

ESTADO DA ARTE EM PRESERVAÇÃO DIGITAL

FEVEREIRO DE 12



VERSÃO

Autores: Miguel Ferreira, Ricardo Saraiva, Eloy Rodrigues

Versão: 1.0

Contribuição: Clara Boavida, José Carvalho

Data de Criação: 2011-09-28

Última Atualização: 8 de Fevereiro de 2012

ÍNDICE

VERSÃO	2
ÍNDICE	3
SUMÁRIO EXECUTIVO	4
INTRODUÇÃO	7
PRESERVAÇÃO DIGITAL	9
Exemplos de obsolescência tecnológica.....	10
Modelo de referência OAIS.....	12
Estratégias de preservação.....	15
Refreshamento.....	15
Emulação.....	16
Migração/conversão.....	19
Encapsulamento.....	20
Confiança e Certificação de repositórios digitais.....	20
Trustworthy repositories audit & certification (TRAC).....	21
Digital Repository Audit Method Based on Risk Assessment (DRAMBORA).....	23
Data Seal of Approval (DSA).....	24
REPOSITÓRIOS DE ACESSO ABERTO E PRESERVAÇÃO DIGITAL	26
Softwares/Plataformas de repositórios.....	30
Projetos e arquiteturas de preservação para repositórios.....	32
Ferramentas e estratégias para a preservação em repositórios.....	41
CONCLUSÕES E RECOMENDAÇÕES	44
ANEXOS	47
Anexo 1: <i>Checklist</i> do TRAC.....	47
Anexo 2: Requisitos do Data Seal of Approval.....	55
BIBLIOGRAFIA	57

SUMÁRIO EXECUTIVO

O presente estudo está inscrito no plano de atividades de 2011 do projeto Repositório Científico de Acesso Aberto de Portugal (RCAAP) e assinala o início do interesse do RCAAP no domínio da preservação digital na esfera dos repositórios de acesso aberto. Para além de recolher informação atualizada sobre o tema e as iniciativas mais relevantes relacionadas com a preservação digital, o objetivo deste documento é também o de informar e orientar futuras atividades e desenvolvimentos relacionados com a preservação digital no âmbito do RCAAP em anos vindouros.

Na sua introdução, o documento começa por contextualizar a crescente importância e interesse do tema da preservação digital na agenda dos repositórios de acesso aberto, e que tem sido objeto de debate nos últimos anos. Independentemente das diversas opiniões quanto à sua centralidade na atividade dos repositórios parece claro que a preservação digital será uma preocupação crescente dos repositórios nos próximos anos, a nível internacional e também em Portugal.

Na segunda secção do estudo intitulada: *"Preservação Digital"* é apresentada uma panorâmica geral do tema, definindo os conceitos mais relevantes e apresentando as principais técnicas preservação digital utilizadas na atualidade. Na terceira secção designada: *"Repositórios de Acesso Aberto e preservação digital"*, o relatório procura contextualizar a preservação digital no âmbito dos repositórios de acesso aberto e identificar os projetos, as arquiteturas e as estratégias mais relevantes neste domínio.

Nas conclusões, que constituem a última secção do documento, constata-se que a preservação digital, no âmbito dos repositórios de acesso aberto, tem evoluído significativamente, como se comprova pelas múltiplas atividades, iniciativas e projetos, que se têm vindo a conhecer nos últimos anos. No entanto, apesar da crescente consciencialização e interesse no que concerne às questões relacionadas com a preservação digital, o número de repositórios com políticas, estratégias e ações consolidadas ainda é residual.

Também em Portugal, dos 35 repositórios atualmente registados no portal RCAAP, nenhum destes repositórios possuirá uma política de preservação formal. Haverá instituições com procedimentos no que concerne aos formatos admissíveis ou que realizam, como normativo interno, migrações de formatos aquando do depósito de documentos, mas ainda sem uma sistematização desejável.

O estudo termina com a exposição de um conjunto de ações e linhas de orientação que poderão integradas no projeto RCAAP, ou desenvolvidas por instituições que nele participam, com o intuito de promover e facilitar o processo de preservação e curadoria digital nos repositórios de acesso aberto em Portugal.

Recomendações e linhas de orientação:

1. Constituir, no âmbito da comunidade RCAAP, um grupo de interesse no domínio da preservação e curadoria digital, com iniciativas, atividades e canais de comunicação próprios, e no quadro do qual se poderiam concretizar a generalidade das recomendações e linhas de orientação apresentadas em seguida;
2. Realizar um recenseamento e caracterização da situação existente nos repositórios de acesso aberto portugueses, no que diz respeito às políticas, procedimentos e estratégias de preservação digital;
3. Avaliar, definir e concretizar um projeto piloto, no domínio da preservação digital, com a participação de vários repositórios portugueses, com o recurso a arquitetura(s) que possa(m) dotar os repositórios participantes de ferramentas abrangentes em termos de preservação digital;
4. Acompanhar e, se possível e adequado, cooperar com iniciativas, serviços e projetos relevantes em curso na área da preservação digital, com o intuito de conhecer, utilizar e promover as boas práticas neste domínio. Considerando a situação portuguesa, deverá ser avaliada desde logo a exequibilidade e os termos de uma possível cooperação entre o projeto RCAAP e/ou os repositórios individualmente com o RODA – Repositório de Objetos Digitais Autênticos;

5. Desenvolver e/ou disseminar documentos de divulgação, formação e suporte, como Briefing papers, modelos de políticas e procedimentos, boas práticas e casos exemplares de preservação digital;
6. Realizar ações de sensibilização, divulgação e formação destinadas a gestores de repositórios e responsáveis institucionais;
7. Avaliar e identificar os recursos e os custos envolvidos nos processos de preservação digital, em especial os relativos aos recursos humanos, quer quanto ao esforço (tempo de trabalho) envolvido, quer quanto às competências requeridas, e eventuais necessidades de formação;
8. Identificar eventuais condicionalismos éticos e legais, no que concerne a ações de preservação que possam incidir em conteúdos já depositados;
9. Sensibilizar, incentivar e apoiar os autores e os gestores dos repositórios para a utilização de formatos de ficheiro que facilitem a preservação a longo prazo, aquando do depósito de documentos no(s) repositório(s).

INTRODUÇÃO

O presente documento foi realizado no âmbito do projeto Repositório Científico de Acesso Aberto de Portugal (RCAAP)¹. A iniciativa RCAAP visa aumentar a visibilidade, acessibilidade e difusão dos resultados da atividade académica e de investigação científica nacional, facilitar o acesso à informação sobre a produção científica nacional em regime de acesso aberto, bem como integrar Portugal num conjunto de iniciativas internacionais neste domínio.

Até ao final de 2009 as atividades do projeto RCAAP focaram-se exclusivamente na criação e desenvolvimento de repositórios de literatura científica, mas desde então o projecto RCAAP alargou o âmbito da sua atividade e serviços aos repositórios de dados científicos (2010) e às revistas de acesso aberto (2011). O plano de trabalho para 2011 assinala também o início do interesse do RCAAP no domínio da preservação digital no âmbito dos repositórios de acesso aberto.

O presente relatório de estado da arte é o único resultado previsto do projeto RCAAP neste domínio em 2011. Para além de recolher informação atualizada sobre o tema e identificar as iniciativas e projetos mais relevantes relacionadas com preservação digital em repositórios de acesso aberto, este documento poderá também informar e orientar futuras atividades e desenvolvimentos relacionados com a preservação digital no âmbito do RCAAP em anos futuros.

A importância e prioridade da preservação digital nos repositórios de acesso aberto têm sido objeto de debate nos últimos anos. As opiniões variam entre os que, num extremo, defendem uma grande prioridade, atenção e investimento às questões da preservação (o que poderia contribuir para tornar os repositórios mais confiáveis e assim ajudar a obter mais conteúdos), até aos que, no outro extremo, defendem que a prioridade tem que

¹ O **projecto RCAAP** é uma iniciativa da UMIC – Agência para a Sociedade do Conhecimento, IP concretizada pela FCCN – Fundação para a Computação Científica Nacional, disponibilizando mais um serviço avançado sobre a Rede Ciência, Tecnologia e Sociedade (RCTS) gerida pela FCCN. A execução do projecto conta ainda com a participação científica e técnica da Universidade do Minho. Toda a informação do projecto pode ser consultada em: <http://projecto.rcaap.pt/>.

continuar a ser aumentado os conteúdos dos repositórios (porque não vale a pena preservar repositórios vazios) e recordam que, relativamente aos artigos científicos, os repositórios são apenas uma via alternativa de acesso, pelo que a responsabilidade principal pela preservação digital tem obrigatoriamente de ser assumida pelos editores das revistas científicas que publicam esses artigos.

No entanto, independentemente do ponto onde nos situemos nesse debate, considerando que existe um número crescente de repositórios institucionais que já recolhe uma percentagem relevante da produção científica dos seus membros, e que uma parte desses conteúdos não é objeto de publicação externa, constituindo os repositórios o local original e principal (e em alguns casos único) de publicação, parece claro que a preservação digital deverá ser uma preocupação crescente dos repositórios nos próximos anos. E isto é verdade para os repositórios em geral, e para os repositórios portugueses em particular.

Para além desta introdução, o relatório está estruturado em mais 2 secções e dois anexos. A secção 2 apresenta uma panorâmica geral da problemática preservação digital, definindo os conceitos mais relevantes e apresentando as principais técnicas utilizadas. A Secção 3 procura contextualizar a preservação digital no âmbito dos repositórios de acesso aberto, e identificar os projetos e iniciativas mais relevantes neste domínio. Finalmente, a secção 3 apresenta algumas conclusões e sugestões de ação para os repositórios portugueses e para o projeto RCAAP. O relatório é complementado ainda por duas listas de verificação e requisitos para a preservação digital.

PRESERVAÇÃO DIGITAL

Apesar das inúmeras vantagens que decorrem da utilização de informação digital, é importante realçar que esta é acompanhada de um problema estrutural que coloca em risco a sua longevidade. A documentação em formatos digitais, embora possa ser copiada infinitas vezes sem perder qualidade, não pode ser consumida senão na presença de um contexto tecnológico, *hardware* e/ou *software*, que capacite o seu destinatário ou potencial interessado (não humano) de a interpretar de forma inteligível. Esta dependência tecnológica torna-o particularmente vulnerável à rápida obsolescência a que a tecnologia está sujeita (Chen, 2001).

Designa-se, assim, por preservação digital o conjunto de processos responsáveis por garantir o acesso continuado à informação digital durante longos períodos de tempo, i.e. períodos de tempo superiores à esperança de vida do ambiente tecnológico necessário à interpretação e/ou reprodução dessa informação (Webb, 2003). Assim, a preservação digital preocupa-se com a capacidade de manter a informação digital acessível, interpretável e autêntica, mesmo na presença de uma plataforma tecnológica diferente daquela inicialmente utilizada no momento da sua criação.

Neste contexto, designa-se por representação digital (ou objeto digital) todo e qualquer objeto de informação que possa ser retratado através de uma sequência de dígitos binários² (Thibodeau, 2002). Documentos de texto, fotografias digitais, diagramas vetoriais, bases de dados, sequências de vídeo e áudio, modelos tridimensionais, páginas Web, jogos e/ou aplicações de *software* são apenas alguns exemplos do que podemos considerar uma representação digital.

No mundo atual, onde cada vez mais organizações e pessoas dependem da informação digital que produzem, torna-se premente a implementação de técnicas e de políticas

² Esta definição é suficientemente lata para acomodar tanto, informação que nasceu num contexto tecnológico digital (objetos nado-digitais), como informação digital obtida a partir de suportes analógicos (objetos digitalizados).

concertadas que vão no sentido de garantir a perenidade e a acessibilidade a este tipo de informação.

Um domínio onde a preservação digital tem particular relevância é a saúde. Com a adoção massificada de ferramentas digitais por parte de médicos e hospitais para gestão de processos clínicos, gestão hospitalar ou meios de diagnósticos, os pacientes estão sujeitos à capacidade destes organismos garantirem o acesso continuado aos seus registos durante toda a sua vida como pacientes. Em alguns hospitais, os registos clínicos são de retenção obrigatória por um período de 25 anos. Isto inclui também todos os meios de diagnóstico, como radiografias, TAC, ECGs ou ecografias, todos eles produzidos e conservados atualmente em formatos digitais.

Outro organismo responsável por manter informação digital durante longos períodos de tempo é a Segurança Social. Este organismo tem como obrigação legal conservar os registos das contribuições efetuadas à tutela por períodos de tempo que poderão ascender às centenas de anos. Geralmente estes registos são conservados até 10 anos após a morte do contribuinte³. Contudo, durante períodos de tempo tão alargados, é comum haver evoluções tecnológicas ao nível do hardware, redes, arquiteturas de software, sistemas de gestão, esquemas e necessidades metadados, e até alterações à tutela responsável pela conservação destes registos. Todos estes eventos apresentam riscos à preservação deste tipo de informação, sendo o seu impacto a nível social e pessoal extremamente elevado.

Exemplos de obsolescência tecnológica

O curso da história tem revelado inúmeros exemplos de obsolescência tecnológica. Um dos casos mais relevantes nas últimas décadas foi o ocorrido em torno dos formatos de vídeo. Como é conhecido, o formato de vídeo Betamax, desenvolvido pela Sony, que no início dos anos 80 dominava cerca de um terço do mercado de vídeo doméstico (IEEE History Center; Nayak & Ketteringham, 1994; Shiraishi, 1985), apesar de oferecer uma

³ **Diário da República** – I Série-B N.º 21—30 de Janeiro de 2006 acessível em: <http://dre.pt/pdf1sdip/2006/01/021B00/06590725.pdf>

qualidade de imagem superior e cassetes de menores dimensões, foi completamente abandonado pelos consumidores, que adoptaram o formato VHS⁴, e no início dos anos 90 era já muito difícil encontrar um dispositivo capaz de apresentar o conteúdo armazenado nas cassetes Betamax. (Nayak & Ketteringham, 1994). E o mesmo se passou nesta última década com o formato VHS, substituído pelos suportes e equipamentos digitais.

Outro exemplo de obsolescência tecnológica, desta vez no domínio digital, reporta-se ao uso das populares disquetes de 3.5 polegadas (Figura 1). Em Março de 2003, o fabricante *Dell Computer Corporation* anunciou que os seus computadores deixariam de integrar dispositivos de leitura compatíveis com este tipo de suportes. Vários fabricantes seguiram de imediato o seu exemplo (Kenney, McGovern, Entlich, Kehoe, & Olsen, 2003).



Figura 1 - Disquete de 3.5 polegadas.

