

Curso Preservação Digital

Rede Sudeste de Repositórios Institucionais

Andréa Gonçalves

ICICT/Fiocruz

24 nov 2020

Objetivos

1. Compreender a importância da preservação digital no contexto institucional e informacional;
2. Conhecer os processos básicos de preservação digital e documentação dos fluxos de trabalho relacionados;
3. Identificar os responsáveis, as funções e os recursos necessários para a implementação de ações de preservação digital;
4. Conhecer ferramentas de apoio aos processos de preservação digital.

O que é Preservação Digital?

A preservação digital constitui-se de um conjunto de ações que devem ser incorporadas em todo o ciclo de vida dos documentos digitais para assegurar sua longevidade e garantir que permaneçam disponíveis, recuperáveis e compreensíveis pelo tempo que se fizer necessário.

Adaptado de BULLOCK, 1999 e CONARQ, 2005

Coisas para ter em mente sobre a preservação digital

- Não é um projeto, é um programa
- Envolve tanto procedimentos técnicos como (principalmente) aspectos políticos e administrativos
- Não existe preservação digital sem a adoção de padrões
- Adoção pode ser incremental
- Ferramentas não são tudo! Concentre-se em estabelecer responsáveis, prioridades, fluxos de trabalho e cronogramas

Principais desafios para a preservação digital

- Falta de funções e responsabilidades definidas ao longo do ciclo de vida do conteúdo e ao longo do tempo
- Falta de política / procedimentos definidos
- Falta de compreensão administrativa / organizacional do valor e da complexidade da preservação digital
- Falta de adesão das partes interessadas e / ou alta administração
- Falta de diretrizes para determinar se você está “fazendo certo” e o que é “bom o suficiente”
- Modelos que não se enquadram nos fluxos de trabalho existentes
- Abordagens isoladas para gestão de objetos digitais

Qual o principal desafio para a preservação digital do seu repositório?

Acesse o site: www.menti.com

Use o código: 40 98 45 3

Elementos da Preservação Digital

1. Políticas, programas e planos
2. Modelos
3. Padrões
4. Estratégias e ferramentas
5. Planejamento e gestão

Políticas, programas e planos

Política	Estratégico	Estabelece os princípios e diretrizes orientadores das ações de preservação para os diferentes acervos e coleções científicos e culturais da instituição
Programa	Orientador	Define os padrões e requisitos mínimos que norteiam as ações de preservação digital na instituição
Plano	Prático	Descreve os procedimentos e práticas para a operacionalização das ações de preservação específicas para uma instituição ou coleção (por exemplo: repositório institucional)

O que é uma Política de Preservação Digital

- Documento que define as diretrizes e objetivos de uma instituição para a implantação de um programa de preservação digital
- Deve estar alinhados com os objetivos e com as políticas da instituição: digitalização, acesso, TI, investimentos, ...
- Guia para a organização e para as equipes
- Deve ser adequada ao longo do tempo
- Os produtores e os consumidores sabem o que esperar da instituição
- Deve abranger todos os elementos relacionados à PD

Fonte: GRACIO, 2019

Principais etapas para criar uma Política de Preservação Digital

- Definir o objetivo da política de PD
- Criar uma equipe multidisciplinar para elaborar a política
- Analisar políticas existentes
- Buscar informações com instituições que já tenham sua política
- Identificar os elementos necessários para a política
- Elaborar uma estrutura do documento
- Elaborar um rascunho
- Discutir o documento com as áreas e pessoas interessadas
- Aprovar a política institucionalmente
- Realizar revisões periódicas

Fonte: GRACIO, 2019



Projeto	Programa
Atividades pontuais	Ações e decisões recorrentes e contínuas
Objetivos imediatos	Visão de longo prazo
Experimentação	Manutenção
Financiamento específico	Sustentabilidade econômica
Independente	Cooperativo
Ferramentas	Fluxos de trabalho e infraestrutura

Fonte: Digital Preservation Network (DPN)

Tabela 1 - Elementos essenciais e itens desejáveis do Plano de Ação de Preservação Digital

Elementos	Itens desejáveis
Descrição do cenário institucional	<ul style="list-style-type: none">○ Políticas de preservação relacionadas○ Obrigações legais○ Contexto de preservação
Descrição da coleção	<ul style="list-style-type: none">○ Identificação e escopo○ Restrições organizacionais○ Restrições técnicas○ Requisitos do usuário
Avaliação de riscos	<ul style="list-style-type: none">○ Possíveis ameaças externas○ Probabilidade e extensão do risco○ Plano para mitigar os riscos
Planejamento de preservação	<ul style="list-style-type: none">○ Metas de preservação○ Estratégias de preservação avaliadas○ Decisão resultante, incluindo o raciocínio da decisão
Ações de preservação	<ul style="list-style-type: none">○ Seleção de componentes para preservação○ Integridade e autenticidade do conteúdo○ Controle de formatos de arquivo○ Requisitos mínimos de metadados○ Gestão de direitos○ Responsabilidades sobre as etapas ou ações○ Regras e condições para execução na coleção
Plano de sucessão	<ul style="list-style-type: none">○ Recriação de metadados e arquivos○ Planejamento de migração

