços disponibilizados pelo Fedora Commons. O Fedora Commons [3, 4] é uma plataforma de código-aberto que disponibiliza um conjunto de serviços básicos que permitem o desenvolvimento de repositórios digitais altamente parametrizáveis. Os serviços disponibilizados pelo Fedora Commons incluem a capacidade de incorporar, aceder e eliminar objectos digitais, bem como estabelecer relações entre estes através de mecanismos ontológicos baseados em RDF. Para obter uma lista completa dos serviços disponibilizados pelo Fedora Commons é favor consultar [5].

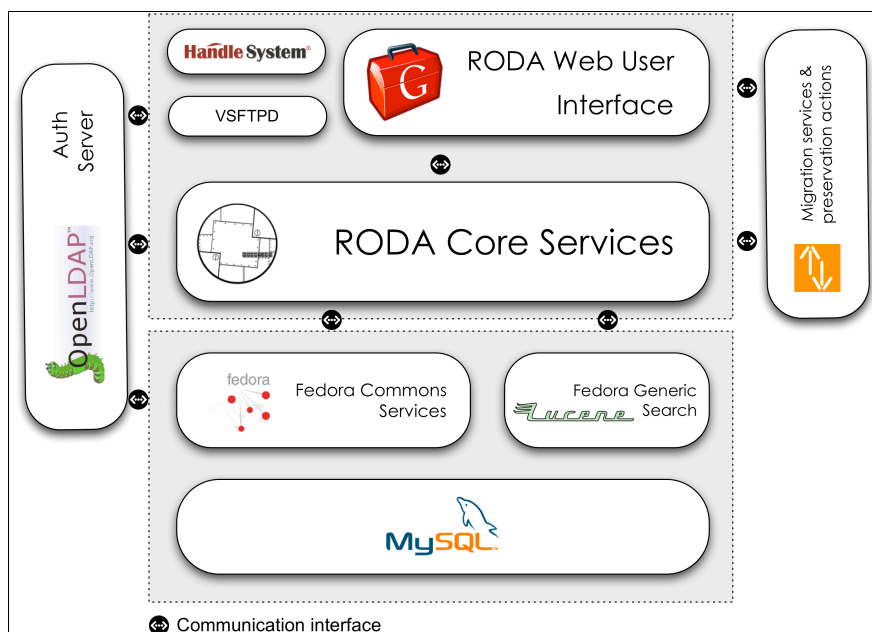
O Fedora é, também, acompanhado de um motor de pesquisa (i.e. Jakarta Lucene) capaz de indexar o conteúdo dos vários objectos armazenados permitindo assim a sua posterior localização e recuperação. O mecanismo de autenticação implementado pelo RODA assenta no recurso a um servidor LDAP (Lightweight Directory Access Protocol).

O RODA guarda um registo de todas as acções realizadas no repositório. Essa informação é armazenada numa base de dados MySQL e em ficheiros de texto armazenados no sistema de ficheiros.

Sobre os serviços anteriormente descritos foram desenvolvidos métodos remotos mais complexos que implementam a lógica de negócio do repositório (i.e. RODA Core Services). Estes métodos encontram-se acessíveis através de vários Web services organizados de acordo com o modelo de referência OAIS [2]. Uma lista dos métodos disponibilizados pelo RODA Core pode ser consultada em <https://roda.dgarq.gov.pt:8443/roda-core> <http://dev.roda.dgarq.gov.pt:8080/roda-core/services>.

Em paralelo com o RODA Core Services existe um conjunto de serviços de migração e serviços que desempenham tarefas de preservação e manutenção, tais como: normalizações de formatos, verificações de integridade, despiste de vírus, tarefas administrativas variadas, lançamento de alertas, calculo de estatísticas, etc.

A interface com o utilizador é assegurada pelo componente RODA Web User Interface (RODA WUI). Este componente faz uso dos serviços disponibilizados pelo RODA Core e apresenta uma interface gráfica que permite aos diferentes utilizadores (i.e. produtores, consumidores, arquivistas, e administradores de sistema) interagir com o repositório.



**Figura 1 - Arquitectura geral do RODA.**

Os subcomponentes que constituem RODA WUI são baseados no Google Web Toolkit e toda a comunicação entre componentes é assegurada por tecnologias AJAX e Web services.

Nome do serviço	Descrição
MySQL	Base de dados onde são guardadas as acções desenvolvidas no RODA
Fedora Commons	Serviços básicos de incorporação de objectos e criação de relações entre objectos
LDAP	Servidor de autenticação
Jakarta Lucene	Serviço de indexação de registos de metainformação
RODA Core Services	Serviços que implementam a lógica de negócio
RODA WUI	Serviços que constituem a interface com o utilizador
Handle System	Servidor de identificadores persistentes
vsFTPD	Servidor de FTP para envio de pacotes de submissão de grandes dimensões
Migration services & preservation actions	Serviços de migração e acções de preservação

**Tabela 1 - Descrição dos vários serviços que constituem o RODA.**

É importante referir que os serviços disponibilizados pelo RODA Core encontram-se acessíveis a partir do exterior da rede da DGARQ, o que possibilita a interoperabilidade com outras aplicações desenvolvidas por terceiros.

A arquitectura do RODA incorpora ainda um servidor de FTP (i.e vsftpd) para envio de pacotes de submissão cujo tamanho exceda os 2 GB de informação.

O RODA implementa ainda um serviço de identificadores persistentes baseado no Handle System [6].

A Tabela 1 resume as funções de cada um dos serviços anteriormente descritos.

## 2.2 Infra-estrutura de hardware

O repositório RODA e os seus constituintes encontram-se distribuídos por vários componentes de hardware de modo a garantir o correcto funcionamento do sistema e a segurança dos dados.

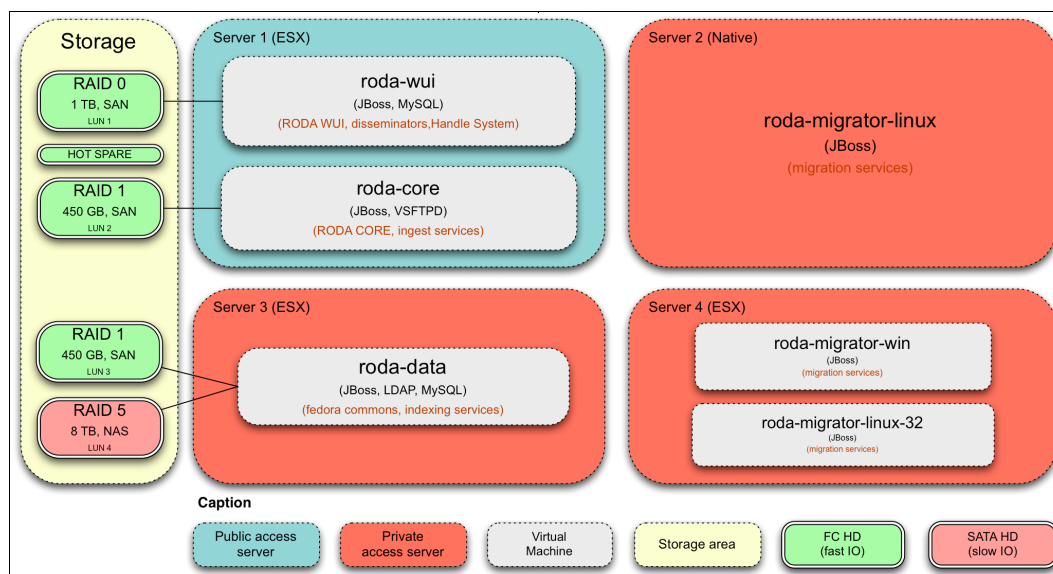
A infra-estrutura que suporta o RODA é constituída por 4 servidores reais que suportam 5 máquinas virtuais. No total, os serviços do RODA encontram-se distribuídos por 6 servidores distintos.

Na implementação do RODA recorreu-se a máquinas virtuais baseadas no sistema operativo ESX da VMWare com o objectivo de maximizar a utilização do hardware existente e aumentar a segurança do sistema. Há, no entanto, um serviço (i.e. *roda-migrator-linux*) que é executado de forma nativa, na máquina real, de modo a maximizar a performance computacional.

Os componentes RODA WUI e RODA Core Services são executados cada um no seu servidor virtual. Estes encontram-se localizados na zona desmilitarizada da rede, ou seja, os serviços implementados nestes servidores podem ser consultados a partir da Internet por qualquer utilizador autenticado (Figura 2). Esses servidores designam-se **roda-wui** e **roda-core**, respectivamente.

Noutro servidor virtual, designado **roda-data**, são executados os serviços responsáveis pelo armazenamento e localização de dados, nomeadamente, o Fedora Commons (correndo sobre o servidor aplicacional JBoss), o servidor de autenticação LDAP, o MySQL e o motor de pesquisa Lucene. Estes componentes encontram-se na zona militarizada da rede, ou seja, existe um filtro que garante que todos os acessos a este servidor provêm exclusivamente do servidor **roda-core**. Não é possível obter acesso directo a estes serviços a não ser através do **roda-core**.