As disquetes acabaram por ser substituídas por CD/DVD graváveis e flash-drives. Porém, mesmo estes encontram-se em vias de extinção. Uma grande parte dos computadores portáteis vendidos atualmente não incorpora leitores de CD/DVD⁵. Os computadores e restantes dispositivos móveis de acesso à informação estão cada vez mais ligados à Internet, em todo o lado e a qualquer momento, o que propicia uma mudança de paradigma que passa pela eliminação dos suportes digitais como forma de troca de informação. Em breve o modo dominante para armazenar e trocar informação será através da "cloud", ou seja, recorrendo a servidores de armazenamento geridos por terceiros e acessíveis através da Internet.

É importante salientar que a obsolescência tecnológica não se manifesta apenas ao nível dos suportes físicos. Toda a informação digital tem necessariamente de respeitar as

⁴ Video Home System.

⁵ Mais informação em: <http://www.pcmag.com/article2/0,2817,2396031,00.asp>

regras lógicas de um formato. Isto permite às aplicações de *software* abrir e processar adequadamente a informação armazenada. À medida que o *software* vai evoluindo, também os formatos por ele suportados vão sendo alvo de atualização. É bastante comum encontrar aplicações de *software* capazes de carregar os ficheiros produzidos por versões anteriores da mesma aplicação. No entanto, essa capacidade raramente vai para além das duas versões precedentes (Kenney et al., 2003).

Modelo de referência OAIS

Em 1990, o Consultative Comitee for Space Data Systems (CCSDS)⁶ iniciou um esforço conjunto com a International Organization for Standardization (ISO)⁷ com o objetivo de desenvolver um conjunto de normas capazes de regular o armazenamento a longo-prazo de informação digital produzida no âmbito de missões espaciais.

Deste esforço nasceu o modelo de referência OAIS (Open Archival Information System), um modelo conceptual que visa identificar os componentes funcionais que deverão fazer parte de um sistema de informação dedicado à preservação digital, bem como as suas interfaces internas e externas e os objetos de informação trocados no seu interior (Consultative Committee for Space Data Systems, 2002; Lavoie, 2004). O modelo foi aprovado como uma norma internacional ISO em 2003 – ISO Standard 14721:2003.

Um dos contributos mais relevantes a destacar desta iniciativa foi a definição de uma terminologia própria que viria a facilitar a comunicação entre os diversos intervenientes envolvidos em processos de preservação digital (Saramago, 2004). A figura 2 ilustra os diferentes componentes funcionais identificados no modelo de referência OAIS, bem como os pacotes de informação trocados no interior do sistema.

⁶ Sítio **Consultative Comitee for Space Data Systems** acessível em: <http://public.ccsds.org/about/default.aspx>

⁷ Sítio **International Organization for Standardization** acessível em: <http://www.iso.org/iso/home.html>

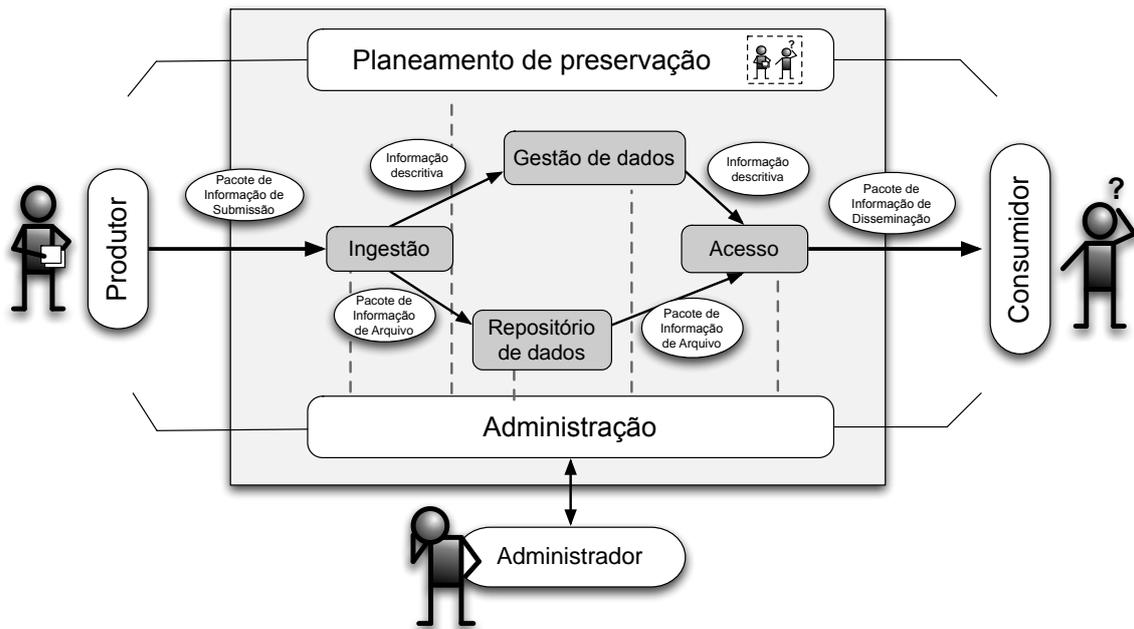


Figura 2 - Modelo de referência Open Archival Information System (Fonte: OAIS).

O **Produtor** é a entidade externa ao repositório responsável por submeter material ao sistema. Este material é aqui representado pelo Submission Information Package (SIP).

Durante o processo de submissão, designado neste contexto por **Ingestão**, o sistema é responsável por verificar a integridade física, lógica e semântica da informação recebida. Nesta fase, é validada toda a **Informação Descritiva** que acompanha a representação digital e que irá suportar a descoberta e localização do material arquivado. Caso esta informação não seja submetida pelo Produtor, deverá ser gerada no interior do sistema. Ainda no contexto da ingestão, são efetuadas todas as operações necessárias para transformar um SIP num Archival Information Package (AIP), i.e. numa estrutura lógica capaz de unificar todos os constituintes da representação digital. É esta estrutura que será alvo de preservação por parte do sistema OAIS.

A **Informação Descritiva**, vulgarmente designada por metainformação, pode ser fornecida pelo produtor ou gerada no interior do sistema. Esta informação é armazenada e gerida pela unidade funcional designada **Gestão de Dados**⁸. Esta unidade deverá, para além de conservar a informação descritiva, permitir estabelecer relações entre a metainformação descritiva e o material preservado (i.e., AIP), assegurar a localização do

⁸ Do inglês *Data Management*.

material e permitir obter relatórios e estatísticas sobre os conteúdos do repositório. Por sua vez, o material a preservar (i.e., o AIP) é armazenado no **Repositório de Dados**⁹. Para além de guardar as representações digitais, esta unidade funcional é responsável por gerir toda a estrutura de armazenamento, garantir que as representações não são degeneradas por mau funcionamento dos suportes físicos, efetuar verificações de integridade e oferecer funcionalidades de salvaguarda e recuperação de dados em situações de desastre, e.g. RAID, cópias de segurança, replicação remota, etc.

A unidade funcional **Planeamento de Preservação** é responsável pela definição de políticas de preservação e pela elaboração de planos de preservação que atuam sobre o material conservado, mantendo-o acessível e de acordo com os requisitos de qualidade e autenticidade definidos pelo gestor do sistema e exigidos pela sua comunidade de interesse¹⁰. Esta unidade é ainda responsável por monitorizar o ambiente externo ao repositório por forma a detetar alterações relevantes no panorama tecnológico ou nos requisitos dos seus utilizadores que possam influenciar a forma como a informação digital está a ser preservada ou deverá ser disseminada. Este serviço tem como missão notificar o responsável pelo Planeamento quando determinadas variáveis externas ao sistema sofreram alterações que poderão colocar em risco a salvaguarda da informação digital.

É importante referir que as funções associadas a este componente de monitorização são vulgarmente desempenhadas por pessoas especializadas em tecnologia e preservação digital. No entanto, existem iniciativas no sentido de automatizar tanto o processo de monitorização, como o de produção e execução de planos de preservação.

⁹ Do inglês *Archival Storage*.

¹⁰ Também conhecido por população potencialmente utilizadora. É de notar que o conceito de comunidade de interesse deverá ser entendido no seu sentido mais lato. Trata-se de um conceito por vezes associado a centros de documentação e bibliotecas especializadas, como é o caso de certas bibliotecas universitárias (e.g. Biblioteca de Física da Universidade do Minho em que a comunidade de interesse são os estudantes e professores de matérias ligadas à Física). Em bibliotecas de carácter geral, como bibliotecas públicas ou nacionais, e na generalidade dos arquivos este conceito não é aplicável ou apenas o será se se considerar que a comunidade de interesse coincide com a totalidade da população.

A unidade funcional designada de **Acesso** estabelece a ponte entre o sistema e a sua comunidade de interesse, i.e., o conjunto de potenciais **Consumidores** do material custodiado. Esta unidade é responsável por permitir a descoberta e localização das representações digitais, bem como preparar as mesmas para entrega ao consumidor.

A informação é entregue sob a forma de **Dissemination Information Packages** (DIP). É de realçar que os DIP poderão ser diferentes dos AIP, ou seja, a informação entregue ao consumidor poderá ser um subconjunto da informação arquivada ou uma versão transformada da mesma (Consultative Committee for Space Data Systems, 2002; Lavoie, 2004). Do mesmo modo, os AIP poderão ser diferentes da informação originalmente submetida ao sistema (SIP). Os SIP são geralmente transformados para estruturas mais fáceis de preservar, havendo contudo a preocupação de preservar também o objeto original.

Por último, a unidade funcional designada de **Administração** é responsável pela operação diária do sistema. Entre as atividades que são da sua responsabilidade destacam-se: a parametrização do sistema, monitorização dos seus processos, o registo de diários de atividades, etc. Este componente interage com todos os restantes de modo a assegurar o correto funcionamento do sistema (Lavoie, 2004).

Estratégias de preservação

São várias as estratégias que podem ser implementadas no sentido de mitigar o problema da obsolescência tecnológica. Segundo Lee et al. estas estratégias podem ser agrupadas em três classes fundamentais: emulação, migração e encapsulamento (Lee, Slattery, Lu, Tang, & McCrary, 2002).

Existe uma outra atividade designada de refrescamento que é transversal a todas as estratégias de preservação.

Refrescamento

Uma representação digital torna-se persistente no momento em que é inscrita num suporte físico de armazenamento (ex. disquete, disco rígido, CD-ROM, etc.). Devido a

esse facto, garantir a integridade do suporte é fundamental para que a informação nele armazenada possa ser lida e posteriormente interpretada. Se o suporte físico se deteriorar ou se tornar obsoleto a ponto de deixarem de existir periféricos capazes de o ler, então a informação nele armazenada perder-se-á de forma irremediável (Hendley, 1998).

O refrescamento consiste na transferência de informação de um suporte físico de armazenamento, para outro geralmente mais atual, antes que o primeiro se deteriore ou se torne irremediavelmente obsoleto (Bearman, 1989; Hendley, 1998; Task Force on Archiving of Digital Information, Commission on Preservation and Access, & Research Libraries Group, 1996; Woodyard, 1998).

O refrescamento não constitui uma estratégia de preservação *per se*. Em vez disso, deverá ser considerado um pré-requisito para o sucesso de qualquer estratégia de preservação (Besser, 2001). A frequente verificação de integridade dos suportes físicos, assim como o seu refrescamento periódico são atividades vitais num contexto de preservação digital.

Emulação

As estratégias de emulação baseiam-se na utilização de *software* especial, designado vulgarmente por Emulador, que é capaz de reproduzir o comportamento de uma plataforma de hardware e/ou software, inicialmente incompatível (Rothenberg, Commission on Preservation and Access, & Council on Library and Information Resources, 1999). A grande vantagem desta abordagem está na capacidade de preservar, com um elevado grau de fidelidade, as características e as funcionalidades da representação digital pois esta será manipulada recorrendo ao *software* originalmente utilizado na sua criação ou reprodução (Lee et al., 2002).

Existem, fundamentalmente, dois tipos de emuladores: emuladores de sistemas operativos e emuladores de *hardware*. Os primeiros focam-se na reprodução de um sistema operativo permitindo a execução de diversas aplicações no contexto de um único

emulador. Um exemplo deste tipo de emuladores é o Wine¹¹, um emulador que permite executar aplicações desenvolvidas para a plataforma Windows, em ambientes Unix.

O segundo tipo de emuladores visa mimar o comportamento de uma plataforma de *hardware*, admitindo que vários sistemas operativos e correspondentes aplicações possam ser executados sobre o mesmo emulador (Granger, 2000; Thibodeau, 2002). Apesar de mais versáteis, este tipo de emuladores obriga à instalação de um sistema operativo, assim como todas as aplicações necessárias ao correto funcionamento. Exemplos deste tipo de emuladores são: *VMware Workstation* (VMWare, 1998) e o *Parallels Desktop* (Parallels, 1995). Trata-se de sistemas atualmente muito utilizados para virtualizar sistemas, i.e., permitir executar concorrentemente várias máquinas virtuais sobre o mesmo *hardware*. Existem também vários emuladores de plataformas consideradas obsoletas, ex. ZX Spectrum (Davidson & Pollard, 2005), Nintendo NES (Krijgsman, 2005), entre outras.

Hendley considera que a emulação apenas deveria ser utilizada em contextos onde a comunidade de interesse valoriza a preservação do ambiente tecnológico original ou ainda em situações em que as representações digitais não são passíveis de ser convertidas para formatos mais atuais e compatíveis com os sistemas utilizados pela sua comunidade de interesse (Hendley, 1998). Outros autores consideram potencialmente arriscado confiar no *software* original como forma de preservar informação digital, uma vez que este pode ser alvo de vírus ou portador de *bugs* que poderão, no futuro, resultar em perdas substanciais de informação (Thibodeau, 2002; Waugh, Wilkinson, Hills, & Dell'oro, 2000). Isto poderá acontecer pois está a preservar-se *software* que já não é suportado por nenhum organismo.

¹¹ Sítio emulador **Wine** acessível em: <http://www.winehq.org/>



Figura 3 - Exemplo de um cenário de emulação.

O uso de emuladores parte do pressuposto de que os utilizadores do futuro serão capazes de operar adequadamente aplicações e sistemas operativos há muito desaparecidos. Por exemplo, assume-se que no futuro os utilizadores possuam a capacidade de operar o sistema MS-DOS (Microsoft Corporation, 1981), pois será um requisito necessário durante o acesso a uma determinada representação digital.

A figura 3 apresenta um cenário de emulação onde um jogo de computador está a ser executado por um emulador de ZX Spectrum, que por sua vez está a ser executado por um emulador de Windows sobre o sistema Mac OS X.

As estratégias de emulação assumem um papel preponderante na preservação de representações digitais com características dinâmicas e/ou interativas (Woodyard, 2000). Este tipo de estratégias é particularmente relevante em contextos em que o objeto a preservar se trata de uma aplicação de *software*. Tal como acontece atualmente com um número crescente de jogos de computador considerados de valor histórico assinalável. A própria British Library está a dar início a um programa de arquivo e preservação de videojogos desenvolvidos no Reino Unido (Crookes, 2010).