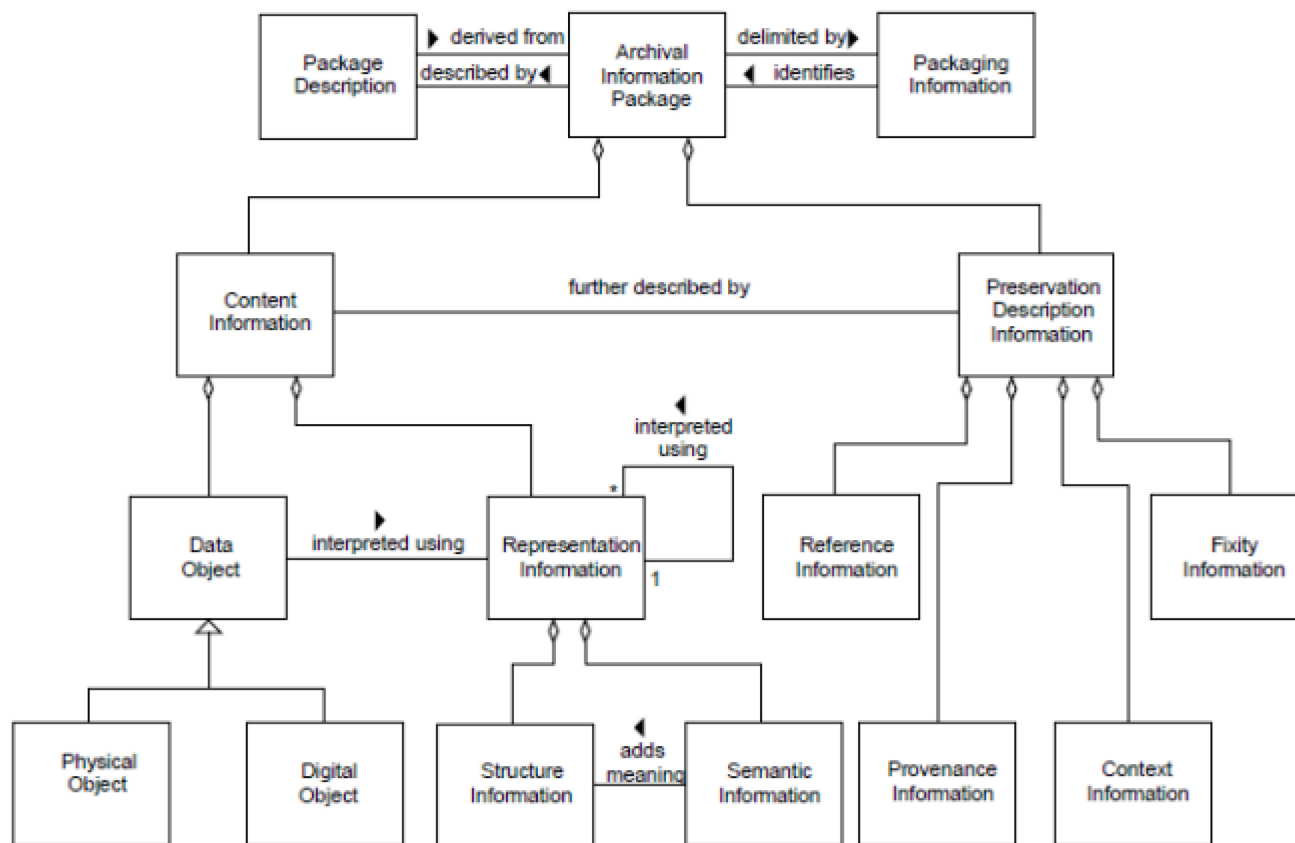
Fonte: NASCIMENTO; QUEIROZ; ARAÚJO, 2019