O RODA é ainda suportado por três servidores adicionais onde são executados os vários serviços de migração. Esses serviços encontram-se distribuídos por várias máquinas para que haja um correcto balanceamento da carga de processamento e devido a algumas dependências tecnológicas impossíveis de contornar (e.g. certos conversores apenas podem ser executados em determinados sistemas operativos, outros carecem de uma arquitectura 32 bits). Este servidores designam-se **roda-migrator-linux** (servidor real), **roda-migrator-win** e **roda-migrator-linux-32** (ambos, máquinas virtuais).



**Figura 2 - Arquitectura de hardware do RODA.**

Os objectos depositados no repositório encontram-se armazenados num *data centre* acessível através de NAS (Network-attached storage) e/ou SAN (Storage area network).

A Tabela 2 resume a distribuição dos vários serviços que constituem o RODA pelos vários servidores existentes.

Servidor	Serviços	Servidores aplicativos
roda-wui	RODA WUI, disseminadores, Handle system	JBoss, MySQL
roda-core	RODA CORE, ingest services	JBoss, vsftpd
roda-data	Fedora commons, indexing services	JBoss, LDAP, MySQL
roda-migrator-linux	migration services	JBoss
roda-migrator-linux-32	migration services	JBoss
roda-migrator-win	migration services	JBoss

**Tabela 2 – Distribuição dos serviços do RODA.**

## 2.3 Sistemas operativos

O RODA é composto por vários serviços especialmente desenvolvidos para trabalhar sobre sistemas operativos Linux 64 bit. Não obstante, partes do sistema operam sobre outros sistemas operativos, nomeadamente, Windows e Linux 32 bit.

A Tabela 3 enumera os diferentes sistemas operativos que suportam os serviços do RODA.



Servidor	Sistema operativo	Tipo
roda-wui	Ubuntu Server 8.10 (64 bit)	Virtual
roda-core	Ubuntu Server 8.10 (64 bit)	Virtual
roda-data	Ubuntu Server 8.10 (64 bit)	Virtual
roda-migrator-linux	Ubuntu Server 8.10 (64 bit)	Real
roda-migrator-linux-32	Ubuntu Server 8.10 (32 bit)	Virtual
roda-migrator-win	Windows Server 2003 (64 bit)	Virtual

**Tabela 3 – Sistemas operativos que suportam o RODA.**

É de notar que a heterogeneidade dos sistemas apresentados representa um teste às capacidades de interoperabilidade do próprio RODA. É fundamental que um sistema de preservação seja interoperável com os sistemas que lhe são contemporâneos.

## 2.4 Infra-estrutura de rede

*Descrição da rede interna e rede externa do RODA em termos dos sistemas que os suportam. Marcas, dispositivos, larguras de banda, etc.*

*Responsabilidade: DGARQ/DSI*

## 3 Segurança

A segurança dos dados é uma peça fundamental num sistema com as características e objectivos do RODA. As secções que se seguem descrevem detalhadamente os mecanismos que foram implementados para garantir que não ocorrem acessos indevidos à informação custodiada.

### 3.1 Segurança de nível físico

Todos os componentes de hardware que constituem o RODA encontram-se acondicionados numa sala apropriada cujo acesso é absolutamente restrito. O conjunto de pessoas com autorização para aceder fisicamente à sala dos servidores está perfeitamente identificado no interior da DGARQ.

Para além das pessoas autorizadas, estão previstas situações em que pessoas estranhas ao serviço tenham necessidade de aceder à sala dos servidores (e.g. operações de manutenção por entidades subcontratadas). Em todas as circunstâncias o estranho deverá ser acompanhado, por alguém pertencente aos quadros da DGARQ com autorização previamente consagrada.

Todos os acessos à sala dos servidores por pessoas estranhas ao serviço carecem de uma autorização escrita por um membro da direcção da DGARQ. Nessa autorização deverá figurar a identificação completa do técnico e a designação da organização a

que pertence, a identificação do técnico da DGARQ que o irá acompanhar e a identificação do responsável por conceder a autorização.

Para além do disposto, todo e qualquer acesso à sala dos servidores é registado, não obstante a sua afectação ao serviço ou não, sendo sempre registada a seguinte informação: nome completo do indivíduo que pretende aceder à sala dos servidores, identificação do mesmo (e.g. BI), data e hora de entrada e saída, nome do acompanhante (caso necessário) e motivo da intervenção.

*A completar pela DGARQ conforme necessário. Descrição das medidas de segurança e acesso físico às máquinas que constituem o RODA. Quem tem autorização para aceder directamente à sala dos servidores e que mecanismos foram implementados para gerir esses acessos.*

*Responsabilidade: DGARQ/DSI*

### 3.2 Segurança da rede

*Como está estruturada a rede em termos de vlans, segurança e endereçamento ip, etc. Indicar os IPs que estão afectos ao RODA e as redes que são por este utilizadas.*

*Responsabilidade: DGARQ/DSI*

### 3.3 Segurança dos dados

O RODA implementa vários mecanismos de segurança dos dados. A capacidade de modificar e/ou eliminar dados preservados pelo RODA é extremamente limitada, sendo apenas possível realizar este tipo de operações pelo administrador do sistema, indivíduo que possui permissões máximas sobre toda a infra-estrutura tecnológica.

Adicionalmente, toda a comunicação entre utilizadores e o RODA é efectuada através de protocolos cifrados do tipo HTTPS/SSL, impedindo a inspecção da informação por terceiros não autorizados.

Para além do disposto, a informação incorporada no RODA não pode ser eliminada. As representações originalmente incorporadas no RODA não podem ser eliminadas por utilizadores do sistema. Essa funcionalidade simplesmente não existe. É, contudo, possível modificar as suas permissões de acesso, fazendo com que determinada informação seja apenas acessível por determinados grupos ou utilizadores.

Todas as acções realizadas por utilizadores do RODA (humanos ou não humanos) são registadas automaticamente pelo sistema em dois locais distintos, numa base de dados MySQL (para rápida consulta) e num ficheiro de texto armazenado no sistema de ficheiros. Este registo permite conhecer todas as actividades desenvolvidas em torno do repositório e apurar responsabilidades caso seja necessário.

O registo de acções não pode ser modificado e apenas está disponível aos administradores do sistema. No registo de acções são guardadas as seguintes

informações: data e hora de ocorrência, descritivo da acção, parâmetros da acção, nome do utilizador e endereço de IP computador por este utilizado. As acções registadas são altamente pormenorizadas e estão directamente relacionadas com os métodos disponibilizados pelo RODA Core. Não há operações no RODA que fiquem por registar. Toda a actividade é registada e guardada *ad eternum*.

Os dados do RODA são alvo permanente de acções de backup para suportes externos, nomeadamente *tapes*. É preciso elaborar este paragrafo com períodos de backup e tecnologias utilizadas e quais os dados que são guardados nos backups.

O acesso aos dados é ainda condicionado por regras de *firewall*. Não é possível aceder directamente aos dados. O acesso tem obrigatoriamente de ser efectuado seguindo o modelo de 3 camadas apresentado na secção 2.1.

Para obter acesso aos dados, um utilizador deverá primeiro recorrer aos serviços disponibilizados pelo RODA WUI ou RODA Core. Estes serviços invocam métodos disponibilizados pelo Fedora Commons que, por sua vez, requisita dados ao sistema de armazenamento remoto. Esta cadeia de eventos não pode ser rompida, pois infra-estrutura de rede não o permite.

### 3.4 Segurança de nível social

Os dados incorporados no RODA encontram-se protegidos por um conjunto de regras e permissões que impedem o acesso a pessoas não autorizadas.

Grupo	Operações disponíveis
Visitantes	Utilizadores que apenas têm acesso às páginas de conteúdo estático, i.e. as páginas que possuem a informação de rosto do repositório.
Consumidores	Utilizadores que, para além de poderem realizar as operações disponíveis para o grupo de visitantes, podem ainda efectuar pesquisas no repositório, navegar na metainformação descritiva e consultar as representações de acesso-livre.
Produtores	Utilizadores que, para além de poderem realizar as operações disponíveis para o grupo de consumidores, podem ainda submeter representações ao repositório (SIP) e consultar a sua lista de submissões.
Arquivistas	Utilizadores que, para além de poderem realizar as tarefas disponíveis ao grupo de consumidores podem ainda aceitar/rejeitar pacotes de submissão, editar metainformação descritiva e criar planos de classificação.
Administradores	Utilizadores que podem realizar todo o tipo de operações no repositório incluindo consultar estatísticas do sistema, criar novos utilizadores e gerir as suas permissões, consultar o registo de acções realizadas no repositório e agendar acções de preservação (e.g. migrações, verificações de integridade, novas tarefas de ingestão, etc.)

**Tabela 4 – Perfis de utilizadores do RODA.**

O RODA define 5 perfis de utilizadores, com níveis crescentes de acesso e responsabilidade. A Tabela 4 enumera esses perfis.

## 4 Processos

O RODA disponibiliza ferramentas para realizar quatro processos fundamentais: ingestão de nova informação no repositório, preservação da informação custodiada, gestão do repositório e disseminação dos materiais armazenados.

As secções que se seguem descrevem detalhadamente o funcionamento interno de cada um destes macro processos.