Migração/conversão

A Migração ou Conversão consiste na “(...) *transferência periódica de material digital de uma dada configuração de hardware/software para uma outra, ou de uma geração de tecnologia para outra subsequente*” (Task Force on Archiving of Digital Information et al., 1996).

As representações digitais são constituídas de informação estruturada numa dada forma, i.e. formato. O formato de um objeto digital define a estrutura pela qual os elementos de informação se encontram organizados. Neste contexto, a migração pode ser vista como o processo responsável pela reorganização dos elementos de informação que constituem uma representação segundo uma nova estrutura (Lawrence, Kehoe, Rieger, Walters, & Kenney, 2000).

Ao contrário das estratégias de preservação já apresentadas, mais focadas na cristalização do objeto digital no seu formato original, as estratégias baseadas em migração centram-se na procura de formatos alternativos para representar o mesmo conteúdo intelectual que se pretende preservar. Trata-se de estratégias orientadas à preservação do objeto conceptual, desvalorizando a preservação do objeto lógico e/ou físico original (Russell, 2000).

A migração tem como principal objetivo garantir que a informação digital permanece compatível com as tecnologias atuais. Um leitor comum é, assim, capaz de consumir essa informação sem necessidade de recorrer a artefactos pouco convencionais como emuladores. Porém, os processos de migração acarretam alguns problemas que deverão ser considerados. Ao efetuar uma migração de formatos, ou uma migração de dados entre sistemas, existe uma grande probabilidade de algumas das propriedades que fazem parte da informação original não serem corretamente transferidas para o formato ou sistema de destino adotado (Hedstrom, 2001; Heslop, Davis, & Wilson, 2002). Isto deve-se, sobretudo, a incompatibilidades estruturais entre os formatos de origem e destino ou à utilização de conversores de baixa qualidade (Ferreira, Baptista, & Ramalho, 2006; Lawrence et al., 2000; Rauber & Aschenbrenner, 2001).

Adicionalmente, não é espectável que uma estratégia de migração possa resolver permanentemente os problemas de preservação. O formato de destino encontra-se, também este, sob constante ameaça de obsolescência, o que significa que será apenas uma questão de tempo até uma nova migração ter de ser ministrada. Não obstante, a migração é de longe a estratégia de preservação mais aplicada, tanto em contextos institucionais, como no domínio doméstico (Lee et al., 2002).

Encapsulamento

Por vezes não é fácil determinar o valor intrínseco de determinados objetos digitais. Poderão passar-se muitos anos até que a comunidade de consumidores revele um particular interesse por uma determinada coleção (Heminger & Robertson, 2004). Esse interesse pode até nunca surgir. Neste tipo de cenários, estratégias de preservação que carecem de uma diligência contínua (e.g. migração) poderão revelar-se demasiado onerosas. As soluções baseadas em encapsulamento procuram resolver este problema, mantendo os objetos digitais inalterados até ao momento em que se tornam efetivamente necessários.

A estratégia de encapsulamento consiste em conservar, juntamente com a informação digital, toda a metainformação necessária e suficiente que permita futuramente desenvolver conversores, visualizadores ou emuladores. Esta informação poderá consistir, por exemplo, numa descrição formal e detalhada do formato preservado e de toda a metainformação técnica relevante que auxilie o técnico a encontrar a melhor plataforma para ler a respetiva informação (Digital Preservation Testbed, 2001). Pode inclusive arquivar-se junto da informação, uma cópia de toda a plataforma tecnológica necessária à sua leitura. Essa é aliás a estratégia adotada pela Biblioteca Nacional da Holanda (Koninklijke Bibliotheek) e por praticamente todas as iniciativas de arquivo da Web atualmente existentes.

Confiança e Certificação de repositórios digitais

Seja qual for a plataforma de suporte ou as estratégias de preservação adotadas, um repositório deve corresponder às expectativas criadas pelos seus utilizadores. Afirmar que

se é capaz de garantir o acesso continuado à informação digital não é suficiente para estabelecer um clima de confiança junto dos vários intervenientes que interagem com o repositório, i.e. produtores, consumidores, operadores do sistema, gestores, entidades de fomento, ou outros (RLG, NARA, & OCLC, 2007; Rodrigues, 2003). Para que um repositório seja verdadeiramente confiável, é fundamental que existam formas de medir e demonstrar essa confiabilidade (RLG et al., 2007).

Trustworthy repositories audit & certification (TRAC)

Em 2003, a RLG (Research Library Group) e a NARA (National Archives and Records Administration) constituíram um grupo de trabalho para analisar questões relacionadas com a confiabilidade e a certificação de repositórios digitais. O objetivo deste grupo de trabalho era estabelecer um conjunto de critérios que permitissem identificar os repositórios que eram capazes de armazenar e fornecer acesso continuado a coleções de material digital. O desafio consistia sobretudo em reunir um conjunto de atributos mensuráveis e delinear um caminho que conduzisse à certificação de repositórios digitais, quer estes se tratassem de arquivos digitais sob a tutela de pequenas instituições académicas ou grandes repositórios de albergados por arquivos ou bibliotecas de âmbito nacional (RLG et al., 2007).

Em 2007 foi publicado o documento *Trustworthy repositories audit & certification: Criteria & Checklist* (RLG et al., 2007), vulgarmente conhecido por TRAC, que reúne um conjunto de requisitos que vão desde a gestão organizacional, às infraestruturas de suporte, e que são considerados vitais no estabelecimento de um clima de confiança em torno de um repositório digital. O documento TRAC foi elaborado com os seguintes objetivos em mente:

- Fornecer uma ferramenta que permita auditar, avaliar e potencialmente certificar repositórios digitais;
- Estabelecer a documentação necessária para a realização de uma auditoria;
- Delinear o processo de certificação;
- Estabelecer metodologias apropriadas para determinar a robustez e a sustentabilidade de um repositório digital.

O documento começa por definir “confiança” no contexto de um repositório digital e identifica os vários elementos que devem ser avaliados ou verificados na prossecução desse objetivo. O documento é apresentado sob a forma de uma *checklist* que serve, numa primeira instância, como ferramenta de autodiagnóstico e, posteriormente, de autoavaliação. A ferramenta permite identificar potenciais omissões e pontos de falha nos sistemas e organizações responsáveis por preservar informação digital (RLG et al., 2007).

A organização responsável por um repositório confiável deverá ser capaz de identificar riscos e prevenir ameaças. Estes poderão ser de várias naturezas (e.g. organizacional, tecnológica, social, ao nível da segurança, etc.), pelo que a monitorização constante do meio ambiente, o planeamento e a manutenção do sistema deverão ser atividades diárias de quem administra o repositório (RLG et al., 2007).

A aplicação do TRAC potencia a confiança junto dos utilizadores do repositório pois estabelece um clima de maior transparência relativamente aos processos implementados pelo repositório e em torno do repositório. Uma auditoria por parte de uma equipa externa de especialistas aumenta a confiança pois evidencia um elevado nível de certeza quanto à utilização de práticas adequadas na operação do repositório e no tratamento da informação que lhe foi confiada (RLG et al., 2007).

A totalidade dos requisitos do TRAC encontram-se enumerados no anexo 1 intitulado: *Anexo 1: Checklist do TRAC.*

Atualmente o TRAC encontra-se num processo de discussão por parte do comité técnico da International Standards Organization (ISO). O objetivo é elevar o TRAC a norma internacional gerida pela ISO. A data prevista para a publicação do documento final sob a forma de norma é Junho de 2012.

Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)

Desenvolvido pelo Digital Curation Centre (DCC)¹² e pelo DigitalPreservationEurope (DPE)¹³ o DRAMBORA nasce da experiência acumulada que resultou de um conjunto de auditorias realizadas ao longo dos anos de 2006 e 2007. As auditorias foram realizadas pelo DCC tendo por base uma versão preliminar do TRAC (Digital Curation Centre & DigitalPreservationEurope, 2007).

O DRAMBORA é um documento e uma ferramenta interativa que sintetiza uma metodologia de autoavaliação, tal como o TRAC, porém foca-se mais em aspetos ligados à gestão estratégica e à organização, e menos em aspetos técnicos relacionados com o repositório e respetiva plataforma tecnológica. O DRAMBORA convida os administradores de repositórios digitais a:

- Elaborar um perfil organizacional, descrevendo e documentando a sua política de depósito, objetivos, responsabilidades, atividades e material custodiado;
- Identificar e avaliar os riscos que poderão impedir a prossecução da sua missão e que ameaçam a salvaguarda dos seus materiais;
- Gerir eficazmente os riscos, mitigando a sua probabilidade de ocorrência;
- Estabelecer planos de contingência eficazes para minimizar os efeitos provocados por riscos que não puderam ser evitados.

O DRAMBORA apresenta, assim, uma abordagem mais focada na identificação e gestão de riscos com o objetivo de racionalizar as incertezas e prevenir ameaças. A utilização da ferramenta é realizada em 6 etapas:

1. Identificação do contexto organizacional;
2. Documentação da política e procedimentos de gestão;
3. Identificação de atividades, recursos e os respetivos responsáveis;
4. Identificação de riscos;

¹² Sítio **Digital Curation Centre** acessível em: <http://www.dcc.ac.uk/>

¹³ Sítio **DigitalPreservationEurope** acessível em: <http://www.digitalpreservationeurope.eu/>

5. Análise e avaliação dos riscos;
6. Gestão de riscos.

As fases iniciais conduzem o administrador/auditor a elaborar um perfil da organização descrevendo e documentando todas as políticas e procedimentos em vigor, bem como as atividades e os objetivos da organização (Digital Curation Centre & DigitalPreservationEurope, 2007). De seguida, são identificados os riscos inerentes a cada elemento que compõe a organização e que poderá ter influência na salvaguarda da informação digital, havendo de seguida uma atividade de gestão desses riscos através da elaboração de planos de contingência que deverão ser ativados caso algum dos riscos se venha a materializar.

O processo permite aos administradores de repositórios identificar e categorizar as áreas onde existe maior probabilidade de falha e prevenir a sua ocorrência (Digital Curation Centre & DigitalPreservationEurope, 2007).

Data Seal of Approval (DSA)

O Data Seal of Approval (DSA)¹⁴ é um “selo de garantia” emitido por um grupo de especialistas (i.e. *DSA board*) que atesta se um repositório é capaz de preservar com qualidade dados científicos¹⁵ para futura referência e processamento, sem que isto acarrete elevados custos ou investimentos para as entidades que os custodiam (DSA Board, n d).

Trata-se, portanto, de um conjunto de boas-práticas que se pretendem que sejam seguidas por organizações responsáveis pela preservação de dados científicos.

Os 16 requisitos que compõem o Data Seal of Approval foram publicados em 2009. O processo de obtenção do “selo” de qualidade não requer a visita de auditores externos ao local do repositório. Por sua vez, o candidato deverá reunir evidências e documentação relevante que demonstre o cumprimento de todos os requisitos

¹⁴ Sítio do **Data Seal of Approval** acessível em: <http://www.datasealofapproval.org>

¹⁵ research datasets.

necessários à obtenção do "selo". Após a reunião de todas as evidências, estas são submetidas através de um formulário em linha e dá-se início ao processo de análise e revisão pelos membros do DSA. O repositório nunca é visitado pelos auditores. Todo o processo é realizado à distância (Joy Davidson, 2011).

Dos 16 requisitos, 3 dizem respeito aos produtores e ao processo de ingestão, 10 à qualidade do repositório e 3 ao acesso à informação por parte dos consumidores. O cumprimento de cada requisito é avaliado numa escala de 0-4. Não obstante, um repositório para se tornar certificado não necessita de obter uma pontuação de 4 em todos os requisitos, havendo, no entanto, patamares mínimos que devem ser respeitados para cada secção (Joy Davidson, 2011).

A totalidade dos requisitos do DSA encontram-se enumerados no anexo 2 intitulado:

Anexo 2: Requisitos do Data Seal of Approval.

REPOSITÓRIOS DE ACESSO ABERTO E PRESERVAÇÃO DIGITAL

A massificação da utilização das tecnologias digitais em instituições de ensino e investigação a nível mundial, bem como as alterações que se vêm produzindo nas formas de armazenar, preservar, aceder e partilhar a informação resultante das atividades científicas (desde os dados primários até às publicações como os artigos de revista), têm vindo a crescer em dimensão e em complexidade, de acordo com a própria evolução no meio científico. Muitos destes resultados da investigação científica possuem um valor duradouro e devem ser preservados para garantir a maximização de investimentos, para que o conhecimento possa ser reutilizado, para que a sua fiabilidade possa ser garantida e a memória organizacional conservada.

Os repositórios de acesso aberto foram concebidos originalmente como uma forma de dar acesso imediato e amplo a trabalhos de investigação resultantes de atividades científicas, mas têm vindo a assumir cada vez mais um papel de curadores da produção científica, exigindo a adoção de políticas específicas e ferramentas para a sua preservação e curadoria¹⁶.

O primeiro repositório de acesso aberto a surgir no panorama internacional foi o arXiv¹⁷, criado em Agosto de 1991, sob a égide de Paul Ginsparg em *Los Alamos National Laboratory* (daí seu nome original, *LANL preprint archive*), e concebido inicialmente como um arquivo para *preprints*¹⁸ na área da Física e posteriormente alargado por forma a incluir disciplinas como a Astronomia, a Matemática, as Ciências da Computação, a Ciência Não-Linear, a Biologia Quantitativa e, mais recentemente, a Estatística.

¹⁶ Do inglês **Curation**. Por curadoria podemos compreender o conjunto de ações que garantem que um conjunto de dados é genuíno, permitindo o seu uso por outros que não os seus produtores. A curadoria pode envolver ações de descrição dos dados, de ligação destes a outros que os tornem inteligíveis, de registo dos usos que tenham e dos resultados a que tenham dado origem. A curadoria envolve também ações de preservação, em que a representação dos dados e os seus metadados tenham de ser modificados.

¹⁷ Sítio do repositório **arXiv** acessível em: <http://arxiv.org/>

¹⁸ Tipicamente um **Preprint** corresponde a um texto digital de um artigo que ainda não foi avaliado e revisto por pares (peer-reviewed) e ainda não foi aceite para publicação por uma revista científica.

No domínio da Física já existia uma “tradição” de troca de cópias entre pares de *preprints*, como uma forma de disseminar e expor mais rapidamente resultados científicos. A utilização de um acervo digital acessível universalmente através da WWW abriu caminho a uma nova forma de divulgação, mais barata, mais fácil de administrar e mais rápida do que os tradicionais sistemas suportados até então no papel.