Modelos

1. Modelos

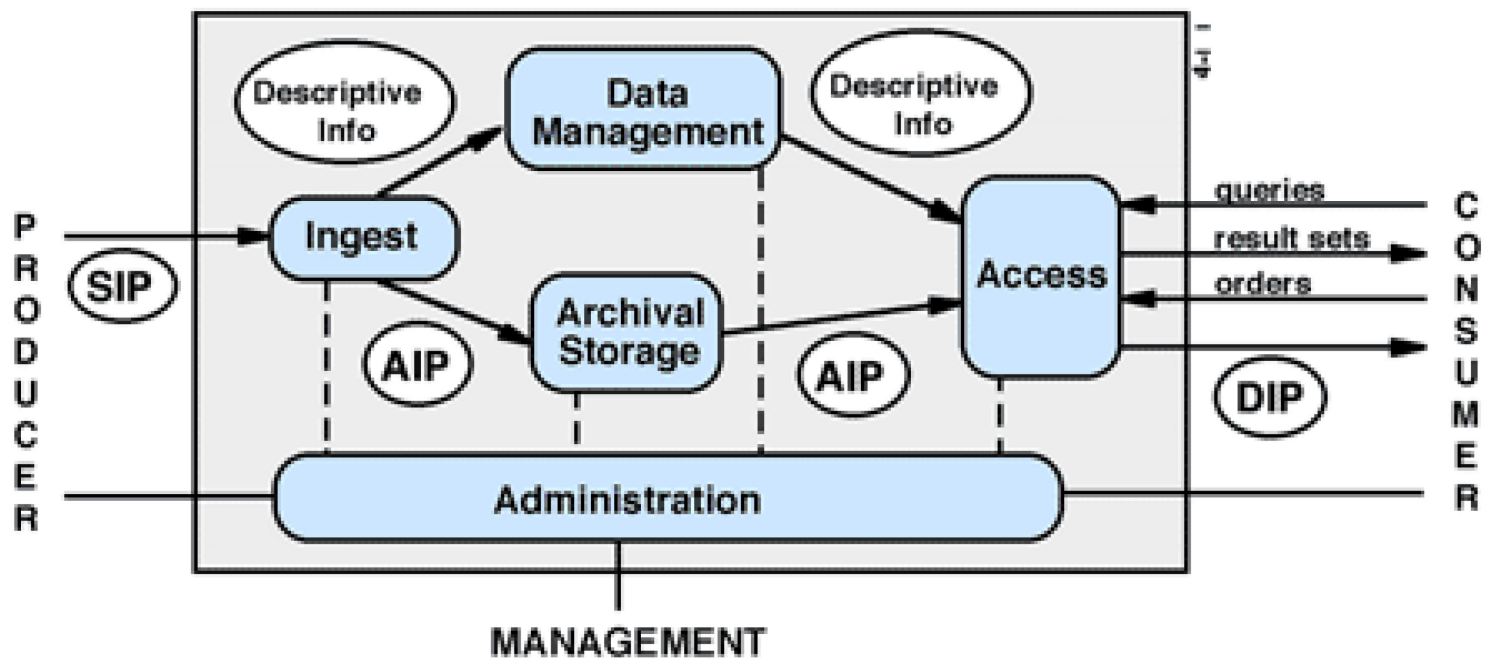
- ISO 14721: 2003 (OAIS – Open Archival Information System)
- NBR 15472:2007 (SAAI – Modelo de referência para um sistema aberto de arquivamento de informação)
- TRAC Certification (Trustworthy Repositories Audit & Certification)
- TDR ISO 16363 (Trustworthy Digital Repository ISO Standard)
- Diversos ciclos de vida de curadoria (que não se parecem com fluxos de trabalho reais)