### 4.1 Ingestão

O processo de ingestão é responsável pela incorporação de novo material no repositório. O repositório é responsável por desenvolver todas as acções necessárias para que estes sejam adequadamente descritos, armazenados, indexados e preservados no seu interior.

Os materiais devem ser inicialmente preparados pelos produtores da informação recorrendo a aplicações desenvolvidas especificamente para esse efeito (e.g. RODA-in). Para mais informações sobre esta actividade é favor consultar o Manual de procedimentos de ingestão.

A informação submetida ao RODA é enviada em pacotes individuais designados SIP (Submission Information Package). Um pacote contém no seu interior uma representação digital da informação a incorporar e um conjunto de registos de metainformação que têm como objectivo descrever a representação e informar o repositório das suas principais características técnicas. Para mais informações sobre a estrutura interna dos pacotes de submissão é favor consultar a secção 5.2 na página 35.

O processo de ingestão é composto por dois subprocessos:

- 1) validação automática;
- 2) validação semântica (manual).

O primeiro destes subprocessos é responsável por validar a estrutura, completude, integridade, sintaxe e a conformidade com a política de ingestão dos vários SIP submetidos pelo produtor. A Tabela 5 enumera as diferentes etapas desta fase de validação.

Se alguma destas etapas falhar, o SIP será automaticamente rejeitado pelo sistema e o produtor será informado, através de correio-electrónico, do motivo da rejeição. O produtor é encorajado a resolver o problema e a voltar a submeter o SIP.

#	Etapas	Descrição	Estado seguinte	Agente
1	Descompressão do SIP	Todos os SIP são comprimidos no formato ZIP. Para dar início ao seu processamento, estes são inicialmente descomprimidos.	Descomprimido	Unpack SIPs

2	Despiste de vírus	Todos os ficheiros contidos no SIP são testados para garantir que não são portadores de vírus. O RODA recorre ao anti-vírus ClamAV para efectuar esta verificação.	Livre de vírus	Virus check
3	Validação sintáctica e estrutural do SIP	Nesta etapa os SIP são verificados para aferir se estão estruturalmente bem formados. Esta etapa é composta por cinco validações distintas.	Bem formado	Check SIP Syntax
3.1	Verificação de sintaxe do envelope METS	Verifica se o envelope METS está bem formado e é válido de acordo com o schema do METS.	-	-
3.2	Verificação de completude do SIP	Verifica se todos os ficheiros referenciados no envelope METS se encontram no interior do SIP.	-	-
3.3	Verificação de integridade dos ficheiros	Todos os ficheiros incluídos no SIP são acompanhados de um <i>checksum</i> baseado no algoritmo MD5. Esse <i>checksum</i> encontra-se no ficheiro METS que envolve todos os ficheiros do SIP. Para cada ficheiro é calculado o <i>checksum</i> e comparado com o MD5 descrito no ficheiro METS.	-	-
3.4	Validação da metainformação descritiva	Verifica se no interior do SIP existe um registo bem formado de metainformação descritiva e válida de acordo com o schema do EAD-C .	-	-
3.5	Verificação de existência de representação	Verifica se existe pelo menos uma representação no interior do SIP. Caso não exista pelo menos uma representação o SIP será rejeitado.	-	-
4	Verificação de permissões	Verifica se o produtor tem permissões para depositar informação no plano de classificação especificado. Esta tarefa evita que utilizadores não autorizados depositem informação em fundos sobre os quais não possuem permissões de depósito.	Autorizado	Check Producer Authorization
5	Incorporação temporária	Os ficheiros fornecidos pelo produtor são incorporados numa área temporária do repositório. Nesta fase são criados temporariamente os objectos Fedora que irão constituir o AIP definitivo. O AIP temporário ainda poderá sofrer alterações, nomeadamente, a metainformação descritiva pode ser alterada e pode ser acrescentada uma representação normalizada da informação depositada.	Incorporado	Create AIP
6	Normalização de formatos	As representações que não se encontrem no formato de normalização serão convertidas durante o processo de ingestão. Para tal será invocado um agente de conversão de acordo com uma tabela de transformação previamente definida.	Normalizado	Normalize format

**Tabela 5 – Etapas da validação automática de SIPs.**

Após a validação automática, os SIP serão alvo de validação semântica. Todos os SIP, sem excepção, são vistoriados manualmente por um arquivista de modo a aferir o cumprimento das seguintes condições:

- Estar previsto no Plano de transferência previamente negociado (ver Acordo de custódia contratado entre o RODA e cada produtor no documento de Política de preservação digital);
- Incluir todos os elementos de metainformação declarados obrigatórios pelo RODA (ver Manual de procedimentos de ingestão);
- Apresentar-se com suficiente e correcto conteúdo nos elementos metainformação descritiva fornecidos;
- Estar correctamente normalizado, ou seja, a representação no formato normalizado<sup>1</sup> não apresentar disformidades aparentes em relação ao formato originalmente submetido pelo produtor e incorporar correctamente todas as propriedades significativas identificadas para a classe das representações a que a informação corresponde (ver documento de Política de preservação digital).

O arquivista deverá rejeitar o SIP no caso de incumprimento de alguma das condições anteriormente descritas. O arquivista é responsável por indicar o motivo da rejeição, informação essa que será enviada para o produtor através de correio-electrónico.

## **4.2 Preservação**

O RODA implementa um conjunto de mecanismos que garantem a correcta preservação e o acesso continuado à informação custodiada.

As secções seguintes procuram descrever detalhadamente as estratégias adoptadas pelo RODA no sentido de assegurar a preservação dos materiais em cada um dos níveis de abstracção a que estes podem ser considerados, i.e. físico, lógico e conceptual e social.

No que diz respeito à preservação de nível organizacional é favor consultar o documento de Política de preservação digital.

### **4.2.1 Nível físico**

No que diz respeito à preservação de nível físico, o RODA garante a preservação da informação através do acondicionamento adequado dos seus suportes físicos de armazenamento em sala apropriada, de ambiente controlado e acesso restrito. O RODA recorre a suportes físicos de armazenamento, nomeadamente, discos e tapes,

---

<sup>1</sup> Toda e qualquer representação apenas será arquivada se se encontrar no formato normalizado definido pela política de preservação em vigor. A conversão da informação para o formato de normalização pode ser efectuada pelo próprio produtor ou pelo repositório durante o processo automático de ingestão.

de elevada qualidade para assegurar que o tempo de vida esperado destes suportes é cumprido. (É preciso referir o Mean Time Between Failure (MTBF) dos suportes físicos).

Adicionalmente, todos os discos incluídos na infra-estrutura implementam mecanismos de redundância de modo a garantir o acesso à informação mesmo em situações de avaria de alguns discos.

A informação que suporta o **roda-wui** é efémera no sentido em que pode voltar a ser gerada a partir da informação que se encontra armazenada. Devido a esse facto, o **roda-wui** apenas implementa RAID 0 de modo a maximizar a sua performance. Este tipo de RAID não implementa tolerância a falhas.

Os serviços disponibilizados pelo **roda-core** são suportados por um sistema de RAID 1, tolerante à falha de 50% dos seus discos.

Os serviços que armazenamento fornecidos pelo **roda-data** implementam RAID 5 (tolerante à falha de 1 disco por cada 3 discos da infra-estrutura).

Para mais informações, consulte a Figura 2 na página 8.

## 4.2.2 Nível lógico

Do ponto de vista da preservação lógica da informação, o RODA incorpora os seguintes mecanismos:

### Cópias de segurança

Todas as representações, metainformação, *logs* e bases de dados que suportam o repositório são replicadas para tapes numa base diária e incremental.

(É preciso completar com a política de backups);

### Verificação de integridade

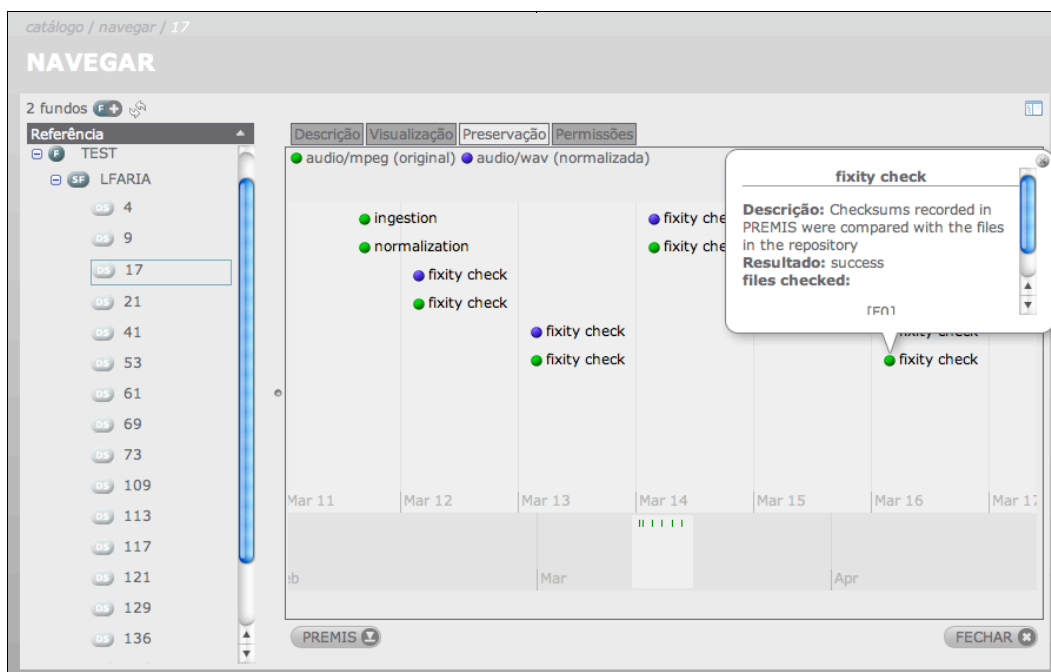
Todas as representações custodiadas no RODA são alvo de verificação permanente e contínua.