O sucesso do arXiv, um repositório disciplinar¹⁹, foi logo seguido pelo lançamento de serviços similares para outras áreas temáticas e instituições de grande dimensão e, eventualmente confluuiu, em 1999, no surgimento da *Open Archives Initiative*²⁰ (OAI), que definiu, entre outros aspetos, um código partilhado para *tags* de metadados e protocolos de interoperabilidade²¹. Um dos resultados subsequentes à primeira reunião da OAI foi a adaptação de um *software* já existente, o *CogPrints*²², para facilitar a criação de um novo tipo de repositórios: os repositórios institucionais²³. Esta plataforma foi designada *EPrints* e foi apresentada publicamente em 2000. Desde então, foram surgindo outros sistemas e plataformas para a criação de repositórios como o DSpace e o Fedora.

¹⁹ **Repositórios disciplinares ou temáticos** são sistemas de informação que arquivam resultados de I&D de uma dada disciplina, de acordo com os princípios do *open access* (acesso aberto).

²⁰ Sítio da **Open Archives Initiative** acessível em: <http://www.openarchives.org/>

²¹ Protocolo **Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)** é um protocolo desenvolvido pela *Open Archives Initiative*, utilizado para distribuir e recolher metadados principalmente no que concerne a descritores de documentos.

²² O **CogPrints** é um arquivo eletrónico na área das ciências cognitivas criado em 1997 e moderado por Stevan Harnad. Na atualidade possui cerca de 4000 artigos em acesso aberto em disciplinas como: a Psicologia, a Neurociência, a Linguística, Ciências da Computação, a Filosofia, a Biologia, a Medicina, a Antropologia, bem como outras áreas das ciências físicas, sociais e matemáticas pertinentes no estudo da cognição.

²³ **Repositórios institucionais** são sistemas de informação que servem para armazenar, preservar e difundir a produção intelectual de uma dada instituição, normalmente uma comunidade universitária. Podem ser criados e mantidos de forma individualizada, ou por grupos de instituições que trabalhem numa base cooperativa.

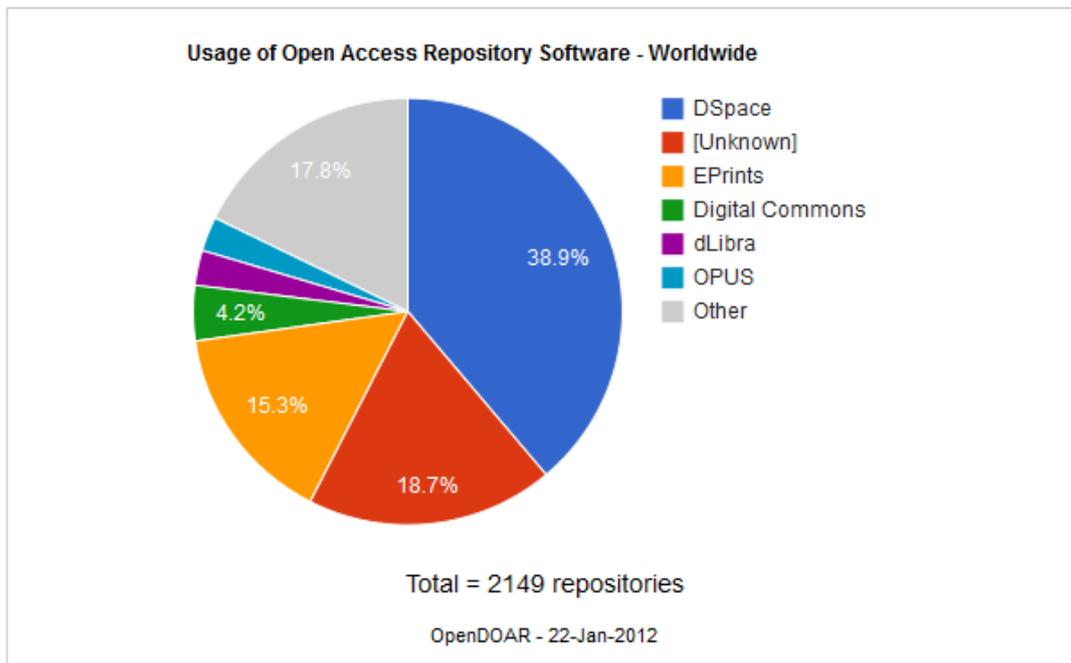


Figura 4 - Utilização de *softwares* de repositórios a nível mundial (fonte: OpenDOAR).

Presentemente existem mais de 2100 repositórios de acesso aberto em todo o mundo, dos quais cerca de 82% (cerca de 1760) são repositórios institucionais e 11,2% (cerca de 240) são repositórios disciplinares.

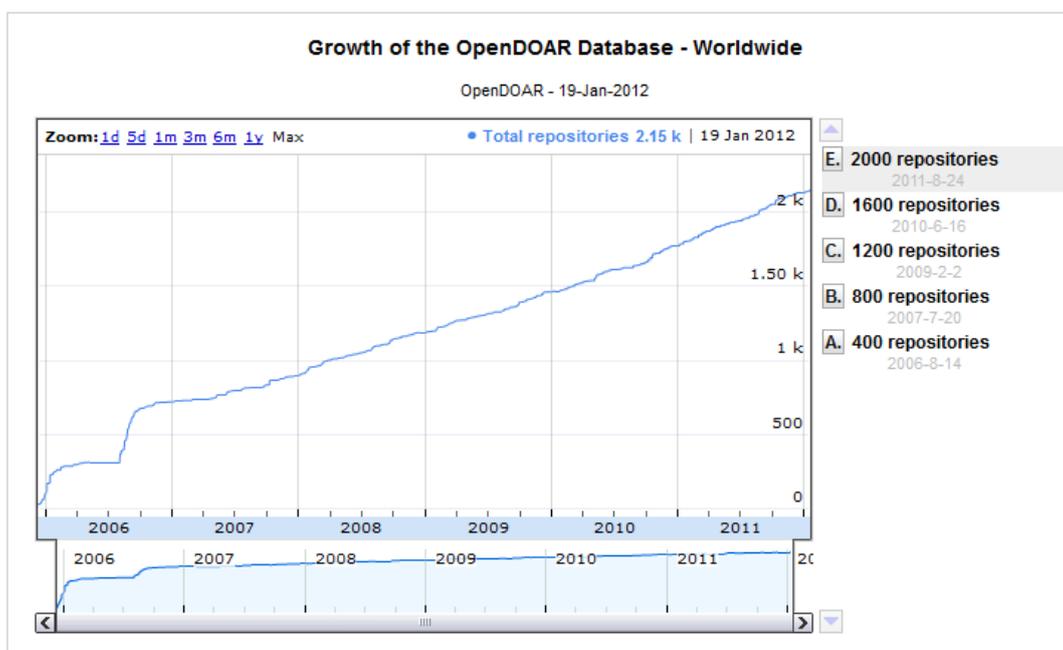


Figura 5 - Evolução do crescimento do número de repositórios a nível mundial (fonte: OpenDOAR).

A questão de saber se os repositórios de acesso aberto devem possuir ou não responsabilidades de preservação tem sido discutida desde há alguns anos. Segundo

algumas correntes de opinião, existe uma forte resistência à ideia, com o argumento de que os repositórios existem exclusivamente com o intuito de acelerar a difusão, de potenciar os resultados de investigação de excelência e que a preservação deve estar mais direcionada para as versões 'oficiais' impressas, registos publicados em revistas ou conservados nas bibliotecas. Existem, naturalmente, argumentos contrários a essa visão. Um deles é o de que o acesso aberto aos também designados *e-prints*²⁴ possibilita aos investigadores um acesso de âmbito mundial a um manancial de informação científica, que de outra forma não teriam acesso porque as suas instituições não possuem acesso a determinadas revistas científicas de acesso pago. Outro argumento é o de que os repositórios institucionais podem conter mais do que meras versões alternativas de artigos científicos, por exemplo, poderão conter versões estendidas de artigos, documentos de conferências inéditos, dados científicos, literatura cinzenta, recursos de ensino e aprendizagem ou mesmo material institucional (registos administrativos e outros).

Na década subsequente ao aparecimento das primeiras plataformas para repositórios de acesso aberto, foram desenvolvidas uma série de ferramentas no âmbito da preservação, desde a elaboração de planos de preservação, políticas para a extração de metadados de preservação de ficheiros, paralelamente com arquiteturas modulares para a ligação de todas as ferramentas em conjunto. Ainda que, no contexto dos repositórios institucionais, a investigação e o desenvolvimento nas áreas de preservação e curadoria, tenha sido lenta, progressivamente tem vindo a crescer e a ganhar maior notoriedade como demonstram a diversidade de projetos e iniciativas que vão decorrendo neste domínio no panorama internacional.

²⁴ Por **e-print** podemos entender a versão digital de um artigo científico com *peer-review* (revisão por pares), antes ou depois da avaliação e publicação.

Softwares/Plataformas de repositórios

Nesta secção abordamos três, de entre os muitos sistemas de RIs existentes, mais conhecidos e utilizados.

EPrints²⁵

O software EPrints é uma das plataformas mais disseminadas em termos de repositórios institucionais. O EPrints tem sido desenvolvido pela "School of Electronics and Computer Science" da Universidade of Southampton, Reino Unido, e tem as suas raízes num software associado ao repositório disciplinar CogPrints. A sua versão mais recente é a 3.3.7 e a partir da versão 3.x, o sistema Eprints passou a incorporar três características que visam melhorar o suporte da preservação na plataforma (Brody, Carr & McSweeney, 2010):

- *History module.* Este módulo fornece para cada objeto de um repositório um *log* das alterações que sofreu o seu registo dentro do sistema do repositório. Atualmente, isto é apenas utilizado para rastrear alterações efetuadas no registo, para fins de controlo, mas pode também ser utilizado para acompanhar ações de preservação, tais como as migrações de formatos.
- *METS e DIDL export plugins.* Estes *plugins* permitem que objetos complexos, ou seja, objetos compostos por mais do que um ficheiro, possam ser exportados como um pacote em formato METS²⁶ ou formato MPEG-21 DIDL (ISO/IEC 21000-2:2003).
- *Creative Commons licensing.* No futuro, poderão subsistir algumas dúvidas sobre se um repositório possui ou não os direitos e as permissões necessárias para incidir ações de preservação em conteúdos depositados. Foi adicionada uma opção de licença, apresentada durante o processo de depósito, que permite aos autores/depositantes conceder explicitamente essas permissões ao repositório. A plataforma armazena essas permissões, conjuntamente com o resto dos metadados do registo.

²⁵ Sítio do **EPrints Software** acessível em: <http://www.eprints.org/>

²⁶ Sítio do formato **METS schema** acessível em: <http://www.loc.gov/standards/mets/>

No EPrints estão previstas mais ações no que concerne à preservação e que serão desenvolvidas como parte do projeto *KeepIt* (descrito à frente). Entre as ferramentas e os serviços propostos²⁷ estão 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.

DSpace²⁸

É a plataforma de repositórios mais disseminada da atualidade e tem uma extensa lista de utilizadores. Foi lançada inicialmente pelo *Massachusetts Institute of Technology* e pelos *Hewlett-Packard Laboratories* em 2002 com o objetivo de fornecer um sistema de repositório para documentos digitais resultantes de investigação ou destinados à educação e distribuído sob uma licença de código aberto. Atualmente é desenvolvido sob um modelo colaborativo em comunidade, com a liderança estratégica da agora designada, *DSpace Foundation*, que se associou com o Fedora Commons para formar uma iniciativa comum sob a chancela *DuraSpace*. A versão mais recente da plataforma DSpace é a versão 1.8.1.

O DSpace realiza preservação *bit-level* em todos os objetos depositados. Relativamente aos formatos não suportados, o DSpace faz distinção entre formatos conhecidos e desconhecidos: aos formatos conhecidos são atribuídos os seus respetivos identificadores de formato, enquanto os formatos desconhecidos são marcados como *bytestream* genérico utilizando o tipo MIME '*application/octet-stream*'. Em termos de preservação, os formatos conhecidos são fechados, mas tão comuns que se possa crer que existirão ferramentas para preservar os arquivos nesses formatos, ao passo que no que concerne aos formatos desconhecidos, não sendo tão usuais, a sua preservação não é dada tão certa.

Fedora Commons²⁹

O Fedora Commons (Flexible Extensible Digital Object Repository Architecture) surgiu em 1997 e, ao contrário do DSpace ou do EPrints, não constitui como uma plataforma específica para repositórios mas sim como uma arquitetura extensível que pode ser

²⁷ Sítio *EPrints Digital Preservation* acessível em: <http://preservation.eprints.org/>

²⁸ Sítio da *DuraSpace Foundation* acessível em: <http://duraspace.org/>

²⁹ Sítio da plataforma *Fedora* acessível em: <http://www.fedora-commons.org/>

utilizada para desenvolver *software* para repositórios. Criada pela Universidade de Cornell, é atualmente mantida pela iniciativa *DuraSpace* tal como o sistema DSpace. Um dos princípios é o da agregação de conteúdos locais, distribuição de objetos digitais e a associação destes a serviços. O sistema inclui ainda um modelo de relações baseado no RDF (Resource Description Framework) do W3C usado para ligar os objetos aos seus componentes. Está disponível em licença de código aberto e tem sido utilizado em diversas aplicações para bibliotecas digitais, arquivos, repositórios institucionais e sistemas de objetos de aprendizagem. A versão atual do Fedora é a versão 3.5.

O *Fedora Commons* criou em 2005 um grupo de trabalho para trabalhar especificamente em questões de preservação, desde então, vários recursos foram adicionados ou melhorados no Fedora em termos de preservação digital, destacando-se os seguintes:

- *Object versioning*. O Fedora suporta versões de fluxos de dados (dados e metadados) e fornecedores de serviços e preserva a ligação entre uma versão do fluxo de dados e a versão correspondente do fornecedor. As diferentes versões são mantidas logicamente dentro do mesmo objeto digital e a informação das relações entre as versões é conservada também nos metadados do objeto.
- *Format characterisation*. Existe um módulo externo para validar os formatos dos ficheiros e extrair metadados a partir de fluxos de dados utilizando *JHOVE*.

Projetos e arquiteturas de preservação para repositórios

Nesta seção iremos abordar alguns dos projetos e iniciativas no domínio da preservação para repositórios e as suas arquiteturas tecnológicas.

CASPAR³⁰

O projeto CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval) foi um projeto integrado, financiado pela UE, que abordou uma série de questões no domínio da preservação e de curadoria em diversas disciplinas. O CASPAR decorreu entre Abril de 2006 e Setembro de 2009, e nessa fase desenvolveu uma arquitetura e um fluxo de trabalho para preservação, conjuntamente com ferramentas integradas na própria arquitetura (Figura 6). Essas ferramentas incluem: REPINF

³⁰ URL sítio **projeto CASPAR** acessível em: <http://www.casparpreserves.eu/>

(Representation Information Toolkit), VIRT (Virtualization), REG (Registry), PACK (Packaging), PDS (Preservation Data Stores), FIND (Finding Aid), KM (Knowledge Manager), POM (Preservation Orchestration Manager), DAMS (Data Access Manager e Security), DRM (Digital Rights Manager) e AUTH (Authenticity) (CASPAR Consortium, 2007).