OAIS Information Model



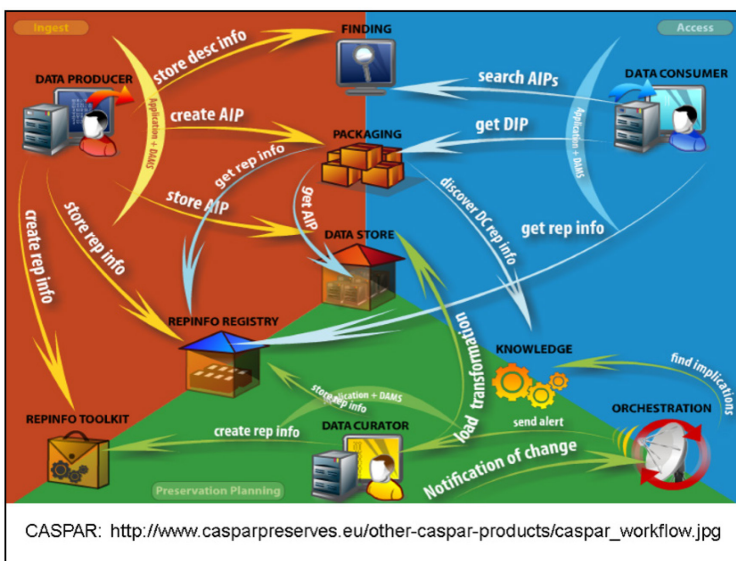
Fonte: Digital Preservation Coalition

OAIS Functional Entities

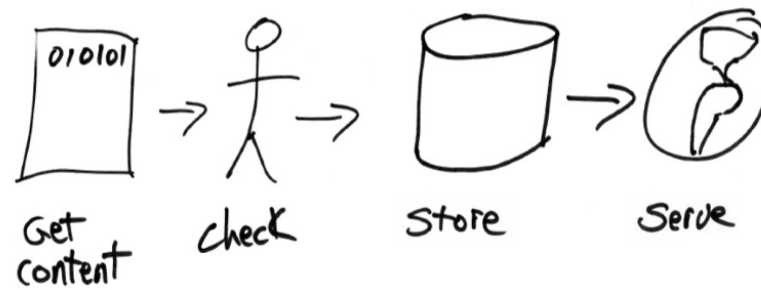


Fonte: NASA.gov

Expectativa



Realidade



Padrões

Para começar, conheça seu repositório!

1. Quantos registros o repositório possui?
2. Quanto espaço de armazenamento ocupa?
3. Qual a projeção de crescimento do repositório?
4. Existe uma lista completa de todos os objetos digitais do acervo?
5. Os objetos digitais seguem um padrão para nomenclatura de arquivos?
6. Quais tipologias documentais contém?
7. Que formatos de arquivo aceita? Quais formatos estão depositados?
8. Quais padrões de metadados são utilizados?
9. Qual o conjunto mínimo de metadados a ser preenchido para cada tipologia?
10. Os fluxos de trabalho do repositório estão documentados?

**Quantas dessas perguntas você consegue responder
sobre o seu repositório?**

Acesse o site: www.menti.com

Use o código: 40 98 45 3

2. Padrões

“Padronização é a base da preservação”
– *Miguel Arellano*

“Não basta ter padrões! Tem que aplicar!”
– *Luana Salles*

Padrões

Recomenda-se a adoção de **padrões** e protocolos abertos e interoperáveis e, sempre que possível, de **formatos de arquivo** abertos e não-proprietários amplamente aceitos, e de estruturas padronizadas de **metadados** descritivos e de preservação.

- **Formatos de arquivo**
 - ✓ Áudio, Dataset, Imagem, Texto, Vídeo
- **Metadados**
 - ✓ Administrativo, Descritivo, Estrutural, Preservação, Técnico

Formatos de arquivo aceitos pelo DSpace

pdf, xml, txt, asc, htm, html, css, doc, docx, ppt, pptx, xls, xlsx, jpeg, jpg, gif, png, tiff, tif, aiff, aif, aifc, au, snd, mp3, wav, mpeg, mpg, mpe, rtf, vsd, fm, bmp, psd, pdd, ps, eps, ai, mov, qt, mpa, abs, mpega, mpp, mpx, mpd, ma, látex, tex, dvi, sgm, sgml, wpd, ra, ram, pcd, ott, odt, oth, odm, odg, otg, odp, otp, ods, ots, odc, odf, odb, odi, oxt, sxw, stw, sxc, stc, sxd, std, sxi, sti, sxd, sxi, sti, sxg, sxm, sdw, sgl, sdc, sda, sdd, sdp, smf, sds, sdm, rdf, mp4, epub, zip

Exemplos de padrões de formato para arquivos

	Preservation Uncompressed, open, supported standards	Access In widespread use, open, small file-size for online hosting
Text	TXT, RTF, ODT, XML	DOC, PDF, ODT
Image	TIFF, PNG	JPEG, PNG
Audio	WAV, FLAC, AIFF	MP3, WMA, QuickTime
Video	MPEG-4, MJPEG 2000	MOV, AVI, WMV