O RODA possui agendada uma tarefa de preservação que é responsável por verificar a integridade da informação armazenada. Este procedimento consiste em calcular os *checksums* MD5 e SHA-1 de cada ficheiro que constitui uma dada representação digital e comparar o resultado dessa operação com os *checksums* que se encontram registados na metainformação técnica associada a cada ficheiro custodiado. Esta metainformação técnica foi automaticamente gerada durante o processo de ingestão.

Se o teste de integridade falhar, o administrador é imediatamente notificado através de correio-electrónico. Diligencias deverão então ser tomadas no sentido de repor a representação a partir do último backup realizado.

O resultado de todas as verificações de integridade realizadas é adicionado à metainformação de preservação associada a cada representação custodiada. Esta

informação pode ser livremente inspecionado no painel de preservação por todos os utilizadores que tiverem permissões de leitura sobre uma representação.



**Figura 3 - Visualização do resultado da verificação de integridade numa representação.**

### Refrescamento de suportes

Todos os discos que suportam o sistema serão imediatamente trocados ao menor sinal de avaria. O RODA utiliza sistemas de RAID 1 e 5 com *hotswap* que permitem a substituição de discos em eminência de falha sem necessidade de desligar o sistema.

O sistema implementa ainda um disco redundante designado *hot spare* que entra imediatamente em funcionamento na eventualidade de um disco da infra-estrutura falhar, assumindo as responsabilidades deste.

### 4.2.3 Nível semântico

A preservação de nível semântico é fundamental para garantir a inteligibilidade dos materiais preservados. Estas acções de preservação são realizadas de modo a fazer cumprir a política de acesso definida no documento de Política de preservação digital.

Numa primeira fase, todas as representações submetidas ao RODA são convertidas para formatos mais adequados à sua preservação. Este processo designa-se vulgarmente por normalização e apenas é realizado caso as representações submetidas pelo produtor não se encontrem já no formato normalizado.



Previamente à sua aceitação, as representações normalizadas serão individualmente inspeccionadas por arquivistas de modo a garantir que as propriedades significativas associadas à respectiva classe de objectos digitais foram devidamente conservadas durante o processo de normalização.

O RODA preserva *ad eternum* a representação original submetida pelo produtor, no entanto, apenas garante o acesso continuado e inteligível à representação normalizada a partir da qual serão produzidas versões para disseminação e consumo. A estratégia que permite ao RODA garantir esta premissa baseia-se na **migração de formatos**.

O RODA permite facilmente adicionar ao sistema novas tarefas de migração de formatos através de plug-ins. Estas tarefas podem ser agendas e aplicadas transversalmente às representações armazenadas. Cada tarefa de migração é responsável por seleccionar as representações sobre as quais deverá operar<sup>2</sup>.

Cada plug-in de migração regista junto da metainformação de preservação a data e hora da operação realizada, o agente responsável pela execução da tarefa (i.e. conversor utilizado) e o resultado da conversão (sucesso ou insucesso). É ainda criada na metainformação de preservação uma relação para o objecto a partir do qual a nova representação foi criada, estabelecendo assim uma cadeia lógica de transmissão de informação desde a sua versão original até à versão mais actual preservada pelo repositório. As representações que serviram de base à criação de novas representações não são eliminadas através deste processo.

Há, no entanto, lugar para que após ordem política sejam eliminadas as representações intermédias cuja preservação não apresente valor acrescentado para o repositório e para os seus utilizadores. Não obstante, a metainformação de preservação nunca é eliminada, mesmo aquela que diz respeito a representações que já não existem no interior do repositório.

---

<sup>2</sup> Por exemplo, o plug-in desenvolvido para o RODA que converte documentos de texto em formato DOCX para PDF/A apenas converte representações que se encontram no formato DOCX, ignorando todas as outras.

A Tabela 6 apresenta um resumo dos conversores suportados pelo RODA.

Classe	Agente	Formato de partida	Formato de chegada
Imagens matriciais (com informação estrutural)	BMP to TIFF	image/mets+bmp	image/mets+tiff
	GIF to TIFF	image/mets+gif	
	JPEG to TIFF	image/mets+jpeg	
	PNG to TIFF	image/mets+png	
	TGA to TIFF	image/mets+tga	
	XMP to TIFF	image/mets+xpm	
	Ico to TIFF	image/mets+ico	
Texto estruturado	DOC to PDF/A	application/msword	application/pdf
	DOCX to PDF/A	application/vnd.openxmlformats-officedocument.wordprocessingml.document	
	RTF to PDF/A	application/rtf	
	ODT to PDF/A	application/vnd.oasis.opendocument.text	
Áudio	AIFF to WAV	audio/aiff	audio/wav
	FLAC to WAV	audio/flac	
	MP3 to WAV	audio/mpeg	
	MP4 to WAV	audio/mp4	
	OGG to WAV	audio/ogg	
Video	AVI to MPEG 2	video/avi	video/mpeg2
	MOV to MPEG 2	video/quicktime	
	MPEG to MPEG 2	video/mpeg	
	MPEG 4 to MPEG 2	video/mp4	
	WMV to MPEG 2	video/x-ms-wmv	

**Tabela 6 - Conversores suportados pelo RODA.**

#### 4.2.4 Nível social

O RODA incorpora mecanismos que previnem a ocorrência de erro humano. O sistema de permissões implementado no RODA impede que utilizadores (humanos

ou não-humanos) efectuem operações para as quais não estão autorizados. A atribuição de permissões e a sua manutenção ao longo do tempo são da responsabilidade do administrador do sistema.

Todos os utilizadores do RODA têm obrigatoriamente de estar registados no sistema para poderem efectuar qualquer operação de consulta, pesquisa, incorporação, administração ou outra.

Todas as actividades desenvolvidas no sistema são registadas em dois locais distintos: ficheiro de texto armazenado no sistema de ficheiros e base de dados relacional (para rápida consulta e apresentação na interface gráfica).

O sistema não permite a execução de qualquer operação que não fique registada. São realizadas cópias de segurança de ambos os registos de acordo com a política de *backup* em vigor.

### **4.3 Gestão**

O RODA implementa um conjunto de procedimentos de gestão que permitem conhecer o estado actual do RODA e efectuar operações de manutenção sobre o repositório.

A generalidade destes procedimentos encontram-se acessíveis exclusivamente aos utilizadores pertencentes ao grupo de Administradores. Os administradores possuem poderes absolutos sobre o repositório, pelo que a atribuição deste tipo de credenciais é limitada.

#### **4.3.1 Edição de metainformação**

A edição de metainformação é uma operação que pode ser realizada por todos os utilizadores pertencentes aos grupos Administradores e Arquivistas.

Esta operação permite editar qualquer registo de metainformação descritiva, incluindo a metainformação fornecida pelo produtor durante a submissão de uma nova representação, no sentido de melhor organizar a informação no interior do repositório e facilitar a sua futura localização.

Todas as alterações aos registos descritivos são automaticamente guardadas pelo sistema, sendo sempre possível consultar uma versão anterior de um registo, saber quem realizou determinada modificação ou repor uma versão anterior do mesmo. Contudo, esta funcionalidade não se encontra disponível na interface gráfica do RODA WUI. A consulta de versões precedentes de registos de metainformação carece ainda de intervenção técnica.

catálogo / navegar / 29

## NAVEGAR

11 fundos

Referência

- DGARQ
- F16
- F40
- F40
- F40
- F40
- F50
- F53
- F54
- TESTES
- lc

Descrição Produtores Permissões

### Identificação

Mostrar campos opcionais

Referência	DGARQ
Referência completa	PT/DGARQ/DGARQ
Id. persistente	http://hdl.handle.net/10384/29
Título	Direcção-Geral de Arquivos
Nível de descrição	fundo
Data inicial	2009 Janeiro 01
Data final	2009 Janeiro 01
Código do país	PT
Código do repositório	DGARQ
Produtor	Miguel Ferreira
Referência do produtor	
Data de aquisição	

### Conteúdo e estrutura

Âmbito e conteúdo	<Indefinido>
-------------------	--------------

GRAVAR SUBNÍVEL MOVER ELIMINAR EAD FECHAR

**Figura 4 – Interface de edição de metainformação.**

Para mais informações sobre o procedimento de edição de metainformação é favor consultar o Manual de procedimentos de administração.

### 4.3.2 Gestão de utilizadores

Os utilizadores do RODA podem ser registados de duas formas distintas: administrativamente pelo gestor do RODA ou através de auto-registo.

