



Repositório UFM

REPOSITÓRIOS DE ACESSO ABERTO E PRESERVAÇÃO DIGITAL

Este folheto informativo sintetiza informação relevante sobre o tema dos repositórios de acesso aberto e a sua relação com a preservação digital. Surge na sequência do estudo “Estado da Arte em Preservação Digital”, apresentado em fevereiro de 2012, no âmbito do projeto Repositório Científico de Acesso Aberto de Portugal (RCAAP).

Para além de uma contextualização geral do conceito de preservação digital procurou-se fazer o seu enquadramento nos repositórios de acesso aberto e exemplificar os projetos, as arquiteturas e as estratégias que têm demonstrado resultados significativos nesta área.

Ferreira, M., Saraiva, R. & Rodrigues, E. (2012). [Estado da Arte em Preservação Digital](http://hdl.handle.net/1822/17049). Acedido em fevereiro 2013 em <http://hdl.handle.net/1822/17049>.

ASPECTOS PRÉVIOS A RETER...

A preservação digital define-se como o conjunto de processos responsáveis por garantir o acesso continuado à informação digital durante longos períodos de tempo.

Estes processos que garantem o acesso à informação ocorrem num contexto tecnológico que envolve sempre, até ao momento, *hardware* e/ou *software*. Inerente a esta dependência surgem as grandes questões da preservação digital que se relacionam com a rápida obsolescência da tecnologia.

A preservação digital preocupa-se com a capacidade de manter a informação digital acessível, interpretável e autêntica independentemente da plataforma tecnológica que faz a sua leitura. Esta informação, que pode surgir de inúmeras formas como texto, fotografias digitais, diagramas vetoriais, base de dados, vídeo e áudio, jogos, etc., é retratada através de uma sequência de dígitos binários designada por representação digital.

ESTRATÉGIAS DE PRESERVAÇÃO

No geral as estratégias de preservação digital agrupam-se em 3 classes. O refrescamento é transversal a todas as estratégias.

Refrescamento – transferência de informação de um suporte físico de armazenamento para outro mais atual antes da sua deterioração. Deve ser considerado pré-requisito de qualquer estratégia de preservação.

Emulação – utilização de um emulador (*software*) que é capaz de reproduzir o comportamento de uma plataforma de *hardware* e/ou *software* inicialmente incompatível. Existem vários emuladores de plataformas obsoletas.

Migração/conversão – transferência periódica de um formato para outro mantendo a informação compatível com as tecnologias atuais. Ao contrário das estratégias anteriores centradas no suporte tecnológico, as estratégias de migração procuram formatos alternativos para representar o mesmo conteúdo.

Encapsulamento – conservação, juntamente com a informação digital, da metainformação necessária e suficiente que permita no futuro desenvolver conversores, visualizadores ou emuladores.

CERTIFICAÇÃO DE REPOSITÓRIOS DIGITAIS

Um repositório *per se* não garante o acesso continuado à informação digital. É fundamental que existam formas de medir e demonstrar confiabilidade aos seus utilizadores. Alguns exemplos são nomeados a seguir:

TRAC significa *Trustworthy repositories audit & certification: criteria & checklist*, e reúne um conjunto de requisitos que vão desde a gestão organizacional, às infraestruturas de suporte. É uma ferramenta que permite auditar, avaliar e potencialmente certificar repositórios digitais. O documento é apresentado sob a forma de uma *checklist*.

DRAMBORA significa *Digital Repository Audit Method Based on Risk Assessment* e é um documento e uma ferramenta interativa que sintetiza uma metodologia de autoavaliação, tal como o TRAC, no entanto,

mais direcionado para a gestão estratégica e para a organização que o anterior. Esta ferramenta permite aos gestores de repositórios identificar e categorizar as áreas com maior probabilidade de falha e prevenir a sua ocorrência.

DAS significa *Data Seal of Approval* e funciona como um “selo de garantia” emitido por um grupo de especialistas que atesta a capacidade de preservação de dados científicos de qualidade por um repositório. Todo o processo é realizado à distância e inclui 16 requisitos que devem ser cumpridos.

REPOSITÓRIOS DE ACESSO ABERTO E PRESERVAÇÃO DIGITAL

Um repositório *per se* não garante o acesso continuado à informação digital. É fundamental que existam formas de medir e demonstrar confiabilidade aos seus utilizadores.

A preservação da produção científica e a curadoria dos dados científicos depositados em repositórios de acesso aberto permitem a definição de estratégias que garantem a reutilização e fiabilidade dos seus conteúdos.

A preservação da produção científica depositada em repositórios de acesso aberto deve ser assegurada caso se verifiquem as seguintes condições¹:

1. Existir grande quantidade de conteúdo digital
2. Ter reunido os conteúdos ao longo dos anos
3. Saber quem é responsável pelos conteúdos

¹ Relatório final do projeto KeepIt (2011) disponível em: <http://repository.jisc.ac.uk/553/1/finalreport-keepit10.pdf>

4. Compreender como os conteúdos são usados atualmente
5. Compreender como os conteúdos serão usados no futuro
6. Compreender como o tipo de conteúdo a preservar pode mudar ao longo do tempo

Os repositórios de acesso aberto podem executar funções de preservação digital quando munidos de ferramentas adequadas de que fazem parte: os planos de preservação, as políticas para extração de metadados de preservação de ficheiros e as arquiteturas modulares. De seguida exemplificam-se algumas dessas ferramentas desenvolvidas para três *softwares* de repositórios:

EPrints inclui características de preservação tais como o *History module*: cada objeto apresenta um *log* das alterações por que passou no sistema; METS e DIDL *export plugins* permitem que objetos compostos por mais do que um ficheiro possam ser exportados como um pacote e *Creative Commons licensing* inclui uma opção para permissão de ações de preservação. Outras ferramentas foram desenvolvidas pelo projeto KeepIt: instalações fiáveis para armazenamento a longo prazo, ferramentas de classificação de ficheiros e formatos, ferramentas de análise de risco e ferramentas para migração de formatos.