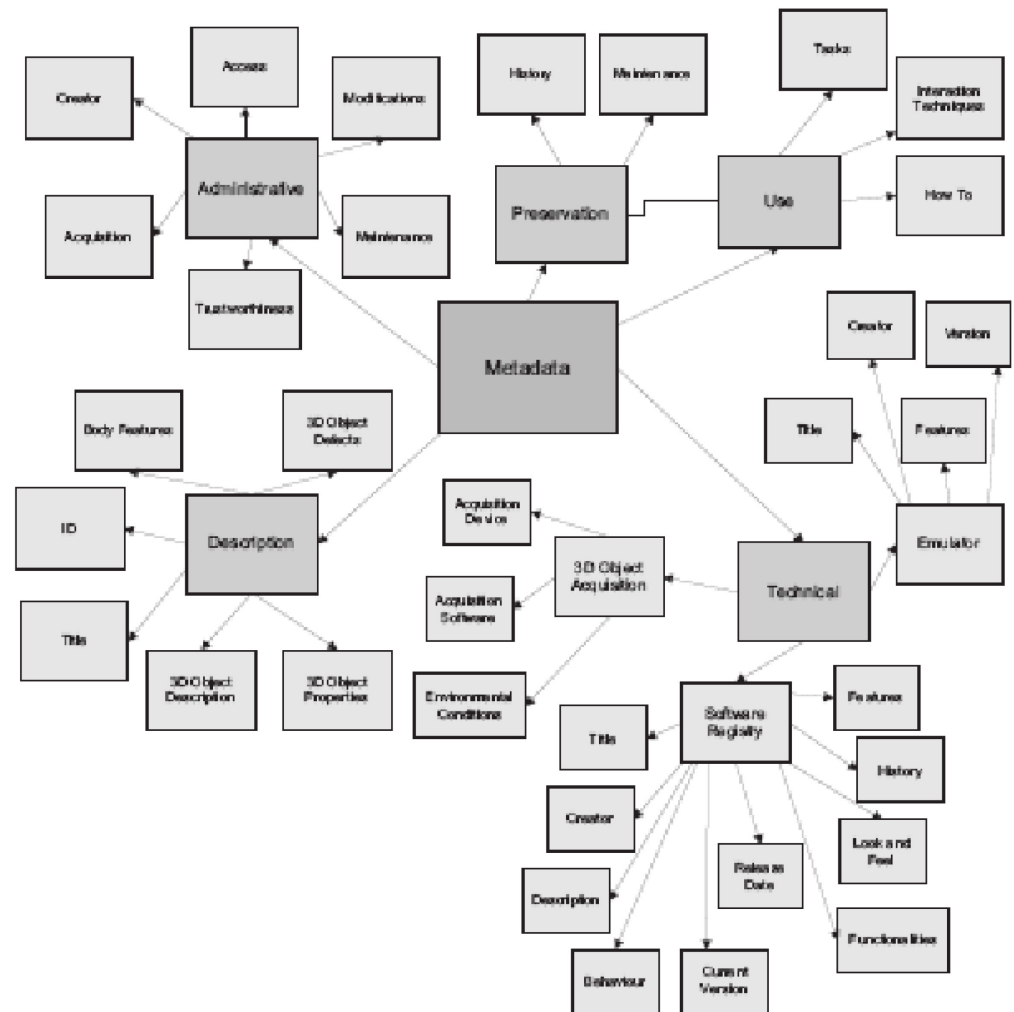
Fonte: Digital Curation Center

Tipos de metadados

- Administrativo
- Descritivo
- Estrutural
- Preservação
- Técnico

Qualidade é melhor que quantidade!

Adote um padrão e siga-o!



Exemplos de (alguns) padrões de metadados

- **Dublin Core (descritivo - geral)**
- Darwin Core (descritivo - coleções biológicas)
- EAD (arquivo)
- MARC21 (descritivo - bibliográfico)
- MODS (descritivo - bibliográfico)
- **METS (estrutural)**
- MIX (técnico - audiovisual)
- **PREMIS (preservação)**
- CCO (descritivo - objetos culturais)
- SPECTRUM (descritivo - museus)
- SCORM (estrutural)
- XML (estrutural, técnico)
- ...

“A melhor coisa sobre os padrões é que existem tantos para escolher!”

Estratégias e ferramentas

Dimensões da Preservação Digital

Física – suportes de armazenamento digital, compatibilidade e manutenção de dispositivos e mídias ópticas e magnéticas

Lógica – formatos de arquivo, cadeia de bits, hardwares e softwares.

Intelectual – conteúdo informacional do objeto digital, integridade e autenticidade, apresentação do documento

Administrativa – políticas, programas, planos e fluxos de trabalho

Adaptado de ARELLANO, 2004 e SANTOS, FLORES, 2017

Infraestrutura tecnológica

- Hardware
 - Digitalização
 - Datacenter (armazenamento e backup)
- Software
 - Repositório confiável
 - Ferramenta de gestão da preservação digital

3. Estratégias de preservação

As estratégias de preservação específicas para cada acervo dependem:

- do formato e da tipologia dos objetos digitais
- da infraestrutura e tecnologia existentes
- da disponibilidade de recursos humanos e materiais
- dos riscos identificados

É possível (e recomendável) adotar mais de uma estratégia de preservação.