A informação sobre os utilizadores é armazenada e gerida num servidor LDAP (ver Arquitectura geral do RODA na página 5).

Após o registo, os utilizadores passam a pertencer automaticamente ao grupo de Consumidores. Este grupo possui permissões básicas que garantem o acesso às funcionalidades de pesquisa básica, pesquisa avançada e visualização de representações de carácter público.

Cabe ao administrador do RODA a atribuição do grupo correcto a cada utilizador (e.g. Produtores, Arquivistas, etc.).

O RODA guarda um registo de todas as acções realizadas pelos seus utilizadores. Após a primeira autenticação, um utilizador nunca mais poderá ser removido do sistema. Pode, no entanto, ser desactivado, impedindo a sua autenticação e, consequentemente, o acesso ao repositório.



Figura 5 – Interface de gestão de utilizadores.

### 4.3.3 Agenda

O RODA implementa um mecanismo que permite agendar tarefas. Estas tarefas poderão ser de ordem diversa, tais como: rotinas de administração, eventos de preservação, avisos, etc. Esta funcionalidade está disponível apenas aos administradores do RODA.

Todas as tarefas ou eventos podem ser geridos através de uma interface gráfica disponível no RODA WUI. Ao agendar uma nova tarefa o utilizador deve definir o nome, a descrição, a data de início, as opções de repetição da tarefa (i.e. de quanto em quanto tempo e quantas vezes deve ser repetida), o agente de software responsável pela desenvolvimento da tarefa e, eventualmente, parâmetros que a tarefa possa necessitar.

O RODA conserva um histórico de todas as tarefas executadas sendo possível conhecer a qualquer momento quem agendou a tarefa, quando foi executada, quando terminou e qual o resultado prático da sua execução (e.g. sucesso, insucesso, etc).

A calendarização e execução de tarefas está a cargo de um componente do sistema designado Quartz<sup>3</sup>.

A qualquer momento é possível modificar o agendamento de uma tarefa de modo a ajustá-la aos requisitos definidos na política de preservação. É também possível pausar uma tarefa, por exemplo, para realizar uma operação de manutenção.

<sup>3</sup> <http://www.opensymphony.com/quartz/>

administração / agenda / agenda

## AGENDA

Agenda Lista

8 eventos

Nome	Data de início	Repetições	Criador
calc statistics	2009-02-13 22:18:42	repetir cada 1 dias para sempre	mferreira
Check producer	2009-02-12 22:07:00	repetir cada 1 minutos para sempre	rcastro
check syntax	2009-02-12 22:06:24	repetir cada 1 minutos para sempre	mferreira
Check virus	2009-02-12 22:09:33	repetir cada 1 minutos para sempre	mferreira
Create AIP	2009-02-12 22:39:00	repetir cada 1 minutos para sempre	rcastro
Monitor FTP	2009-02-12 22:08:26	repetir cada 1 minutos para sempre	mferreira
Normalization	2009-02-12 22:08:00	repetir cada 1 minutos para sempre	rcastro
unpack	2009-02-13 22:17:54	repetir cada 1 minutos para sempre	mferreira
unpack	2009-02-13 22:17:54	repetir cada 1 minutos para sempre	mferreira

**Listar**

agendados  
pausados  
não agendados  
a correr  
concluídos  
todos

**Nome**

NOVA  
EDITAR  
REMOVER  
PAUSAR  
RESUMIR

### 4.3.4 Listagens

O RODA apresenta ao utilizador informação variada sob a forma de listagens. Por exemplo, é possível obter listagens de tarefas executadas por utilizadores, pacotes de submissão em processamento, etc.

Todas as listagens apresentadas pelo RODA podem ser descarregadas em formato PDF (para impressão) ou CSV (Comma Separated Values, compatível com a generalidade das folhas de cálculo, e.g. Excel).

Uma vez que a maior parte destas listagens pode crescer *ad eternum*, por uma questão de escalabilidade, a exportação deste tipo de informação está limitado a 1000 registos de informação. Este número é configurável.

# Relatório RODA: Lista de pacotes

<https://roda.di.uminho.pt:443>

## Identificação do relatório

Data de geração	2009-02-26 17:42:50
Gerado por	Miguel Ferreira (utilizador: mferreira, e-mail: mferreira@dsi.uminho.pt)
Filtro de conteúdo	Estado tem ser igual a SIP_NORMALIZED, e Completo igual a falso
Intervalo do conteúdo	Mostrando todos os 6 elementos
Ordem do conteúdo	A ordenar por datetime (descendente), e Identificador (descendente)

## Pacote 33

Nome original:	TESTES_ding_ACCESS26642950520434922750.sip
Data de submissão:	2009-02-26T12:42:53.00Z
Estado:	SIP_NORMALIZED
Percentagem:	100%
Identificador:	roda:966
Acesso:	<a href="https://roda.di.uminho.pt:443/#dissemination.browse.966">https://roda.di.uminho.pt:443/#dissemination.browse.966</a>
Produtor:	lfaria

## Pacote 32

Nome original:	TESTES_ding_ACCESS39205292825474096066.sip
Data de submissão:	2009-02-25T15:41:51.00Z
Estado:	SIP_NORMALIZED
Percentagem:	100%
Identificador:	roda:921
Acesso:	<a href="https://roda.di.uminho.pt:443/#dissemination.browse.921">https://roda.di.uminho.pt:443/#dissemination.browse.921</a>
Produtor:	lfaria

**Figura 6 - Exemplo de uma listagem em formato PDF.**

Actualmente, o RODA suporta as seguintes listagens:

### Listagem de pacotes

Permite aos utilizadores responsáveis pela ingestão (i.e. os Arquivistas) identificar os pacotes de informação submetidos pelos produtores de informação e qual o seu estado actual. O sistema é capaz de fornecer listagens individuais para os pacotes em processamento, processados, rejeitados, aceites ou todos.

### Listagem de utilizadores/grupos

O administrador pode obter uma listagem com todos os utilizadores ou grupos registados no sistema. Esta listagem específica não pode ser exportada para PDF ou CSV.

### Registo de acções de um utilizador

Descreve todas as actividades desenvolvidas por um utilizador em torno do repositório. Este relatório tem como missão permitir ao administrador do sistema conhecer a actividade de um dado utilizador.

**Agenda de tarefas**

Permite ao administrador do repositório obter uma listagem de todas as tarefas agendadas no RODA.

**Histórico de tarefas**

Permite ao administrador do repositório obter uma listagem de todas as tarefas já executadas no RODA.

**Relatório de acções**

Descreve todas as actividades desenvolvidas em torno do repositório. Este relatório tem como missão permitir ao administrador do sistema saber quem fez o quê no repositório.

**4.3.5 Estatísticas e relatórios de auditoria**

O RODA produz um conjunto de estatísticas e relatórios de auditoria que permitem ao administrador aferir o “estado de saúde” do sistema e agir em conformidade (Figura 7).

Todos os dados utilizados para produzir os gráficos estatísticos apresentados ou os relatórios de auditoria podem ser descarregadas em formato PDF ou CSV (para processamento externo numa folha de cálculo).





**Figura 7 – Consulta de estatísticas produzidas pelo RODA.**

O repositório produz actualmente os seguintes relatórios e estatísticas:

### **Estado do repositório**

1. N° total de registos descritivos
2. N° total de fundos
3. N° total de representações
4. N° total de depósitos
5. N° total de registos de metainformação de preservação (representações)
6. N° total de registos de metainformação de preservação (eventos)
7. Distribuição de classes de representação
8. Distribuição de formatos

### **Estatísticas de ingestão**

1. % de pacotes de submissão completos
2. Estado dos pacotes de submissão
3. Tempo médio de processamento automático de ingestão
4. Tempo médio de tarefas manuais de ingestão
5. Tempo mínimo de processamento automático de ingestão

6. Tempo máximo de processamento automático de ingestão
7. Tempo mínimo de tarefas manuais de ingestão
8. Tempo máximo de tarefas manuais de ingestão

#### **Informação sobre produtores**

1. Data da última submissão de um produtor
2. Distribuição do estado dos depósitos de um produtor

#### **Estatísticas sobre utilizadores**

1. N° total de utilizadores
2. N° total de grupos de utilizadores
3. Distribuição dos 5 maiores grupos de utilizadores

#### **Estatísticas de acesso**

1. N° total de autenticações
2. N° total de páginas visualizadas
3. N° total de eventos anómalos (interface gráfica)
4. N° total de consultas de metainformação descritiva
5. N° total de consultas de metainformação de preservação
6. N° total de pesquisas básicas efectuadas
7. N° total de pesquisas avançadas efectuadas
8. Distribuição de disseminadores acedidos
9. N° total de derivadas geradas
10. N° total de registos de utilizadores
11. N° total de endereços electrónicos confirmados
12. N° total de palavras-passe recuperadas

#### **Estatísticas sobre acções de utilizadores**

1. Criação de registos descritivos
2. Edição de metainformação descritiva
3. Remoção de registos

4. Alterações na hierarquia descritiva
5. Criação de registos de utilizadores
6. Criação de registos de grupos de utilizadores
7. Edição de registos de utilizadores
8. Edição de registos de grupos de utilizadores
9. Remoção de registos de utilizadores
10. Remoção de registos de grupos de utilizadores
11. Alteração da palavra-passe de registos de utilizadores
12. Aceitação/Rejeição de pacotes de submissão
13. Criação de eventos
14. Edição de eventos
15. Remoção de eventos

### **Informação sobre o estado da infra-estrutura**

O RODA implementa um sistema de monitorização que faculta ao administrador informação variada sobre estado da infra-estrutura que suporta o repositório. Actualmente o RODA possui suporta para relatórios produzidos pelo sistema Munin<sup>4</sup>.