DSpace realiza preservação *bit-level* em todos os objetos depositados; faz distinção entre formatos conhecidos e desconhecidos; os formatos conhecidos são fechados e a preservação é assegurada, nos desconhecidos não é garantida a preservação.

Fedora Commons inclui o recurso *Object versioning* que suporta versões de fluxos de dados e fornecedores de serviços e preserva a ligação entre ambas as versões. O *Format characterisation* valida os formatos dos ficheiros e extrai metadados a partir de fluxos de dados utilizando JHOVE.

CONCLUSÕES E RECOMENDAÇÕES

A preservação digital, no âmbito dos repositórios de acesso aberto, tem evoluído significativamente nos últimos anos. Ao nível internacional têm vindo a registar-se múltiplas atividades, iniciativas e projetos. Variam entre projetos de menor dimensão promovidos por grupos de investigação até projetos de grande dimensão internacional.

Existem em Portugal linhas orientadoras que facilitam o processo de preservação e curadoria digital nos repositórios de acesso aberto.

Em Portugal, apesar de não existirem repositórios com políticas e estratégias consolidadas, estão definidas linhas de orientação que facilitam o processo de preservação e curadoria digital nos repositórios de acesso aberto:

1. Constituir, na comunidade RCAAP, um grupo de interesse em preservação e curadoria digital.
2. Realizar um recenseamento e caracterização da situação existente nos repositórios de acesso aberto em Portugal.
3. Avaliar, definir e concretizar um projeto-piloto.
4. Cooperar com iniciativas, serviços e projetos relevantes em curso.
5. Disseminar documentos de divulgação, formação e suporte.
6. Realizar ações de sensibilização, divulgação e formação.
7. Avaliar e identificar os recursos e os custos envolvidos nos processos de preservação digital.
8. Identificar eventuais condicionalismos éticos e legais.
9. Sensibilizar, incentivar e apoiar os autores e os gestores dos repositórios para a utilização de formatos que facilitem a preservação.

PROJETOS E ARQUITETURAS DE PRESERVAÇÃO PARA REPOSITÓRIOS

CASPAR

Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval desenvolveu uma arquitetura e um fluxo de trabalho para preservação. <http://www.casparpreserves.eu>

CRIB³¹

Conversion and Recommendation of Digital Object Formats é uma arquitetura de serviço orientado (SOA), desenvolvida na Universidade do Minho, concebida para ajudar instituições com património cultural a implementar soluções de preservação baseadas em migração de formatos. <http://crib.dsi.uminho.pt>

PANIC

Preservation web services Architecture for New media, Interactive Collections and scientific data desenvolveu uma arquitetura que pudesse ser adicionada a repositórios e que lhes permitisse suportar a preservação de objetos digitais. <http://www.itee.uq.edu.au/~eresearch/projects/panic>

PLANETS

Preservation and Long-term Access through NETworked Services construiu serviços práticos de curadoria e ferramentas para bens culturais e científicos. <http://www.planets-project.eu>

PRESERV e KeepIt

Desenvolveu um serviço de ingestão OAIS para repositórios EPrints. <http://preserv.eprints.org>
<http://ie-repository.jisc.ac.uk/553/1/finalreport-keepit10.pdf>

RepoMMan and REMAP

Ambos incorporaram um repositório institucional nos fluxos de trabalho de académicos e administrativos. <http://www.hull.ac.uk/esig/repomman> e <http://www.hull.ac.uk/remap>

Seamless Flow

Criou uma infraestrutura ativa de preservação de registos eletrónicos governamentais.

SHAMAN

Sustaining Heritage Access through Multivalent Archiving estabeleceu as bases conceituais e técnicas para uma infraestrutura de preservação digital distribuída em rede. <http://shaman-ip.eu>

SHERPA DP

Modelo de prestação de serviços no domínio da preservação em repositórios. <http://www.sherpa.ac.uk/projects/sherpadp.html>

SCAPE

Sistemas de preservação digital escaláveis. Portugal está representado pela KEEP SOLUTIONS. <http://www.scape-project.eu/>

FERRAMENTAS E ESTRATÉGIAS PARA A PRESERVAÇÃO EM REPOSITÓRIOS

Modelo de referência OAIS

Open Archival Information System define o modelo conceptual que identifica os componentes funcionais, as interfaces internas e externas e os objetos de informação trocados no seu interior. O modelo deu origem à ISO Standard 14721:2003.

AONS II

Automated Obsolescence Notification System II permite ao gestor de repositório efetuar avaliações de risco aos formatos de ficheiros contidos no seu repositório. <http://apsr.anu.edu.au/aons2/index.htm>

Plato

Utiliza código aberto que suporta e automatiza o processo de especificação de requisitos, avaliando as possíveis soluções e constrói um plano para a preservação de um determinado conjunto de objetos digitais. <http://www.ifs.tuwien.ac.at/dp/plato>



Informação compilada e produzida pelos
Serviços de Documentação
da Universidade do Minho (2013).