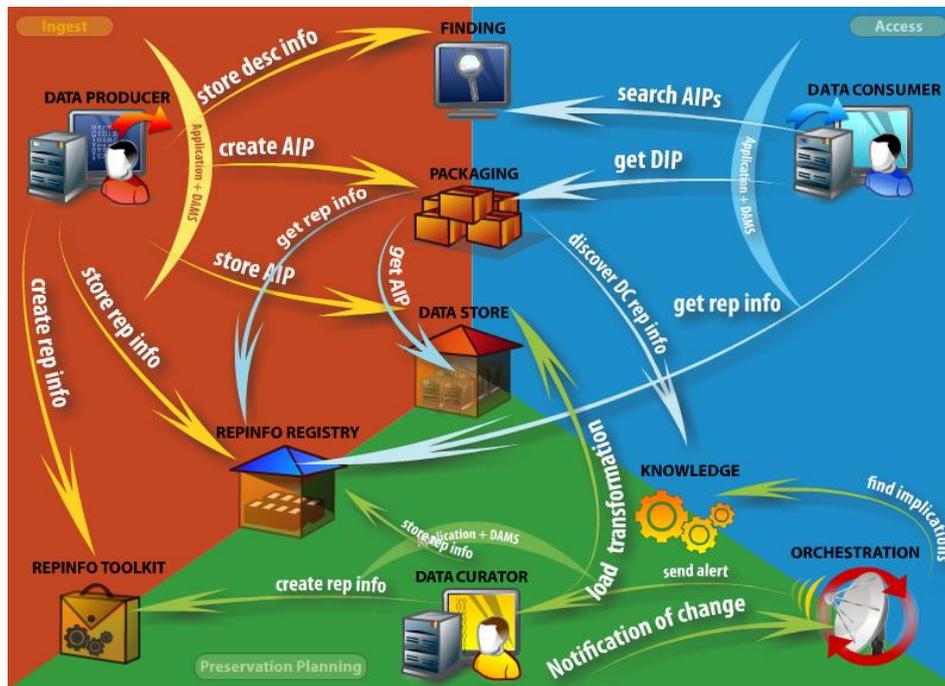


Figura 6 - Diagrama (workflow) de preservação CASPAR (fonte sítio CASPAR).

O projeto criou três infraestruturas de teste nas quais a metodologia de preservação CASPAR foi aplicada a diferentes disciplinas. Estas infraestruturas de teste tinham como objetivos: garantir que o fluxo de trabalho de preservação CASPAR poderia ser utilizado com sucesso em diversos contextos e com diversos tipos de dados; determinar a informação de representação e informação de descrição de preservação necessárias para apoiar o uso contínuo desses diferentes tipos de dados a longo prazo e investigar a adaptação de diferentes técnicas de preservação em diferentes circunstâncias - migração, emulação, conservação de código fonte, preservação de *hardware*, a reconstrução de software a partir de documentação. As infraestruturas de teste possibilitaram ainda a oportunidade para produzir ferramentas para auxiliar na recolha de metadados de preservação e produção de pacotes para submissão de informação.

CRiB³¹

A CRiB (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. O sistema CRiB (Figura 7) possui serviços para identificar o formato de um objeto digital, monitorizar formatos quase obsoletos e compará-los com as soluções de migração existentes e em seguida produzir recomendações e estratégias de migração adequadas. As recomendações produzidas pelo sistema levam em consideração as exigências de preservação específicas de cada instituição/organização.

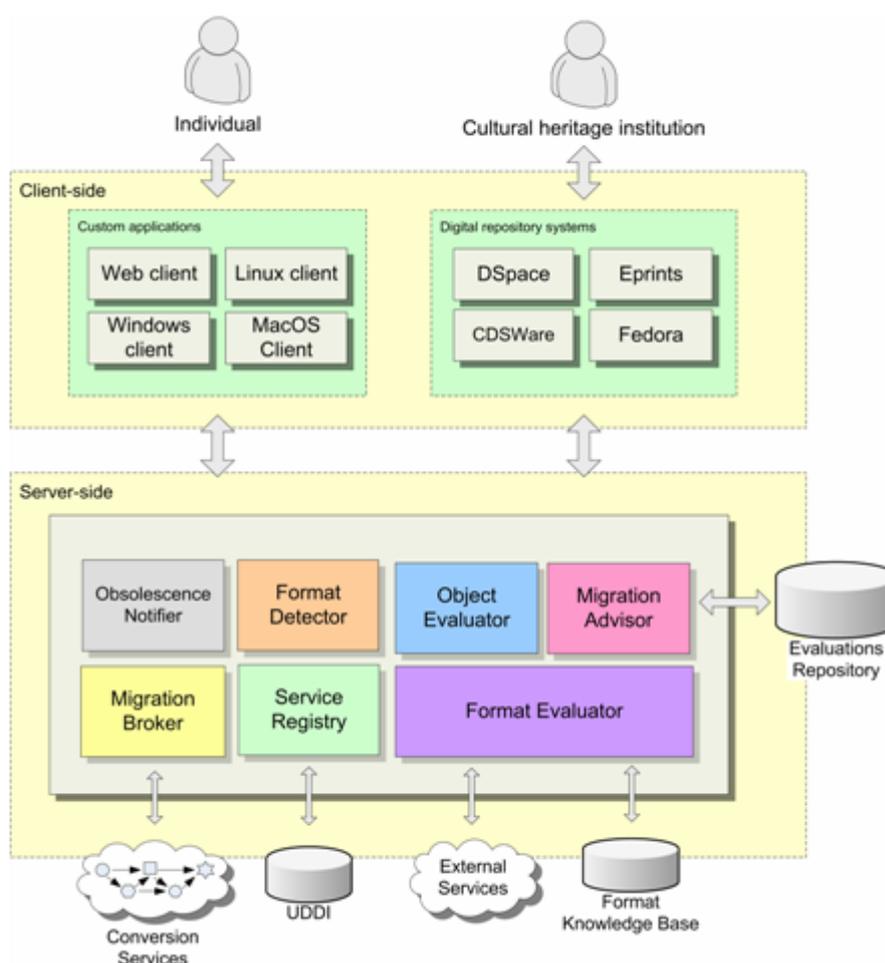


Figura 10 - Visão da arquitetura CRiB (Fonte sítio CRiB).

Com base na arquitetura CRiB, a Direção Geral de Arquivos (DGARQ) de Portugal, já desenvolveu e criou o RODA (Repositório de Objetos Digitais Autênticos)³², um

³¹ Sítio projeto **CRiB** acessível em: <http://crib.dsi.uminho.pt>

³² Sítio do repositório **RODA** acessível em: <https://roda.dgarq.gov.pt/>

repositório multimédia, assente na plataforma Fedora Commons, com o qual a DGARQ espera incorporar documentos eletrónicos de instituições nacionais de forma controlada assegurando a sua gestão ao longo do tempo, bem como o seu acesso aos utilizadores. Está também a ser estudada a integração do CRiB com outros sistemas, por exemplo, Becker et al. (2008) descrevem como a CRiB poderá ser utilizada em conjunto com um serviço de planeamento de preservação, como a ferramenta *PLATO* (abordada no próximo capítulo).

A CRiB assenta em *web services* e é capaz de realizar as seguintes atividades em plataformas de repositórios com o DSpace, EPrints e Fedora:

- Recomendação de migrações alternativas e adequadas, que se ajustem às exigências de preservação específicas de cada instituição em particular;
- Conversão de objetos digitais para codificações atualizadas, que a maioria dos utilizadores seja capaz de interpretar;
- Avaliação do resultado de migração, comparando o objeto digital original com os seus homólogos convertidos e identificação das propriedades importantes que não foram corretamente preservadas;
- Geração de relatórios de migração em formato adequado para inclusão nos metadados de preservação dos objetos migrados;

PANIC³³

O projeto PANIC (Preservation web services Architecture for New media, Interactive Collections and scientific data) foi conduzido pelo *Distributed Systems Technology Centre* (DTSC) e pela Universidade de Queensland, entre 2003 e 2006. O objetivo do projeto consistia na elaboração de uma arquitetura que pudesse ser adicionada a repositórios e que lhes permitisse suportar a preservação de objetos digitais. A arquitetura do sistema PANIC (Figura 8) é baseada em três processos principais: a captura de metadados de preservação, deteção e notificação de obsolescência e serviço de descoberta e invocação de preservação. Para ajudar na captura de metadados de preservação, surgiu o projeto PREMINT (Preservation Metadata Input Tool), que fornece uma interface amigável para escrever metadados de preservação em pacotes de *METS* e *MPEG-21 DIDL*. Também codificou entidades *PREMIS* como ontologias *OWL*. Para apoiar a deteção e notificação de obsolescência, o PANIC desenvolveu três bases de dados: uma versão do *software* do registo, um formato do registo e um registo de formato recomendado.

³³ Sítio do projeto **PANIC** acessível em: <http://www.itee.uq.edu.au/~eresearch/projects/panic/>

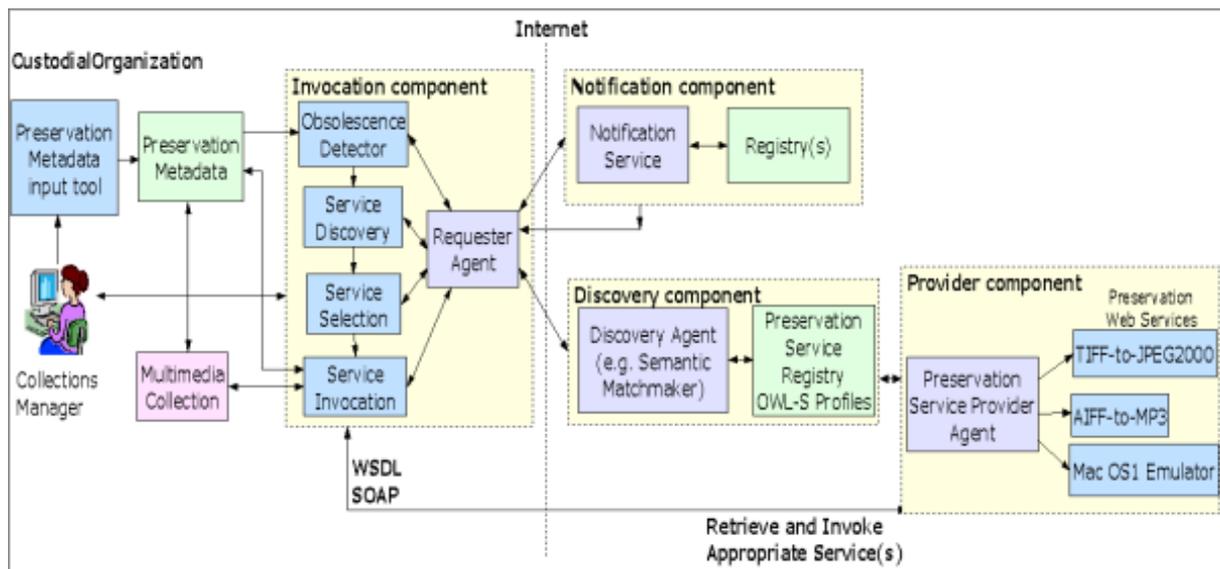


Figura 8 - Visão da arquitetura PANIC (Fonte sítio PANIC).

PLANETS³⁴

O Planets (Preservation and Long-term Access through NETworked Services) foi um projeto previsto para quatro anos e iniciado em Junho de 2006. Tratou-se de um projeto financiado pela União Europeia (EU) com vista a construir serviços práticos de curadoria e ferramentas para bens culturais e científicos. Foi coordenado pela *British Library* e envolveu bibliotecas nacionais, arquivos nacionais, bibliotecas de universidades e organizações comerciais de toda Europa. A arquitetura de preservação desenvolvida pelo projeto é conhecida como o *Planets Interoperability Framework*. Esta arquitetura foi desenhada e projetada para funcionar em paralelo com os *softwares* de repositórios existentes, ingerir material diretamente do repositório e tornar os resultados de preservação disponíveis para nova ingestão no repositório.

A um nível superior, a arquitetura Planets centra-se num servidor gateway, sendo que este possui dois conjuntos de interfaces: Portal Serviços, que fornece um meio seguro para os utilizadores interagirem com o servidor, e a *API WFlow*, que permite ao servidor interagir com serviços de preservação como uma parte de um fluxo de trabalho de preservação. Os principais componentes subjacentes ao servidor gateway são o *Core Registry* (contendo informações de representação e informações significativas sobre os tipos de propriedades de objetos digitais e formatos, conjuntamente com ferramentas de caracterização e ação de preservação), o *Service Registry* (contendo informações sobre

³⁴ Sítio do projeto **Planets** acessível em: <http://www.planets-project.eu/>

serviços compatíveis com o Planets) e do *Data Registry* (contendo os objetos digitais modificados pelo sistema).

PRESERV³⁵ and Keepit

O Projeto PRESERV resultou de uma colaboração entre as Universidades de Oxford e Southampton e os *The National Archives* (TNA), no Reino Unido, e decorreu em duas fases distintas compreendidas entre Fevereiro 2005 e Março de 2009. O objetivo original do projeto consistia no desenvolvimento de um serviço de ingestão OAIS para repositórios EPrints ligados ao serviço *TNA PRONOM* para a identificação e verificação de formatos de ficheiros. No decurso do projeto, os desenvolvimentos passaram a prever outros tipos de repositórios

Entre os resultados do PRESERV encontram-se um conjunto de três modelos para a ligação de serviços de preservação, com repositórios institucionais, expresso em termos do Modelo de Referência OAIS (Hitchcock, Brody, Hey & Carr, 2007a). O trabalho do projeto na integração de serviços de preservação em repositórios levou ainda a duas inovações. A primeira foi a inclusão de um *history module* na v3.x EPrints, que fornece um log das alterações que um documento sofreu ou o seu registo dentro do sistema do repositório. O segundo, resultante do trabalho na aplicação de PRONOM a dois repositórios piloto, foi uma ideia para um *web service* baseado em OAI-PMH para a produção de perfis de formato de ficheiros para vários repositórios ao mesmo tempo.

Com o término da segunda fase, Preserv 2, as estratégias, políticas e serviços do projeto foram desenvolvidas e colocadas em prática numa série de repositórios como parte do projeto KeepIt³⁶, um projeto financiado pelo *Joint Information Systems Committee* (JISC) que teve início em Abril de 2009.