O sistema implementado apresenta informação acumulada ao longo do tempo sobre a memória disponível, a taxa de ocupação e de utilização dos discos, a carga média no processador, a utilização da rede, a carga no servidor aplicacional JBoss, o estado da base de dados MySQL, entre outros (Figura 8).

---

<sup>4</sup> <http://munin.projects.linpro.no/>

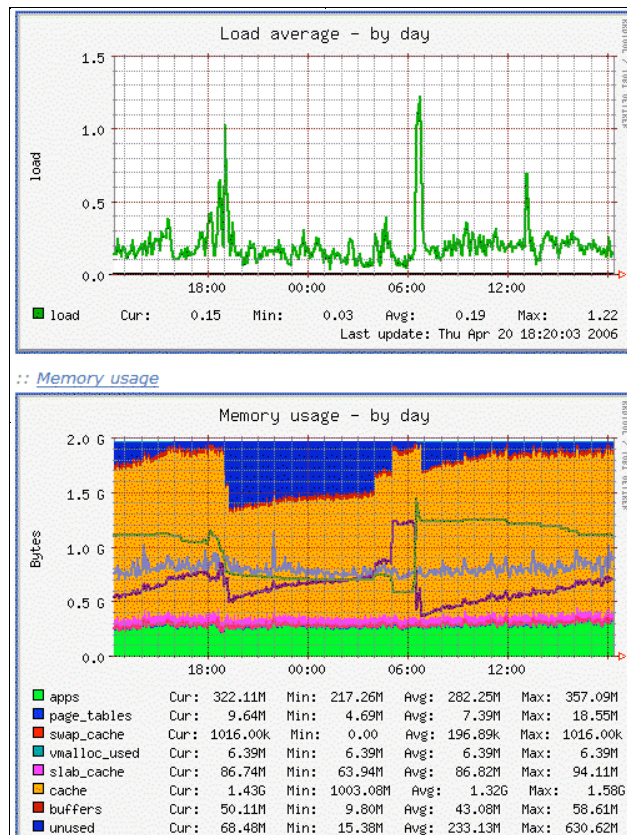


Figura 8 - Monitorização do processador e memória.

#### 4.3.6 Relatório de acções

O RODA oferece uma funcionalidade que permite ao administrador conhecer toda a actividade desenvolvida em torno do repositório. O administrador, através de interface própria disponível através do RODA WUI, poderá filtrar o registo de acções por tipo de acção e por datas extremas.

O registo armazena a seguinte informação para todas as acções realizadas no repositório:

- Data e hora de execução
- Designação da acção
- Parâmetros da acção
- Nome do utilizador
- Endereço IP do utilizador

administração / relatório de ações

## RELATÓRIO DE AÇÕES

Ação: 
 Data inicial: 
 Data final: 
 APLICAR

82783 entradas no registro

Data e hora	Ação	Parâmetros	Utilizador
2009-02-26 10:45:34.0	IngestMonitor getSIPs	contentAdapter: ContentAdapter(filter=Filter ([OneOfManyFilterParameter(name=state, values=[ACCEPTED]), SimpleFilterParameter(name=complete, value=true))), sorter=Sorter ([SortParameter(name=datetime, descending=true), SortParameter(name=id, descending=true))), sublist=Sublist(firstElementIndex=9, maximumElementCount=30))	mferreira
2009-02-26 10:45:33.0	IngestMonitor getSIPsCount	filter: Filter ([OneOfManyFilterParameter(name=state, values=[ACCEPTED]), SimpleFilterParameter(name=complete, value=true)])	mferreira
2009-02-26 10:45:33.0	IngestMonitor getSIPs	contentAdapter: ContentAdapter(filter=Filter ([OneOfManyFilterParameter(name=state, values=[ACCEPTED]), SimpleFilterParameter(name=complete, value=true))), sorter=Sorter ([SortParameter(name=datetime, descending=true), SortParameter(name=id, descending=true))), sublist=Sublist(firstElementIndex=0, maximumElementCount=9))	mferreira

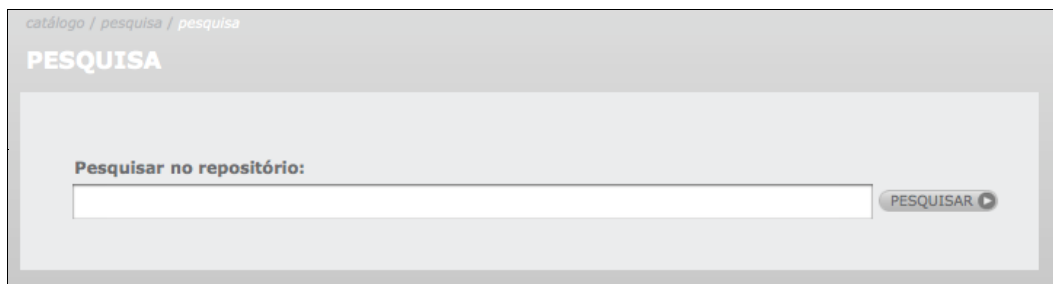
## 4.4 Disseminação

As funcionalidades de disseminação oferecidas pelo sistema estabelecem a ponte entre a comunidade de interesse do RODA e as representações nele custodiadas. Esta unidade funcional permite aos utilizadores do RODA localizar representações, visualizá-las, descarregá-las e conhecer os seus relacionamentos semânticos com outras representações.

### 4.4.1 Localização de representações

O RODA disponibiliza aos seus consumidores três modalidades de localização de informação: pesquisa básica (Figura 9), pesquisa avançada (Figura 10) e navegação sobre fundos documentais (Figura 11).

A pesquisa básica e a pesquisa avançada são as formas primordiais de localização de informação. A funcionalidade de navegação deve ser utilizada sobretudo para encontrar localização relacionada que fora publicada sensivelmente na mesma data.



**Figura 9 - Interface de pesquisa básica.**

A pesquisa básica permite encontrar um conjunto de palavras em todos os elementos de metainformação descritiva suportados pelo RODA, e.g. referência, nível de descrição, código do repositório, identificador persistente, título, datas, âmbito e conteúdo, e história administrativa e biográfica, etc.

A pesquisa avançada oferece um conjunto acrescido de parâmetros de consulta, nomeadamente: selecção de níveis de descrição, datas extremas e atributos a pesquisar.

catálogo / pesquisa / pesquisa avançada

## PESQUISA AVANÇADA

**Selecione os níveis de descrição que pretende incluir na pesquisa**

☒ todos os níveis  
☐ escolher níveis

<input checked="" type="checkbox"/> fundo	<input checked="" type="checkbox"/> classe	<input checked="" type="checkbox"/> serie	<input checked="" type="checkbox"/> documento composto
<input checked="" type="checkbox"/> sub-fundo	<input checked="" type="checkbox"/> sub-classe	<input checked="" type="checkbox"/> sub-serie	<input checked="" type="checkbox"/> documento simples

**Selecione o intervalo de tempo em que pretende pesquisar**

☐ sem restrições temporais  
☒ especificar intervalo

desde  Janeiro   até  Janeiro

**Preencha as palavras que pretende encontrar**

☐ todos os campos  
☒ escolher campos

☒ todas as palavras  
☐ frase exacta  
☐ pelo menos uma das palavras  
☐ sem as palavras

<input type="checkbox"/> Referência completa	<input type="checkbox"/> Citação	<input type="checkbox"/> Características físicas e requisitos técnicos
<input type="checkbox"/> Título	<input type="checkbox"/> História administrativa e biográfica	<input type="checkbox"/> Condições de acesso
<input type="checkbox"/> Produtor	<input type="checkbox"/> História custodial	<input type="checkbox"/> Condições de reprodução
<input type="checkbox"/> Referência do produtor	<input type="checkbox"/> Modalidades de aquisição	<input type="checkbox"/> Unidades de descrição relacionadas
<input type="checkbox"/> Detalhes específicos	<input type="checkbox"/> Âmbito e conteúdo	<input type="checkbox"/> Materiais associados
<input type="checkbox"/> Descrição física	<input type="checkbox"/> Organização e ordenação	<input type="checkbox"/> Bibliografia
<input type="checkbox"/> Dimensões	<input type="checkbox"/> Avaliação, selecção e eliminação	
<input type="checkbox"/> Aspecto ou aparência	<input type="checkbox"/> Ingressos adicionais	
<input type="checkbox"/> Extensão		

**PESQUISAR**

**Figura 10 - Pesquisa avançada.**

A navegação em fundos documentais permite consultar registos descritos de forma hierárquica, conforme o plano de classificação definido.

catálogo / navegar / 29

## NAVEGAR

12 fundos

Referência	Título	Data inicial	Data final
DGARQ	Direcção-Geral de Arquivos	2009-01-01	2009-01-01
374	Teste	2001-01-01	2003-01-01
386	Teste de referencia do produtor	2000-01-01	2100-01-01
F16	Teste fonds	2009-01-12	2009-01-15
F40	Teste fonds	2009-01-12	2009-01-15
F40	Teste fonds	2009-01-12	2009-01-15
F40	Teste fonds	2009-01-12	2009-01-15
F40	Teste fonds	2009-01-12	2009-01-15
F40	Teste fonds	2009-01-12	2009-01-15
F50	Teste fonds	2009-01-12	2009-01-15
F53	Teste fonds	2009-01-12	2009-01-15
F54	Teste fonds	2009-01-12	2009-01-15
TESTES	Fundo de testes	2000-01-01	2001-01-01
lc	teste ppd	0001-01-01	0002-01-01

Figura 11 - Navegação em fundos documentais.