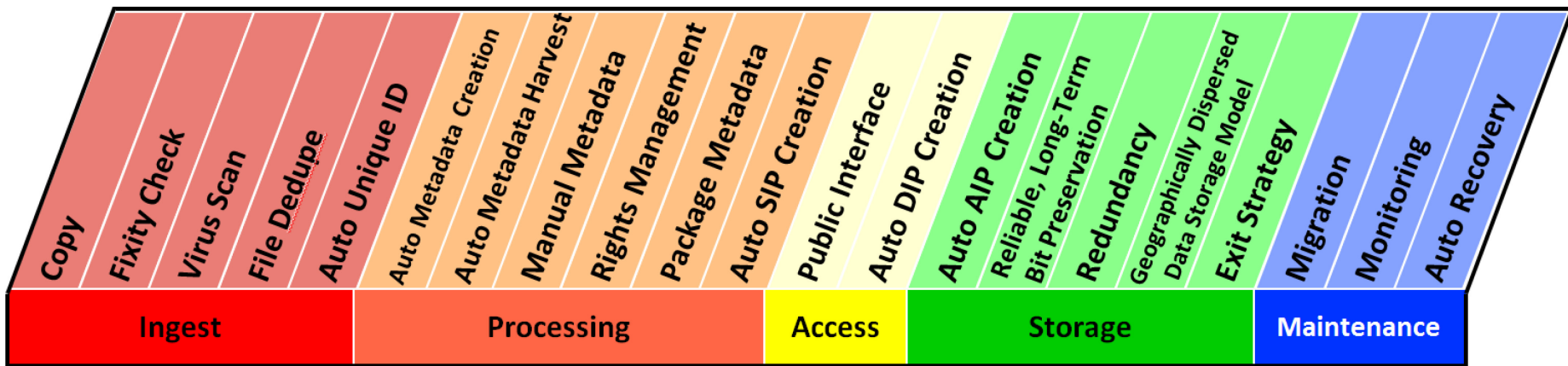
Estratégias de preservação comuns

- Migração de formatos de arquivo
- Emulação de ambiente de software
- Refrescamento de suportes e mídias
- Encapsulamento de tecnologias
- Preservação da tecnologia

Algumas coisas para ter em mente

- Nem todas as ferramentas e serviços são iguais. Começar com pouco é bom o suficiente!
- As opções de ferramentas não precisam ser definitivas. Defina o que você precisa agora, pensando no mais tarde.
- As novas ferramentas de hoje são as órfãs de amanhã. Concentre-se nos fluxos de trabalho!
- Adote padrões, documente tudo, e mantenha em dia!

Processos que precisam acontecer ou ser consideradas ao longo do caminho da preservação digital (segundo o modelo OAIS)...



Fonte: <http://digitalpowrr.niu.edu/tool-grid/>

Algumas funções de preservação em repositórios

- Escaneamento de vírus
- Cópia (*backup*)
- ID único
- Metadados manuais
- Verificação de fixidez (*checksum*)
- Criação de AIP

Funções de preservação em repositórios DSpace

- Exportação de pacotes de armazenamento de informação (AIP)
- Função nativa do Dspace
- Compatível com modelo OAIS e diversas ferramentas de preservação
- Permite exportar itens individuais, coleções, comunidades ou todo o conteúdo do repositório, incluindo a estrutura e perfis de usuário
- Armazenamento local ou remoto dos AIPs para preservação e restauração dos itens (metadados e objetos digitais) e estrutura do repositório
- AIPs podem ser utilizados como pacotes de submissão (SIP) a um sistema de preservação digital ou como pacotes de disseminação (DIP)

Porém...

Pacotes de armazenamento de informação (AIP) do Dspace não incluem:

- Verificação de formato
- Conversão de formatos
- Escaneamento de vírus
- Metadados de preservação
- Monitoramento de obsolescência

Ferramentas e sistemas de preservação digital

Existem ferramentas de front-end / processamento como...

Archivematica
Curator's Workbench
Data Accessioner
BitCurator

E existem serviços de armazenamento / preservação de back-end como...

MetaArchive
DuraCloud
Amazon Glacier
Fixity
Internet Archive

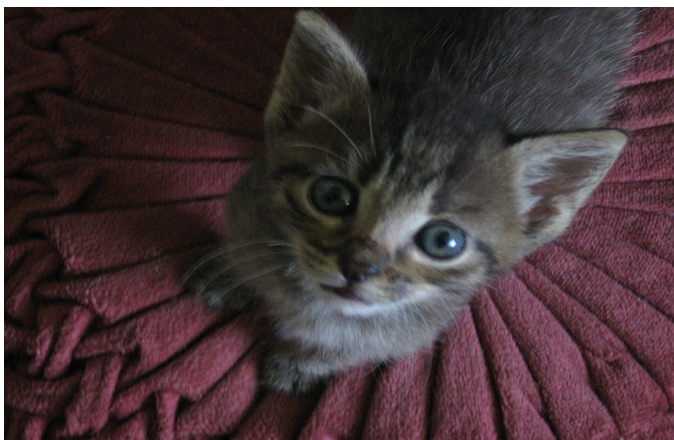


Existem até alguns serviços que fazem praticamente tudo como....