³⁵ Sítio do projeto **PRESERV** acessível em: <http://preserv.eprints.org/>

³⁶ Final project report from the JISC **KeepIt** Project: <http://ie-repository.jisc.ac.uk/553/1/finalreport-keepit10.pdf>

RepoMMan³⁷ and REMAP³⁸

O projeto RepoMMan (Repository Metadata and Management) decorreu entre 2005 e 2007 e tentou incorporar o uso de um repositório institucional nos fluxos de trabalho de utilizadores académicos e administrativos, focando-se em ações de gestão (depósito, acesso, partilha e publicação itens) e os metadados necessários para as apoiar. Este projeto teve continuidade num projeto subsequente de dois anos designado REMAP (Records Management and Preservation).

O projeto REMAP investigou como instituições de Ensino Superior no Reino Unido poderiam utilizar um repositório digital para apoiar a gestão de registos e preservação digital. Para o efeito, estendeu o modelo de fluxo de trabalho RepoMMan para incluir o ciclo de vida completo dos objetos digitais, e introduziu uma ferramenta de organização para trabalhar com serviços de conservação externos, bem como uma camada de notificação com três tipos de alerta no repositório (baseado em eventos, tempo decorrido e estado).

Seamless Flow

O programa Seamless Flow decorreu entre 2005 e 2008 sob a égide do *The National Archives* (TNA) do Reino Unido, com o intuito de criar de uma infraestrutura ativa de preservação de registos eletrónicos governamentais. A estrutura foi baseada em três atividades principais: caracterização, planeamento e ações de preservação.

Quando um objeto é depositado num repositório digital, a ferramenta *DROID* é usada para identificar o formato. O sistema verifica o formato no registo *PRONOM* e descobre as ferramentas adequadas para verificar se o objeto é bem-formado (sintaticamente compatível com as especificações do formato) ou válido (bem formado e em conformidade com as restrições semânticas). No final, a ferramenta *JHOVE* é utilizada para extrair metadados de preservação do objeto e gerar informações sobre suas propriedades significativas.

Os objetos dentro do repositório são verificados regularmente para fatores de risco com recurso ao registo *PRONOM*. Se em algum momento o fator de risco para um dado objeto ou uma classe de objetos ultrapassar um patamar crítico, o sistema verifica o

³⁷ Sítio do Projeto **RepoMMan** acessível em: <http://www.hull.ac.uk/esig/repomman/>

³⁸ Sítio do Projeto **REMAP** acessível em: <http://www.hull.ac.uk/remap/>

registo *PRONOM* para determinar formatos alternativos apropriados para os objetos em situação de risco e possíveis vias de migração. Logo que o plano de preservação seja dado como adequado para a situação, o sistema executa o plano. Após a execução do plano, os ficheiros recém-migrados são caracterizados e comparados com os originais para garantir que nenhuma das propriedades significativas foram danificadas.

SHAMAN³⁹

O projeto SHAMAN (Sustaining Heritage Access through Multivalent Archiving) foi financiado pela UE e visou estabelecer as bases conceituais e técnicas para uma infraestrutura de preservação digital distribuída em rede. O SHAMAN teve início em 2008 e foi delineado para uma duração de três anos, entre os entregáveis deste projeto contam-se um conjunto de serviços básicos e implementações de ferramentas de preservação de referência. Os serviços incluem a integração de dados em rede, uma biblioteca digital, um arquivo persistente, uma representação de contexto, serviço de anotação e preservação, um serviço de análise linguística profunda e representação semântica e tecnologias de anotação. Conjuntamente com uma seleção de ferramentas de preservação, esses serviços foram integrados em ambientes aplicativos e testado em três domínios distintos: instituições de memória cultural, design e engenharia e e-Science (Figura 9).

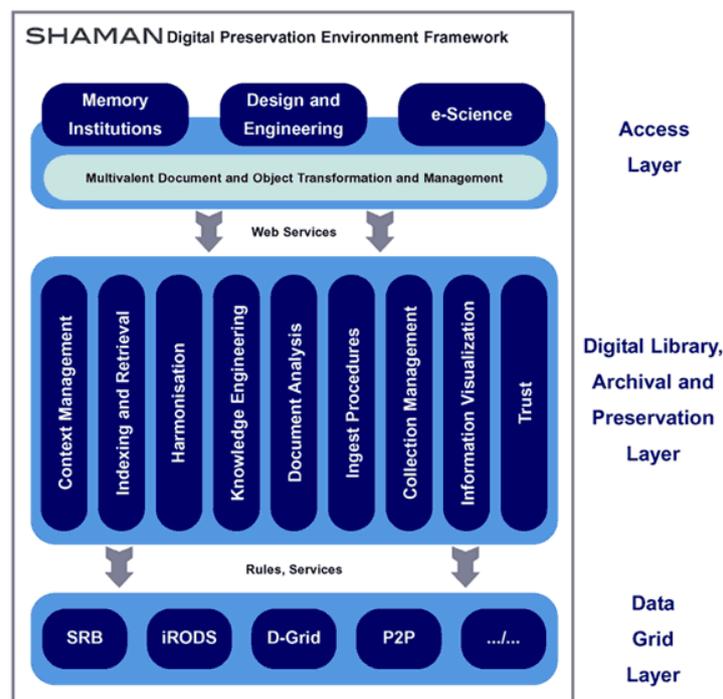


Figura 9 - Visão da arquitetura SHAMAN (Fonte sítio SHAMAN).

³⁹ Sítio do projeto SHAMAN acessível em: <http://shaman-ip.eu/>

SHERPA DP⁴⁰

O projeto inicial SHERPA DP investigou a prestação de serviços no domínio da preservação em repositórios operados pelo consórcio SHERPA. O projeto desenvolveu um modelo para a prestação de um serviço de preservação baseado tipicamente em repositórios institucionais: em termos de OAIS, a função de acesso do fornecedor de conteúdos alimenta a entidade de ingestão do fornecedor de serviços e vice-versa. O SHERPA DP desenvolveu um serviço de demonstração, criando uma arquitetura em que ferramentas de software como o *Fedora Commons*, *DROID* e *JHOVE*, e esquemas como *METS* e *PREMIS*, pudessem ser ligadas (Figura 10). Foram realizados mais trabalhos para desenvolver um modelo de gestão e custos com base nos resultados *Projeto LIFE*.

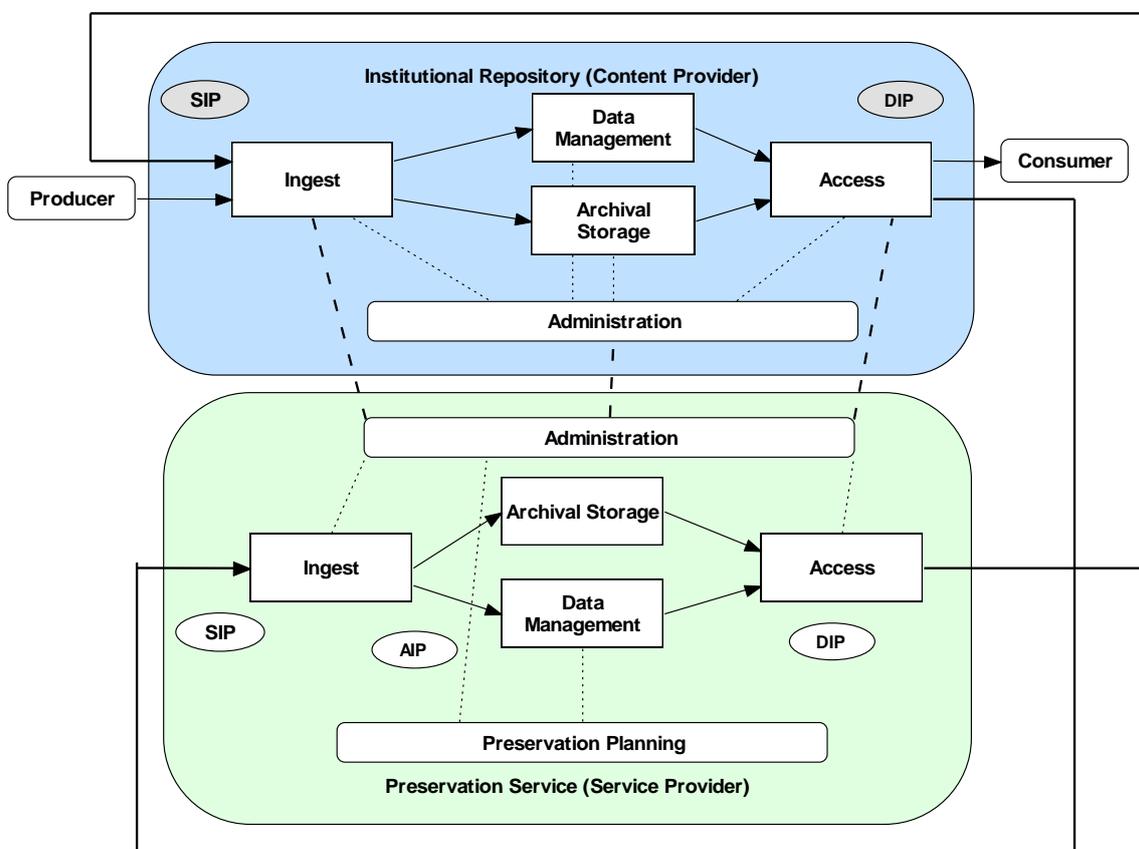


Figura 10 - Visão da arquitetura SHERPA DP (Fonte: Knight & Anderson, 2007).

O projeto SHERPA DP teve início em Março de 2005, sendo sucedido em Março de 2007 pelo projeto SHERPA DP2⁴¹. Na sua segunda fase o projeto expandiu o modelo de serviço de preservação de um duplo sistema de fornecedor para um modelo mais

⁴⁰ Sítio do projeto **SHERPA DP** acessível em: <http://www.sherpa.ac.uk/projects/sherpapdp.html>

⁴¹ Sítio do projeto **SHERPA DP2** acessível em: <http://www.sherpdp.org.uk/>

desagregado. O SHERPA DP2 também investigou diferentes métodos pelos quais os fornecedores de serviços de preservação poderiam obter conteúdos dos seus clientes, investigando as exigências de preservação dos fornecedores de conteúdos e estendeu as especificações de metadados SHERPA DP para abarcar mais tipos de conteúdos.

SCAPE⁴²

O projeto SCAPE (SCAlable Preservation Environments) é um projecto europeu em curso, financiado sob a égide do 7.º Programa Quadro da União Europeia (FP7 ICT-2009.4.1), que visa avançar o estado da arte no domínio da preservação digital. Para isso, reúne especialistas de bibliotecas, laboratórios de investigação, universidades, empresas, entre outros. O objetivo é investigar e desenvolver sistemas de preservação digital escaláveis capazes de lidar com milhões de objetos em tempo útil.

No consórcio deste projeto constam instituições de renome mundial, como a Universidade Técnica de Viena, Universidade Técnica de Berlim, Universidade Marie Curie, Universidade de Manchester e as Bibliotecas Nacionais do Reino Unido, Áustria, Dinamarca e Holanda.

Nesta iniciativa, existe também uma participação do sector empresarial em atividades de investigação & desenvolvimento, estando representado por três empresas: a KEEP SOLUTIONS (Portugal), a Ex Libris Group e a Microsoft Research.

Ferramentas e estratégias para a preservação em repositórios

Qualquer sistema que lide com dados deverá manter informações adicionais sobre esses dados se pretende geri-los de forma eficaz. Sistemas que visam suportar a preservação possuem necessidades de metadados muito específicas e que resultam de diversos projetos e iniciativas.

Uma das iniciativas mais importantes ao nível da preservação digital é o *OAIS Reference Model terms Representation Information*, pois permite que um objeto digital seja interpretado e compreendido. Mas, para além do OAIS, foram surgindo outras iniciativas que procuram fornecer uma orientação mais abrangente sobre os metadados a recolher.

⁴² Sítio do Projeto **SCAPE** acessível em: <http://www.scape-project.eu/>

A iniciativa *PREMIS*⁴³, por exemplo, coloca enfoque na preservação, mas pretende ser aplicável a todos os objetos digitais. Outras iniciativas e projetos como o *CAIRO* (Complex Archive Ingest for Repository Objects)⁴⁴ focam-se em determinados tipos de objetos digitais, mas consideram uma vasta gama de aplicações para metadados. Um grande desenvolvimento nesta área foi a introdução de perfis de aplicação do Dublin Core, que podem ser delimitados de acordo com o tipo de conteúdo e aplicação.

De acordo com o modelo de referência OAIS, o planeamento de preservação é uma atividade que engloba a monitorização ambiental do repositório, a revisão dos conteúdos do repositório à luz desse acompanhamento e a elaboração de planos para a migração de conteúdos para novos formatos ou atualizar a maneira como é processada ou divulgada aos consumidores.

Abordamos em seguida três ferramentas que auxiliam no planeamento de preservação, a saber: AONS II, Plato, e PRONOM ROAR.

AONS II⁴⁵

O AONS II (Automated Obsolescence Notification System II) é uma ferramenta baseada numa plataforma independente de código aberto configurável e que fornece automaticamente informações de registos de autoridade internacionais para apoiar no planeamento de preservação. Atualmente existe a possibilidade de obter informações de sítios como *PRONOM* e a *Library of Congress Sustainability of Digital Formats*, sendo previsíveis outros recursos em termos futuros.

O objetivo primordial para a AONS era o de constituir-se como um serviço de notificação de obsolescência para a arquitetura PANIC. Com o AONS II, o enfoque passou a ser o de suporte de uma infraestrutura nacional federada, bem como de repositórios locais e de repositórios organizacionais em rede. Permitindo especificamente, que um gestor de repositório possa efetuar avaliações de risco aos formatos de ficheiros contidos no seu repositório, tanto no momento de introdução, como no momento de execução de operações em lote execução ou programadas. O utilizador também pode configurar um perfil de risco para um repositório de modo a que quando se registem alterações em

⁴³ Sítio da iniciativa **PREMIS** acessível em: <http://www.loc.gov/standards/premis/>

⁴⁴ Sítio do projeto **CAIRO** acessível em: <http://cairo.paradigm.ac.uk/>

⁴⁵ Sítio da ferramenta **AONS II** acessível em: <http://www.apsr.edu.au/aons2/>

determinados indicadores, o AONS II envie uma notificação recomendando que seja efetuada uma avaliação de risco ou ações de preservação (Pearson & Walker, 2007).

Plato⁴⁶

O Plato é uma ferramenta Web de 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. Esta ferramenta implementa a metodologia de planeamento de preservação *Planets* (Strodl & Becker, 2007) e integra os registos e serviços para ações de preservação e caracterização. É baseado num trabalho anterior realizado pelo *DELOS Digital Preservation Cluster* e baseia-se na análise de utilização para avaliar o desempenho de várias soluções face a requisitos e objetivos bem definidos. A metodologia pode ser aplicada a qualquer tipo de estratégia - migração, emulação, normalização, etc. - e foi validada numa série de estudos de caso.