#### 4.4.2 Visualização de representações

O RODA implementa várias formas de consulta de representações, nomeadamente:

##### 4.4.2.1 Descarregar

Disseminador para descarregar a representação tal como ela está guardada no interior do repositório. Caso a representação seja composta por um só ficheiro, esse ficheiro será enviado para o consumidor. Alternativamente, se esta for composta por vários ficheiros, então um zip é criado e enviado.

##### 4.4.2.2 Imagens

SimpleViewer é um visualizador de imagens baseado em Adobe Flash e com uma interface optimizada para fotos. Este disseminador utiliza os conversores existentes no RODA-Migrator para converter imagens em imagens optimizadas para o SimpleViewer e definir a sua configuração.

##### 4.4.2.3 Recursos paginados

FlashPageFlip é visualizador para imagens e documentos de texto baseado em Adobe Flash que apresenta uma interface semelhante a um livro. Este disseminador utiliza os conversores existentes no RODA-Migrator para converter documentos de texto ou imagens para imagens optimizadas para o respectivo disseminador e definir a sua



configuração. Este disseminador incorpora funcionalidades de zoom, visualização em ecrã completo, impressão de páginas, alteração de papel de parede, controlo por teclado, entre outras.

#### **4.4.2.4 Multimédia Player**

Disseminador optimizado para consultar representações baseadas em áudio e vídeo. Este disseminador usa os conversores existentes no RODA-Migrator para criar recursos em FLV ou MP3. O multimédia player utilizado, baseado em Adobe Flash, denomina-se JW FLV Media Player.

#### **4.4.2.5 Bases de dados**

Para consultar bases de dados é disponibilizado um sistema designado phpMyAdmin. Este sistema oferece ao utilizador uma poderosa capacidade de navegação, pesquisa e exportação de informação incluída em bases de dados.

Este disseminador usa os conversores existentes no RODA-Migrator para inserir uma base de dados em formato DBML no DBMS MySQL. O disseminador também permite visualizar o dicionário de dados da base de dados disseminada.

## 5 Organização e estrutura da informação

No âmbito do RODA há um conjunto alargado de informação que é conservada e gerida pelo sistema. Entre outros tipos de informação, o RODA é responsável por gerir metainformação, SIPs, AIPs e DIPs.

As secções que se seguem caracterizam cada uma destas unidades de informação.

### 5.1 Metainformação

De modo a suportar a sua actividade, o RODA implementa diversos tipos de metainformação, nomeadamente, metainformação descritiva, de preservação, técnica e estrutural.

A metainformação descritiva cumpre o propósito de manter os materiais bem organizados e facilmente localizáveis. Esta metainformação é indexada pelo repositório e é sobre esta metainformação que são realizadas todas as pesquisas por parte dos utilizadores.

A metainformação de preservação tem como objectivo garantir a autenticidade dos materiais. É junto desta metainformação que são registados os resultados de todas as acções de preservação realizadas sobre representações.

A metainformação técnica agrupa um conjunto de dados específicos do formato da representação digital. Este tipo de metainformação tem como principal objectivo detectar eventuais falhas no processo de preservação e auxiliar a tomada de decisão quanto à estratégia de preservação a implementar.

A metainformação estrutural permite descrever as relações existentes entre os diversos constituintes de uma representação, permitindo, assim, processar e interpretar a mesma de forma coerente e correcta.

O RODA faz uso de todos estes tipos de metainformação. A tabela que se segue enumera as normas implementadas pelo RODA em cada um destes contextos.

Metainformação	Norma	Utilização
Descritiva	EAD <sup>5</sup>	Descrição de representações e respectivos planos de classificação
Preservação	PREMIS <sup>6</sup>	Registo de todas as acções de preservação

---

<sup>5</sup> <http://www.loc.gov/ead/>

<sup>6</sup> <http://www.oclc.org/research/projects/pmwg/>  
[http://www.niso.org/kst/reports/standards?step=2&gid=None&project\\_key=b897b0cf3e2ee526252d9f830207b3cc9f3b6c2c](http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key=b897b0cf3e2ee526252d9f830207b3cc9f3b6c2c)

		aplicadas a representações digitais
Técnica	ANSI/NISO Z39.87	Registo de características técnicas específicas de formatos de imagem
Estrutural	METS <sup>7</sup>	Envelope que acompanha pacotes de submissão de informação
Estrutural	RDF <sup>8</sup>	Criação de relações entre objectos de descrição, preservação e representações

**Tabela 7 - Tipos de metainformação gerida pelo RODA.**

## 5.2 Caracterização dos SIP

O SIP (*Submission Information Package*) é o formato utilizado pelo produtor para transferir informação para o repositório.

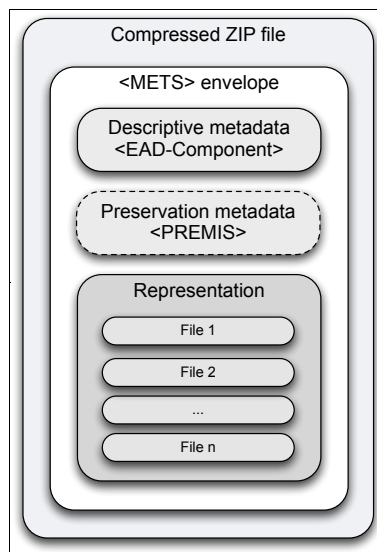
Estes pacotes de informação são compostos por uma ou mais representações, um registo de metainformação descritiva por cada representação e restante metainformação de preservação e técnica (se existente). Todos esses ficheiros são acompanhados de um envelope METS para lhes dar ordem e estrutura. Finalmente, todos esses ficheiros são comprimidos num ficheiro ZIP, ficheiro esse que constitui efectivamente o SIP a enviar pelo produtor. A Figura 12 ilustra a estrutura de um SIP.

Dentro do SIP deverá existir um registo de metainformação descritiva por cada representação incluída. Essa metainformação deve seguir o formato EAD-Component, i.e. um subconjunto da norma EAD que diz respeito à descrição de um único nível ou unidade de informação, ao invés de uma colecção completa. Trata-se efectivamente do conteúdo de um elemento <c> da norma EAD.

Todavia, é também possível encontrar metainformação de preservação no interior de um SIP. Apesar de esta informação ser raramente preenchida pelos produtores, é bastante útil em situações em que o SIP a incorporar tem origem noutro sistema de informação, i.e. em situações de sucessão ou migração de dados entre sistemas.

<sup>7</sup> <http://www.loc.gov/standards/mets/>

<sup>8</sup> <http://www.w3.org/RDF/>



**Figura 12 - Estrutura do SIP.**

### 5.3 Caracterização dos AIP

Após o processo automático de ingestão, onde são verificadas todas as características técnicas do SIP, este será alvo de escrutínio por parte da equipa de arquivistas. Estes irão verificar se todos os aspectos diplomáticos associados ao SIP se encontram de acordo com o contrato de pré-ingestão e com a política de preservação em vigor.

Após a verificação manual de cada SIP, este poderá ser aceite ou rejeitado. No primeiro caso, procede-se então à incorporação do SIP. Ao longo desse processo, o SIP irá sofrer um conjunto de transformações nos seus constituintes para que possa devidamente armazenado e preservado pelo repositório. A nova estrutura designa-se por *Archival Information Package* (AIP).

A estrutura do AIP baseia-se no modelo de dados proposto pelo PREMIS (Figura 13). Cada entidade intelectual é descrita por um registo de metainformação, designado por **objecto de descrição**, que mais não é do que uma instância do EAD-C anteriormente descrito. Os objectos de descrição relacionam-se entre si através de relações do tipo *child-of* de modo a construir a árvore de classificação que descreve um fundo. Estas relações são criadas recorrendo aos mecanismos relacionais facilitados pelo motor de RDF incorporado no Fedora Commons.