Preservica
Dspace Direct (uses DuraCloud)
ArchivesDIRECT

Habilidade técnica disponível + financiamento anual para preservação =
gama de ferramentas a considerar

O que quer dizer “gratuito”?



NÃO É



O software de código aberto requer recursos para instalar, manter e melhorá-lo.

Digital POWRR Tool Evaluation Grid	Ingest					Processing						Access		Storage					Maintenance			Other		
	Copy	Fixity Check	Virus Scan	File Dedupe	Auto Unique ID	Auto Metadata Creation	Auto Metadata Harvest	Manual Metadata	Rights Management	Package Metadata	Auto SIP Creation	Public Interface	Auto DIP Creation	Auto AIP Creation	Native Lockless Bitstream	Redundancy	ETDS	Exit Strategy	Migration	Monitoring	Auto Recovery	Open Source	Clear Documentation	Cost
**Archivematica	x	x	x		x	x	x	x	x	x		x	x					x			x	x	Free	
Archive-It	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x		x		x	x	Varies	
Chronopolis	x	x		x	x	x	x	x	x	x			x	x	x	x	x	x	x	x		x	\$1,500 - \$2,200	
**DuraCloud	x	x		x	x	x	x	x			x			x	x	x	x		x	x	x	x	Varies	
Fedora3/Islandora7	x	x			x	x	x	x	x	x	x	x	x	x	x	x			x		x	x	Free	
**MetaArchive (A private LOCKSS Network)	x	x		x	x			x		x			x	x	x	x			x	x		x	Varies	
Rosetta(Ex Libris)	x	x	x	x	x	x	x	x	x	x	x	x	x	x			x	x	x	x		x	Varies	
Preservica(Tessella)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	Varies	
Roda	x	x	x		x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	Free	

Fonte: POWRR Tool Grid (2013) - <https://digitalpowrr.niu.edu/digital-preservation-101/tool-grid/>

Planejamento e Gestão

Você conhece seu repositório? Parte 2

1. Como é feita a verificação de vírus e outras ameaças nos arquivos digitais?
2. Quem é o responsável por essa verificação?
3. Com que periodicidade são geradas as cópias de segurança (backup) do acervo?
4. Quantas cópias de segurança são geradas para cada objeto digital?
5. Onde são armazenadas as cópias de backup?
6. Quem são os responsáveis pelos procedimentos de backup?
7. Há algum plano de recuperação de desastres em caso de perda/corrupção dos objetos digitais?
8. Existem recursos alocados para a preservação digital do repositório a curto ou médio prazo?

Planejamento da preservação

1. Conheça seu repositório
2. Conheça seus fluxos de trabalho
3. Conheça os responsáveis por cada tarefa
4. Defina as suas prioridades e estratégias de preservação digital
5. Defina as tarefas, os responsáveis e a periodicidade das ações de preservação digital
6. Documente tudo!!!

Fluxos de trabalho e responsabilidades

- Descrição do passo-a-passo para realizar uma tarefa de modo que ela possa ser reproduzida por outra pessoa
- Determinar os responsáveis por cada passo
- Pode ser simples ou complexo
- Pode ser mudado ou melhorado com a experiência adquirida ou por novas circunstâncias e recursos
- Varia muito de acordo com o acervo, equipe, recursos, tecnologia, complexidade, etc.
- Deve estar conectado aos demais processos institucionais

Fluxos de trabalho e responsabilidades

Comece com o que você tem...

- “O que precisa ser feito?”
- “Quais são os recursos (humanos, infraestrutura, financeiros) necessários?”
- “Que recursos estão disponíveis?”

- Comece com um fluxo de trabalho simples. Teste, revise e melhore conforme necessário.

Fluxos de trabalho e responsabilidades

Use narrativas para criar um passo-a-passo do processo

Narrativa do fluxo de trabalho	Processo
<p>Após o descarregamento ou transferência dos conteúdos advindos de cartões de câmeras digitais, pendrive, FTP etc, os arquivos passam pelo processo de verificação de vírus e reconhecimento por meio da ferramenta XYZ para garantir a integridade do arquivo.</p>	<ol style="list-style-type: none">1. O responsável recebe a(s) mídia(s) com os arquivos.2. O responsável realiza a leitura da mídia para determinar quais arquivos estão incluídos na mídia.3. O responsável passa os arquivos pelo processo de verificação de vírus.4. O responsável abre todos os arquivos para verificar sua legibilidade.5. O responsável carrega os arquivos na ferramenta XYZ para gerar um checksum para cada arquivo recebido.