A ferramenta *Plato* suporta a identificação de formatos de ficheiros via *DROID*, a caracterização de conteúdos via *XCL* e comparação de objetos através do serviço *Planets XCDL*. Suporta o uso de um modelo e fragmentos de uma biblioteca para reutilização de requisitos em diferentes conjuntos de objetos e pode importar e exportar planos de preservação em formato XML.

PRONOM-ROAR

O PRONOM-ROAR é um serviço que fornece perfis de formato de ficheiros, em mais de 200 repositórios indexados pelo *Registry of Open Access Repositories* (ROAR)⁴⁷. O ROAR indexa esses repositórios recolhendo registos de metadados através do protocolo OAI-PMH. O PRONOM-ROAR gera os perfis de formato de ficheiros descarregando todos os ficheiros associados a esses registos e identifica seu formato utilizando *DROID*. O serviço tem algumas limitações, por exemplo, não inclui ficheiros com mais de 2MB em termos de tamanho, ou ficheiros armazenados num servidor diferente da interface OAI-PMH (Brody, Carr, Hey, Brown & Hitchcock, 2007).

⁴⁶ Sítio da ferramenta **PLATO** acessível em: <http://www.ifs.tuwien.ac.at/dp/plato/>

⁴⁷ Sítio **Registry of Open Access Repositories** acessível em: <http://roar.eprints.org/>

CONCLUSÕES E RECOMENDAÇÕES

O presente estudo revela que a preservação digital, no âmbito dos repositórios de acesso aberto, tem evoluído significativamente nos últimos anos. No panorama internacional, têm vindo a registar-se múltiplas atividades, iniciativas e projetos, em especial nos últimos cinco anos. São iniciativas com origens e âmbitos diversificados, desde projetos menor dimensão promovidos por grupos de investigação até projetos de grande dimensão internacional.

No entanto, e por outro lado, apesar da crescente consciencialização e interesse no que concerne às questões relacionadas com a preservação digital, é ainda reduzido o número de repositórios com políticas e estratégias consolidadas neste domínio, pelo que, na maioria dos repositórios, as ações relacionadas com a preservação digital são quase inexistentes ou incipientes.

Como demonstram alguns estudos recentes⁴⁸, ainda subsistem dúvidas, incertezas e lacunas quanto aos papéis e responsabilidades, isto é, quem deverá ser responsável pela preservação e curadoria, a qualidade e interoperabilidade dos repositórios, ou a inexistência de enquadramento jurídico apropriado em termos de preservação digital.

Na vertente técnica, constata-se que ferramentas para suporte a tarefas de preservação digital estão em franco desenvolvimento e amadurecimento. Por exemplo, projetos como o CASPAR, Planets e SHAMAN apresentam já uma série de ferramentas úteis para a curadoria de um ciclo de vida completo, enquanto outros como o SHERPA DP e o CRiB demonstram como essas ferramentas poderão ser integradas para formar um conjunto de ferramentas abrangente. Na vertente mais estratégica, verifica-se também que ferramentas de auditoria e de certificação já atingiram patamares desenvolvimento em que já podem ser utilizadas em termos práticos.

Em Portugal, dos 35 repositórios atualmente registados no portal RCAAP⁴⁹, a totalidade utiliza como suporte a plataforma DSpace e pelo que será conhecido, nenhum destes repositórios possuirá uma política de preservação formal. Para além das funcionalidades

⁴⁸ Cf. "Online survey on scientific information in the digital age". Acessível em linha em:

http://ec.europa.eu/research/science-society/document_library/pdf_06/survey-on-scientific-information-digital-age_en.pdf

⁴⁹ **Directório RCAAP** de repositórios portugueses acessível em: <http://www.rcaap.pt/directory.jsp>

intrínsecas à própria plataforma de suporte ao repositório, haverá também instituições com procedimentos no que concerne aos formatos admissíveis ou que realizam, como normativo interno, migrações de formatos aquando do depósito de documentos (ex. conversões para o formato PDF), mas ainda sem uma sistematização desejável.

No quadro atual, considerando os desenvolvimentos no domínio da preservação e curadoria, e a sua incipiente aplicação em repositórios de acesso aberto, existem certamente oportunidades para a investigação, o desenvolvimento e o teste de novos serviços e tecnologias de qualidade profissional. Possivelmente, ainda será necessário mais algum tempo para que a preservação e curadoria estejam totalmente integrados no trabalho diário dos repositórios portugueses. Esta transição começa a vislumbrar-se, mas será importante que o *momentum* atual dos repositórios portugueses seja prosseguido, para que possamos ser cautelosamente otimistas sobre a sistematização da preservação digital em Portugal.

Tendo em conta o acima exposto, sugerem-se, de seguida, algumas ações e linhas de orientação que poderão ser avançadas com o intuito de promover e facilitar o processo de preservação e curadoria digital nos repositórios de acesso aberto em Portugal.

Recomendações e linhas de orientação:

1. Constituir, no âmbito da comunidade RCAAP, um grupo de interesse no domínio da preservação e curadoria digital, com iniciativas, atividades e canais de comunicação próprios, e no quadro do qual se poderiam concretizar a generalidade das recomendações e linhas de orientação apresentadas em seguida;
2. Realizar um recenseamento e caracterização da situação existente nos repositórios de acesso aberto portugueses, no que diz respeito às políticas, procedimentos e estratégias de preservação digital;
3. Avaliar, definir e concretizar um projeto piloto, no domínio da preservação digital, com a participação de vários repositórios portugueses, com o recurso a arquitetura(s) que possa(m) dotar os repositórios participantes de ferramentas abrangentes em termos de preservação digital;
4. Acompanhar e, se possível e adequado, cooperar com iniciativas, serviços e projetos relevantes em curso na área da preservação digital, com o intuito de conhecer, utilizar e promover as boas práticas neste domínio. Considerando a situação portuguesa, deverá ser avaliada desde logo a exequibilidade e os termos

- de uma possível cooperação entre o projeto RCAAP e/ou os repositórios individualmente com o RODA – Repositório de Objetos Digitais Autênticos;
5. Desenvolver e/ou disseminar documentos de divulgação, formação e suporte, como *Briefing papers*, modelos de políticas e procedimentos, boas práticas e casos exemplares de preservação digital;
 6. Realizar ações de sensibilização, divulgação e formação destinadas a gestores de *repositórios* e responsáveis institucionais;
 7. Avaliar e identificar os recursos e os custos envolvidos nos processos de preservação digital, em especial os relativos aos recursos humanos, quer quanto ao esforço (tempo de trabalho) envolvido, quer quanto às competências requeridas, e eventuais necessidades de formação;
 8. Identificar eventuais condicionalismos éticos e legais, no que concerne a ações de preservação que possam incidir em conteúdos já depositados;
 9. Sensibilizar, incentivar e apoiar os autores e os gestores dos repositórios para a utilização de formatos de ficheiro que facilitem a preservação a longo prazo, aquando do depósito de documentos no(s) repositório(s).

Vista a experiência acumulada pelo projeto Repositório Científico de Acesso Aberto de Portugal (RCAAP) no domínio dos repositórios e de iniciativas de Acesso Aberto em Portugal, será crível que a concretização das linhas de ação acima identificadas possam ser apoiadas e estimuladas através do projeto RCAAP.

ANEXOS

Anexo 1: *Checklist* do TRAC

A - Organizational Infrastructure

A 1 - Governance & organizational viability

A 1.1 - Repository has a mission statement that reflects a commitment to the long-term retention of, management of, and access to digital information.

A 1.2 - Repository has an appropriate, formal succession plan, contingency plans, and/or escrow arrangements in place in case the repository ceases to operate or the governing or funding institution substantially changes its scope.

A 2 - Organizational structure & staffing

A 2.1 - Repository has identified and established the duties that it needs to perform and has appointed staff with adequate skills and experience to fulfil these duties.

A 2.2 - Repository has the appropriate number of staff to support all functions and services.

A 2.3 - Repository has an active professional development program in place that provides staff with skills and expertise development opportunities.

A 3 - Procedural accountability & policy framework

A 3.1 - Repository has defined its designated community(ies) and associated knowledge base(s) and has publicly accessible definitions and policies in place to dictate how its preservation service requirements will be met.

A 3.2 - Repository has procedures and policies in place, and mechanisms for their review, update, and development as the repository grows and as technology and community practice evolve.

A 3.3 - Repository maintains written policies that specify the nature of any legal permission required to preserve digital content over time, and repository can demonstrate that these permissions have been acquired when needed.

A 3.4 - Repository is committed to formal, periodic review and assessment to ensure responsiveness to technological developments and evolving requirements.

A 3.5 - Repository has policies and procedures to ensure that feedback from producers and users is sought and addressed over time.

A 3.6 - Repository has a documented history of the changes to its operations, procedures, software, and hardware that, where appropriate, is linked to relevant preservation strategies and describes potential effects on preserving digital content.

A 3.7 - Repository commits to transparency and accountability in all actions supporting the operation and management of the repository, especially those that affect the preservation of digital content over time.

A 3.8 - Repository commits to defining, collecting, tracking, and providing, on demand, its information integrity measurements.

A 3.9 - Repository commits to a regular schedule of self-assessment and certification and, if certified, commits to notifying certifying bodies of operational changes that will change or nullify its certification status.

A 4 - Financial sustainability

A 4.1 - Repository has short- and long-term business planning processes in place to sustain the repository over time.

A 4.2 - Repository has in place processes to review and adjust business plans at least annually.

A 4.3 - Repository's financial practices and procedures are transparent, compliant with relevant accounting standards and practices, and audited by third parties in accordance with territorial legal requirements.

A 4.4 - Repository has on-going commitment to analyse and report on risk, benefit, investment, and expenditure (including assets, licenses, and liabilities).

A 4.5 - Repository commits to monitoring for and bridging gaps in funding.

A 5 - Contracts, licenses & liabilities

A 5.1 - If repository manages, preserves and/or provides access to digital materials on behalf of another organization, it has and maintains appropriate contracts or deposit agreements.

A 5.2 - Repository contracts or deposit agreements must specify and transfer all necessary preservation rights, and those rights transferred must be documented.

A 5.3 - Repository has specified all appropriate aspects of acquisition, maintenance, access, and withdrawal in written agreements with depositors and other relevant parties.

A 5.4 - Repository tracks and manages intellectual property rights and restrictions on use of repository content as required by deposit agreement, contract, or license.

A 5.5 - If repository ingests digital content with unclear ownership/rights, policies are in place to address liability and challenges to those rights.

B - Digital Object Management

B 1 - Ingest: acquisition of content

B 1.1 - repository identifies properties it will preserve for digital objects.

B 1.2 - repository clearly specifies the information that needs to be associated with digital material at the time of its deposit (i.e., SIP)

B 1.3 - Repository has mechanisms to authenticate the source of all materials.

B 1.4 - Repository's ingest process verifies each submitted object (i.e., SIP) for completeness and correctness as specified in B 1.2.

B 1.5 - Repository obtains sufficient physical control over the digital objects to preserve them.

B 1.6 - repository provides producer/depositor with appropriate responses at predefined points during the ingest processes.

B 1.7 - repository can demonstrate when preservation responsibility is formally accepted for the contents of the submitted data objects (i.e., SIPs).

B 1.8 - Repository has contemporaneous records of actions and administration processes that are relevant to preservation (Ingest: content acquisition).

B 2 - Ingest: creation of the archival package

B 2.1 - repository has an identifiable, written definition for each AIP or class of information preserved by the repository.

B 2.2 - repository has a definition of each AIP (or class) that is adequate to fit long-term preservation needs.

B 2.3 - Repository has a description of how AIPs are constructed from SIPs.

B 2.4 - Repository can demonstrate that all submitted objects (i.e., SIPs) are either accepted as whole or part of an eventual archival object (i.e., AIP), or otherwise disposed of in a recorded fashion.

B 2.5 - repository has and uses a naming convention that generates visible, persistent, unique identifiers for all archival objects (i.e., AIPs).

B 2.6 - If unique identifiers are associated with SIPs before ingest, the repository preserves the identifiers in a way that maintains a persistent association with the resultant archived object (e.g., AIP).

B 2.7 - Repository demonstrates that it has access to necessary tools and resources to establish authoritative semantic or technical context of the digital objects it contains (i.e., access to appropriate international Representation Information and format registries).

B 2.8 - Repository records/registers Representation Information (including formats) ingested.

B 2.9 - repository acquires preservation metadata (i.e., PDI) for its associated Content Information.

B 2.10 - Repository has a documented process for testing understanding understandability of the information content and bringing the information content up to the agreed level of understanding.

B 2.11 - Repository verifies each AIP for completeness and correctness at the point it is generated.

B 2.12 - Repository provides an independent mechanism for audit of the integrity of the repository collection/content.

B 2.13 - Repository has contemporaneous records of actions and administration processes that are relevant to preservation (AIP creation).

B 3 - Preservation planning

B 3.1 - Repository has documented preservation strategies.

B 3.2 - repository has mechanisms in place for monitoring and notification when Representation Information (including formats) approaches obsolescence or is no longer viable.

B 3.3 - Repository has mechanisms to change its preservation plans as a result of its monitoring activities.

B 3.4 - Repository can provide evidence of the effectiveness of its preservation planning.

B 4 - Archival storage & preservation/maintenance of AIPs

B 4.1 - Repository employs documented preservation strategies.

B 4.2 - Repository implements/responds to strategies for archival object (i.e., AIP) storage and migration.

B 4.3 - Repository preserves the Content Information of archival objects (i.e., AIPs).

B 4.4 - Repository actively monitors integrity of archival objects (i.e., AIPs).

B 4.5 - Repository has contemporaneous records of actions and administration processes that are relevant to preservation (Archival Storage).

B 5 - Information management

B 5.1 - Repository articulates minimum metadata requirements to enable the designated community(ies) to discover and identify material of interest.

B 5.2 - Repository captures or creates minimum descriptive metadata and ensures that it is associated with the archival object (i.e., AIP).

B 5.3 - Repository can demonstrate that referential integrity is created between all archived objects (i.e., AIPs) and associated descriptive information.

B 5.4 - Repository can demonstrate that referential integrity is maintained between all archived objects (i.e., AIPs) and associated descriptive information.

B 6 - Access management

B 6.1 - Repository documents and communicates to its designated community(ies) what access and delivery options are available.

B 6.2 - Repository has implemented a policy for recording all access actions (includes requests, orders, etc.) that meet the requirements of the repository and information producers/depositors.

B 6.3 - Repository ensures that agreements applicable to access conditions are adhered to.