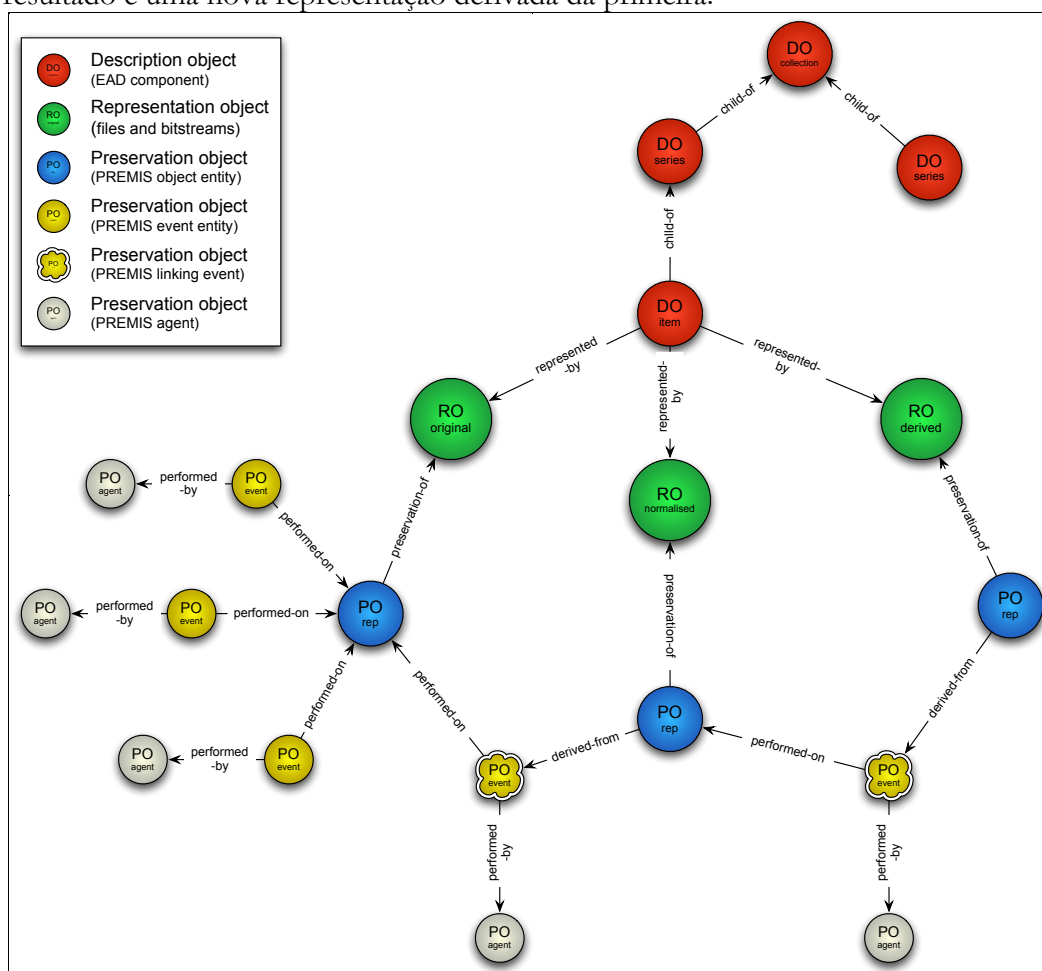
Os objectos de descrição do tipo *documento* ou *documento composto* possuem relações com um ou mais **objectos de representação**. Um objecto de representação é um objecto Fedora que incorpora todos os ficheiros e sequências binárias que constituem uma representação digital.

Finalmente, estes objectos encontra-se relacionados com um terceiro tipo de objectos designados por **objectos de preservação**. Estes têm como objectivo

documentar a proveniência da representação digital e o historial de eventos de preservação que ocorreram no interior do repositório. Os objectos de preservação podem ser de quatro tipos distintos: objecto, evento, evento com relacionamento, e agente.

Os **objectos de preservação** do tipo **objecto** registam a estrutura da representação e informação técnica sobre os ficheiros que as constituem. Os objectos do tipo **agente** descrevem qualquer pessoa, organização ou aplicação de software que tenha sido responsável pela realização de um **evento**. Por sua vez, um **evento** agrega toda a informação relativa a uma acção de preservação (e.g. verificação de integridade, migração, etc.), bem como o resultado da mesma (e.g. sucesso, insucesso, etc.).

Os objectos do tipo **evento com relacionamento** são utilizados para documentar **eventos** que originam novas representações. Exemplos disto são os eventos de migração. Tratam-se de eventos que são aplicados a uma representação e cujo resultado é uma nova representação derivada da primeira.



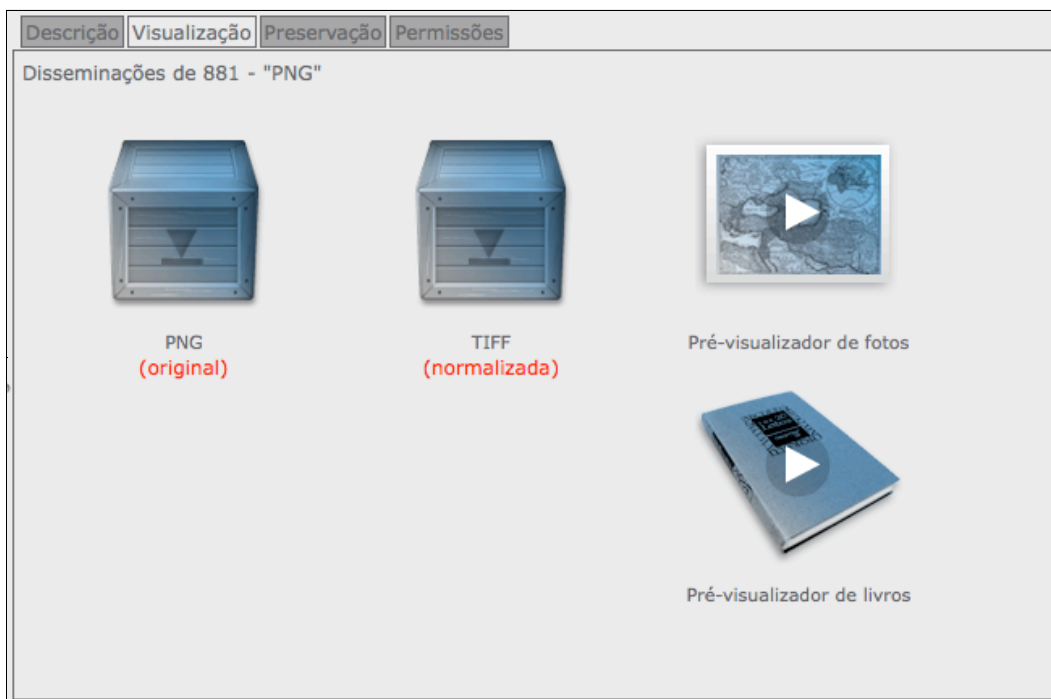
**Figura 13 – Estrutura AIP.**

## 5.4 Caracterização dos DIP

O RODA faculta o acesso às representações digitais de duas formas distintas: 1) apresentação das representações através de visualizadores Web ou disponibilizando uma hiperligação para que o consumidor possa descarregar a respectiva representação. Esta última opção permite ao consumidor utilizar o software que considerar mais conveniente para abrir e interpretar a representação descarregada (Figura 14).

Ambas as opções encontram-se devidamente assinaladas através de sinalética apropriada, i.e. descarregar (seta apontada para baixo) e visualizar (triângulo virado para a direita).

É importante referir que para determinadas classes de representações é possível haver mais do que um visualizador disponível. O utilizador é livre de utilizar o visualizador que mais lhe convier ou agradar (Figura 14).



**Figura 14 - Formas de disseminação.**

Em certos casos, representações submetidas ao RODA serão alvo de migração de formatos. Nesse caso, o consumidor terá a possibilidade de descarregar a representação originalmente submetida pelo produtor ou a versão que convertida a partir desta.

## 5.5 Política de identificadores

Todos os objectos incorporados no RODA, quer sejam objectos de descrição, de preservação, representações, ficheiros ou outros são identificados por uma referência unívoca gerada automaticamente pelo próprio sistema. Esta referência é do tipo numérico e sequencial.

Para acesso e referenciação são utilizados identificadores persistentes baseados no Handle System. Um identificador persistente tem a seguinte estrutura:

**Protocolo://servidor de resolução/identificador do repositório/identificador do objecto**

Um exemplo de um identificador persistente é <http://hdl.handle.net/10384/374>. O protocolo utilizado é o **http**, o servidor de resolução é o **hdl.handle.net**, o identificador do repositório é o **10384** e o identificador do objecto no interior do repositório é o **374**.

Durante a associação de metainformação descritiva aos pacotes de informação submetidos ao repositório, o produtor é convidado a atribuir uma referência à respectiva representação. Essa referência é mantida pelo sistema num atributo designado **referência do produtor**.

## 6 Referências

1. Ambacher, B., et al., *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. 2007, OCLC & CRL.
2. Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS) - Blue Book*. 2002, Washington: National Aeronautics and Space Administration.
3. The Fedora Project. *The Fedora Project - An Open-Source Digital Repository Management System*. [cited 2004].
4. University of Virginia and Cornell University. *Fedora Commons Web site*. [cited 2008 2008-04-21]; Available from: <http://www.fedora.info/>.
5. Fedora Commons. *Fedora Access and Management Web Services - API Documentation*. [cited 2007-11-05]; Available from: <http://www.fedora.info/definitions/api/>.
6. Corporation for National Research Initiatives. *CNRI Handle System*. [cited 2004; Available from: <http://www.handle.net/>].