Exemplo de plano de ação

Tabela 5 - Ações, tarefas, periodicidade e responsabilidades do plano de preservação digital do repositório Arca

Ação de preservação	Tarefa	Periodicidade	Responsável
Validação de metadados	Verificar a presença de metadados mínimos em todos os registros ingressados no repositório	Contínuo	- NAACs - Bibliotecários das unidades - Bibliotecário/Bolsista Arca
	Realizar curadoria dos metadados de registros existentes	Contínuo	- Bibliotecário/Bolsista Arca
Padronização de formatos	Verificar se os formatos de arquivo depositados estão em conformidade com as diretrizes do Manual de Digitalização e do Plano Operativo do Arca	Contínuo	- NAACs - Bibliotecários das unidades - Bibliotecário/Bolsista Arca
Nomenclatura de arquivos	Verificar se o nome dos objetos digitais depositados está de acordo com o padrão previsto no Manual de Tratamento de Dados do Arca	Contínuo	- NAACs - Bibliotecários das unidades - Bibliotecário/Bolsista Arca
Verificação de antivírus	Executar programa antivírus em todos os objetos digitais depositados no repositório	Diária	- Equipe TI/ICICT
	Enviar alerta caso seja identificada a presença de vírus em algum arquivo	Quando aplicável	- Equipe TI/ICICT
	Revisar alerta de antivírus e providenciar substituição/remoção do arquivo infectado	Quando aplicável	- Bibliotecário Arca

Redes de cooperação

- LOCKSS / Cariniana
- MetaArchive Cooperative
- Open Preservation Foundation
- Portico

- Outras iniciativas de apoio
 - Arquivo Nacional
 - Digital Preservation Coalition
 - InterPARES
 - POWRR

Lembrando...

- A preservação digital não é um projeto, é um programa
- Envolve tanto procedimentos técnicos como (principalmente) aspectos políticos e administrativos
- Não existe preservação digital sem a adoção de padrões
- Adoção pode ser incremental
- Ferramentas não são tudo! Concentre-se em estabelecer responsáveis, prioridades, fluxos de trabalho e cronogramas

Qual a sua impressão sobre esse curso?

Acesse o site: www.menti.com

Use o código: **68 22 81 2**

Referências

- ARELLANO, Miguel Angel. Preservação de documentos digitais. *Ciência da Informação*, Brasília, v. 33, n. 2, p. 15-27, maio/ago. 2004. Disponível em: <http://www.scielo.br/pdf/ci/v33n2/a02v33n2.pdf>. Acesso em: 05 jan. 2019.
- BULLOCK, Alison. Preservation of digital information: issues and current status. 22 abr. 1999. Disponível em: <http://www.collectionscanada.ca/9/1/p1-259-e.html>. Acesso em: 28 jul. 2005.
- CONARQ. Carta para Preservação do Patrimônio Arquivístico Digital. 2005. Disponível em: http://www.conarq.gov.br/images/publicacoes_textos/Carta_preservacao.pdf. Acesso em: 10 abr. 2019.
- GRACIO, José Carlos Abbud. Políticas de preservação digital. In: BIREDIAL-ISTEC, 2019.
- NASCIMENTO, Andréa Gonçalves do; QUEIROZ, Claudete Fernandes de; ARAÚJO, Luciana Danielli de. Garantindo acervos para o futuro: Plano de preservação digital para o Repositório Institucional Arca. *Ciência da Informação*, v. 48, n. 3, 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4924>. Acesso em: 20 mar. 2020.
- SANTOS, Henrique Machado; FLORES, Daniel. Os impactos da obsolescência tecnológica frente à preservação de documentos digitais. *Brazilian Journal of Information Science: research trends*, v. 11, n. 2, 2017. Disponível em: <http://200.145.171.5/revistas/index.php/bjis/article/view/5550>. Acesso em: 11 mar. 2019.

Referências

Digital Preservation Handbook, 2nd Edition, 2015, disponível em <http://handbook.dpconline.org>

Digital Preservation Workflow Curriculum. Digital Preservation Network (DPN)
<http://www.dpn.org>

“From Theory to Action: A pragmatic approach to digital preservation strategies and tools”.
Digital POWRR, disponível em <https://digitalpowrr.niu.edu/>

Steps in a digital preservation workflow, 2012, disponível em
http://downloads.alcts.ala.org/ce/03072012_Digital_Preservation_Workflow_Slides.ppt



Instituto de Comunicação e Informação Científica e Tecnológica em Saúde

www.facebook.com/fiocruz.icict

[twitter.com/@Icict_fiocruz](https://twitter.com/Icict_fiocruz)

www.youtube.com/videosaudefio

www.icict.fiocruz.br