B 6.4 - Repository has documented and implemented access policies (authorization rules, authentication requirements) consistent with deposit agreements for stored objects.

B 6.5 - Repository access management system fully implements access policy.

B 6.6 - Repository logs all access management failures, and staff review inappropriate "access denial" incidents.

B 6.7 - Repository can demonstrate that the process that generates the requested digital object(s) (i.e., DIP) is completed in relation to the request.

B 6.8 - Repository can demonstrate that the process that generates the requested digital object(s) is correct in relation to the request.

B 6.9 - Repository demonstrates that all access request result in a response of acceptance or rejection.

B 6.10 - Repository enables the dissemination of authentic copies of the original or objects traceable to originals.

C - Technologies, Technical Infrastructure, & Security

C 1 - System infrastructure

C 1.1 - Repository functions on well-supported operating systems and other core infrastructural software.

C 1.2 - Repository ensures that it has adequate hardware and software support for backup functionality sufficient for the repository's services and for the data held, e.g., metadata associated with access controls, repository main content.

C 1.3 - Repository manages the number and location of copies at all digital objects.

C 1.4 - Repository has mechanisms in place to ensure any/multiple copies of digital objects are synchronize.

C 1.5 - repository has effective mechanisms to detect bit corruption or loss.

C 1.6 - Repository reports to its administration all incidents of data corruption or loss, and steps taken to repair/replace corrupt or lost data.

C 1.7 - Repository has defined processes for storage media and/or hardware change (e.g., refreshing, migration).

C 1.8 - Repository has a documented change management process that identifies changes to critical processes that potentially affect the repository's ability to comply with its mandatory responsibilities.

C 1.9 - Repository has a process for testing the effect of critical changes to the system.

C 1.10 - Repository has a process to react to the availability of new software security updates based on a risk-benefit assessment.

C 2 - Appropriate technologies

C 2.1 - Repository has hardware technologies appropriate to the services it provides to its designated community (ies) and has procedures in place to receive and monitor notifications, and evaluates when hardware technology changes are needed.

C 2.2 - Repository has software technologies appropriate to the services it provides to its designated community (ies) and has procedures in place to receive and monitor notifications, and evaluates when software technology changes are needed.

C 3 - Security

C 3.1 - Repository maintains a systematic analysis of such factors as data, systems, personnel, physical plant, and security needs.

C 3.2 - Repository has implemented controls to adequately address each of the defined security needs.

C 3.3 - Repository staffs have delineated roles, responsibilities, and authorizations related to implementing changes within the system.

C 3.4 - Repository has suitable written disaster preparedness and recovery plan(s), including at least one off-site backup of all preserved information together with an off-site copy of the recovery plan(s).

Anexo 2: Requisitos do Data Seal of Approval

1. The data producer deposits the research data in a data repository with sufficient information for others to assess the scientific and scholarly quality of the research data and compliance with disciplinary and ethical norms.
2. The data producer provides the research data in formats recommended by the data repository
3. The data producer provides the research data together with the metadata requested by the data repository
4. The data repository has an explicit mission in the area of digital archiving and promulgates it
5. The data repository uses due diligence to ensure compliance with legal regulations and contracts including, when applicable, regulations governing the protection of human subjects.
6. The data repository applies documented processes and procedures for managing data storage
7. The data repository has a plan for long-term preservation of its digital assets
8. Archiving takes place according to explicit workflows across the data life cycle
9. The data repository assumes responsibility from the data producers for access and availability of the digital objects
10. The data repository enables the users to utilize the research data and refer to them

11. The data repository ensures the integrity of the digital objects and the metadata
12. The data repository ensures the authenticity of the digital objects and the metadata
13. The technical infrastructure explicitly supports the tasks and functions described in internationally accepted archival standards like OAIS
14. The data consumer complies with access regulations set by the data repository
15. The data consumer conforms to and agrees with any codes of conduct that are generally accepted in higher education and research for the exchange and proper use of knowledge and information
16. The data consumer respects the applicable licenses of the data repository regarding the use of the research data

BIBLIOGRAFIA

- Bearman, D. (1989). *Archival Methods*. Pittsburgh: Archives and Museum Informatics.
- Besser, H. (2001). Digital Preservation of Moving Image Material? *The Journal of the Association of Moving Image Archivists*, 1(2), 39-55.
- Brody, T., Carr, L. & McSweeney, P. (2010, February 8). *Preservation support*. Retrieved January 19, 2012, from the University of Southampton Web site: http://wiki.eprints.org/w/Preservation_Support
- Brody, T., Carr, L., Hey, J. M. N., Brown, A. & Hitchcock, S. (2007, December). PRONOM-ROAR: Adding format profiles to a repository registry to inform preservation services. *International Journal of Digital Curation*, 2(2). Retrieved January 19, 2012, from: <http://www.ijdc.net/ijdc/article/view/53/>
- CASPAR Consortium. (2007). *CASPAR overall component architecture and componente model* (Report CASPAR-D1301-TN-0101-1_1). CASPAR Project. Retrieved January 19, 2012, from: http://www.casparpreserves.eu/Members/celrc/Deliverables/caspar-overall-component-architecture-and-component-model-1/at_download/CASPAR-D1301-TN-0101-1_1.pdf
- Chen, S.-S. (2001). The Paradox of Digital Preservation. *IEEE Computer*, 34(3), 24-28.
- Consultative Committee for Space Data Systems. (2002). *Reference model for an Open Archival Information System (OAIS)*. Blue Book. Also published as ISO 14721:2003. Retrieved from: <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- Crookes, D. (2010). British Library hints at videogame archiving plan. *The Independent*. London. Retrieved from <http://www.independent.co.uk/life-style/gadgets-and-tech/news/british-library-hints-at-videogame-archiving-plan-2125428.html>
- Davidson, A., & Pollard, A. (2005). Jasper - ZX Spectrum Emulator. Retrieved from <http://www.spectrum.lovely.net/>
- Digital Curation Centre, & DigitalPreservationEurope. (2007). *Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)*.
- Digital Preservation Testbed. (2001). *Migration: Context and Current Status*. The Hague.
- DSA Board. (n.d.). Data Seal of Approval. 2009. Retrieved October 7, 2011, from <http://www.datasealofapproval.org/>
- Ferreira, M., Baptista, A. A., & Ramalho, J. C. (2006). A Foundation for Automatic Digital Preservation. *Ariadne*, (48). Retrieved from <http://www.ariadne.ac.uk/issue48/ferreira-et-al/>
- Ferreira, M., Baptista, A. A. & Ramalho, J. C. (2009). *CRiB: Conversion and recommendation of digital object formats*. Retrieved January 19, 2012, from the RepositóriUM: <http://hdl.handle.net/1822/6195>
- Granger, S. (2000). Emulation as a Digital Preservation Strategy. *D-Lib Magazine*, 6(10).
- Hedstrom, M. (2001). Digital Preservation: Problems and Prospects. *Digital Library Network (DLnet)*, (20).
- Heminger, A. R., & Robertson, S. B. (2004). A Delphi Assessment of the Digital Rosetta Stone Model. *37th Annual Hawaii International Conference on System Sciences (HICSS'04)*. Big Island, Hawaii.

- Hendley, T. (1998). *Comparison of Methods & Costs of Digital Preservation*. West Yorkshire: British Library Research and Innovation Center.
- Heslop, H., Davis, S., & Wilson, A. (2002). *An Approach to the Preservation of Digital Records*. Canberra, Australia: National Archives of Australia.
- Hitchcock, S., Brody, T., Hey, J. M. N. & Carr, L. (2005). Preservation for institutional repositories: Practical and invisible. In *Proceedings of PV 2005: Ensuring long-term preservation and adding value to scientific and technical data*. Royal Society of Edinburgh, Edinburgh, UK. Retrieved January 19, 2012, from: <http://eprints.soton.ac.uk/18774/>
- Hitchcock, S., Brody, T., Hey, J. M. N. & Carr, L. (2007a). Digital preservation service provider models for institutional repositories: Towards distributed services. *DLib Magazine*, 13(5/6). Retrieved from: <http://www.dlib.org/dlib/may07/hitchcock/05hitchcock.html>
- Hitchcock, S., Brody, T., Hey, J. M. N. & Carr, L. (2007b). *Preservation metadata for institutional repositories: Applying PREMIS*. University of Southampton. Retrieved January 17, 2012, from: <http://preserv.eprints.org/papers/presmeta/presmeta-paper.html>
- Hitchcock, S., Brody, T., Hey, J. M. N. & Carr, L. (2007c). *Survey of repository preservation policy and activity*. University of Southampton. Retrieved January 15, 2012, from: <http://preserv.eprints.org/papers/survey/survey-results.html>
- Hitchcock, S., Hey, J. M., Brody, T. & Carr, L. (2007d). *Laying the foundations for repository preservation services: Final report from the PRESERV project*. JISC. Retrieved from: <http://eprints.ecs.soton.ac.uk/18147/>
- Hitchcock, S., Tarrant, D. & Carr, L. (2009). *Towards repository preservation services: Final report from the JISC Preserv 2 project*. JISC. Retrieved from: <http://eprints.ecs.soton.ac.uk/18148/1/preserv2-finalreport.pdf>
- Hunter, J. & Choudhury, S. (2006). PANIC: an integrated approach to the preservation of composite digital objects using Semantic Web services. *International Journal on Digital Libraries*, 6(2), 174–183. doi:10.1007/s00799-005-0134-z
- Joy Davidson. (2011). *ADS and the Data Seal of Approval – case study for the DCC*. Retrieved from <http://www.dcc.ac.uk/resources/case-studies/ads-dsa>
- Kenney, A. R., McGovern, N. Y., Entlich, R., Kehoe, W. R., & Olsen, E. (2003). Digital Preservation Management. *Implementing Short-term Strategies for Long-term Problems*. Cornell University Library. Retrieved from <http://www.library.cornell.edu/iris/tutorial/dpm/>
- King, R., Schmidt, R., Jackson, A., Wilson, C., & Steeg, F. (2009). The Planets Interoperability Framework: An Infrastructure for Digital Preservation Actions. In M. Agosti, J. Borbinha, S. Kapidakis, C. Papatheodorou, & G. Tsakonas (Eds.), *ECDL09 Proceedings of the 13th European conference on Research and advanced technology for digital libraries* (Vol. 5714/2009, pp. 425-428). Springer-Verlag. Retrieved January 17, 2012, from: http://planets-project.eu/events/bern-2009/pre-reading/docs/King_etc_InteroperabilityFramework.pdf
- Knight, G. & Anderson, S. (2007). *SHERPA DP: Final report of the SHERPA DP project*. JISC. Retrieved January 23, 2012, from: http://www.sherpadp.org.uk/documents/sherpadp_finalreport.rtf
- Krijgsman, G. (2005). Emulator Zone. Retrieved from <http://www.emulator-zone.com>
- Lavoie, B. F. (2004). *The Open Archival Information System Reference Model: Introductory Guide*. Dublin, USA: Digital Preservation Coalition.

- Lawrence, G. W., Kehoe, W. R., Rieger, O. Y., Walters, W. H., & Kenney, A. R. (2000). Risk Management of Digital Information: A file format investigation. Washington, DC: Council on Library and Information Resources.
- Lee, K.-H., Slattery, O., Lu, R., Tang, X., & McCrary, V. (2002). The State of the Art and Practice in Digital Preservation. *Journal of Research of the National Institute of Standards and Technology*, 107(1), 93-106.
- Microsoft Corporation. (1981). MS-DOS.
- Parallels. (1995). Parallels Desktop Web site. Retrieved from <http://www.parallels.com>
- Pearson, D. & Walker, M. (2007, November 30). *Report of the Format Notification and Obsolescence Service (AONS II)*. Australian Partnership for Sustainable Repositories. Retrieved January 23, 2012, from: <http://www.apsr.edu.au/aons2/report.pdf>
- RLG, NARA, & OCLC. (2007). *Trustworthy repositories audit & certification: Criteria & Checklist* (p. 94).
- Rauber, A., & Aschenbrenner, A. (2001). Part of Our Culture is Born Digital - On Efforts to Preserve it for Future Generations. *TRANS - On-line Journal for Cultural Studies*, 10.
- Rodrigues, M. de L. T. S. (2003). *Preservação Digital de Longo Prazo - Estado da arte e boas práticas em repositórios digitais*. Instituto Superior de Ciências do Trabalho e da Empresa.
- Rothenberg, J., Commission on Preservation and Access, & Council on Library and Information Resources. (1999). *Avoiding technological quicksand: finding a viable technical foundation for digital preservation: a report to the Council on Library and Information Resources* (p. vi, 35 p.). Washington, DC: Council on Library and Information Resources.
- Russell, K. (2000). Digital Preservation and the CEDARS Project Experience. *International Conference on Preservation and Long Term Accessibility of Digital Materials* (pp. 139-154). York, England.
- Saramago, M. de L. (2004). Metadados para preservação digital e aplicação do modelo OAIS. *VIII Congresso da BAD*. Estoril, Portugal.
- Strodl, S. & Becker, C. (2007, July 31). *Report on methodology for specifying preservation plans* (Deliverable PP4/D1). Planets Project. Retrieved January 17, 2012, from: http://www.ifs.tuwien.ac.at/dp/plato/docs/Planets_PP4-D1_Final.pdf
- Task Force on Archiving of Digital Information, Commission on Preservation and Access, & Research Libraries Group. (1996). *Preserving digital information: report of the Task Force on Archiving of Digital Information* (p. 59). Washington, D.C.: Commission on Preservation and Access.
- Thibodeau, K. (2002). Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years. In C. on L. and I. Resources (Ed.), *The State of Digital Preservation: An International Perspective*. Washington D.C.: Documentation Abstracts, Inc. - Institutes for Information Science.
- VMWare. (1998). VMWare Workstation Web site. Retrieved from <http://www.vmware.com/>
- Wagh, A., Wilkinson, R., Hills, B., & Dell'oro, J. (2000). Preserving Digital Information Forever. *Fifth ACM Conference on Digital Libraries* (pp. 175-184). San Antonio, Texas: Association for Computing Machinery.
- Webb, C. (2003). Guidelines for the Preservation of Digital Heritage. United Nations Educational Scientific and Cultural Organization - Information Society Division.

Wilczek, Eliot & Glick, Kevin (2006) Fedora and the Preservation of University Records Project: Reports and Findings, Tufts University and Yale University, Final Narrative Report to National Historical Publications and Records Commission, September 27, 2006 <<http://dca.tufts.edu/features/nhprc/reports/index.html>>.

Woodyard, D. (1998). Farewell my Floppy: a strategy for migration of digital information. Retrieved from <http://www.nla.gov.au/nla/staffpaper/valadw.html>

Woodyard, D. (2000). Digital Preservation: The Australian Experience. *Third Conference Digital Library: Positioning the Fountain of Knowledge*. Malaysia.