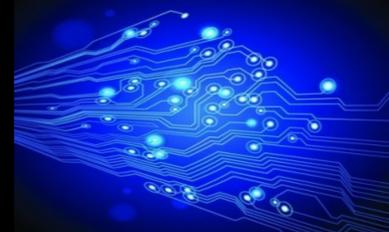


Infraestrutura de uma Rede de Preservação Digital Distribuída

Arthur Heleno Lima Rodrigues de Souza



Resumo

A cooperação entre instituições em prol da salvaguarda do patrimônio cultural em meio digital tem provado ser uma alternativa viável a iniciativas centralizadas, uma vez que portam melhores condições em integridade e segurança dos dados e ainda contam com redução dos custos operacionais (por instituição). No entanto, a infraestrutura distribuída de preservação digital possui uma maior complexidade e exige mais empenho das instituições envolvidas. Dentre estes desafios, é possível citar: a fragilidade intrínseca do armazenamento digital, a rápida obsolescência da tecnologia digital, a dificuldade em garantir a integridade, a autenticidade e o acesso dos documentos digitais. Desta forma, o presente estudo visa apresentar dados sobre os processos de implantação, manutenção, escolhas gerenciais, infraestrutura e protocolos de redes colaborativas. A pesquisa utiliza dados obtidos na Rede de Serviços de Preservação Digital - Cariniana, que possui vínculos de colaboração com diversas instituições brasileiras e internacionais e que adotou a ferramenta LOCKSS. O software foi desenvolvido pela Universidade de Stanford para gerar a infraestrutura lógica em preservação dos documentos digitais.

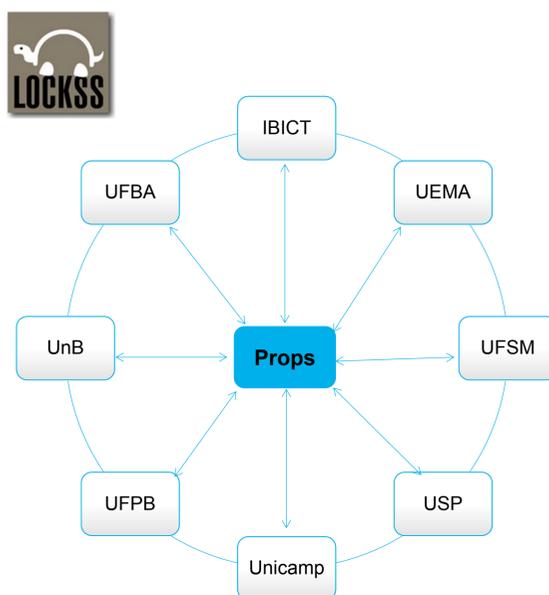
Introdução

A principal característica da preservação digital distribuída é o trabalho colaborativo e a dispersão geográfica dos servidores a fim de reduzir riscos e perdas de dados devido à falhas e acidentes. Desta forma, torna-se necessário firmar uma cooperação entre instituições para criar a rede de preservação geograficamente dispersa. A Rede de Serviços de Preservação Digital Cariniana hoje possui 8 servidores LOCKSS em rede coletando periódicos eletrônicos para a preservação. Há também o processo de tratamento dos documentos para prepara-los para preservação que acontece em parceria com a equipe LOCKSS da Stanford University e os editores. O sistema de preservação adotado foi desenvolvido para preservar periódicos publicados na ferramenta OJS/SEER, mas é capaz de integrar com outras ferramentas como o D-SPACE, utilizado como repositório de diversos tipos de conteúdo, como os livros eletrônicos do programa Livro Aberto.

Objetivos

- Apresentar a infraestrutura de uma rede de preservação digital distribuída;
- Informar sobre escolhas de ferramentas e dos objetos digitais candidatos a preservação;
- Estudar e calcular o valor médio de armazenamento por volume de periódico;
- Comparar resultados de hoje com o objetivo da rede;

Rede de Preservação Digital Distribuída (Cariniana)



Armazenamento de Dados

Uma das preocupação mais frequentes é o crescimento do armazenamento de uma rede preservação digital distribuída. Desta forma, conduziu-se uma pesquisa coletando valores estatísticos das caixas LOCKSS, ou servidores, e calculando a média de utilização de bytes por volume.

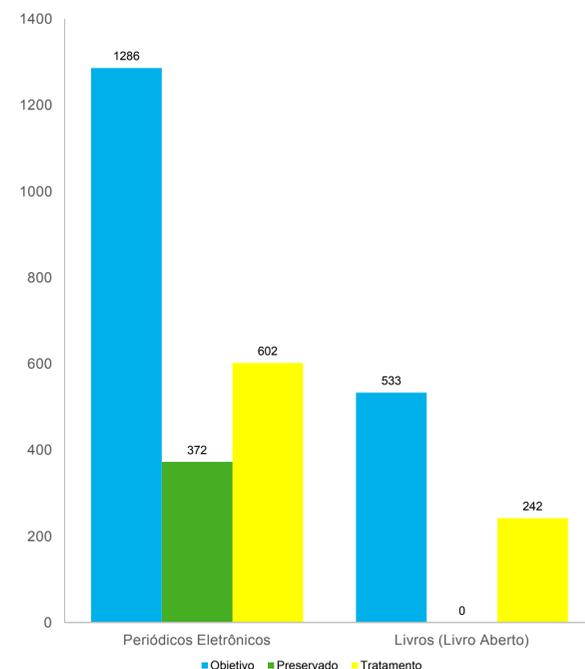
Adições	Volumes	Armazenam ento em disco	Armazenam ento / Volumes
1	15	345 MB	23 MB
2	82	1558 MB	19 MB
3	51	969 MB	19 MB
4	43	860 MB	25 MB
5	311	7151 MB	24 MB
6	593	13046 MB	24 MB
7	913	20999 MB	23 MB
Total	2008	45 GB	$\frac{157}{7} = 22,5$

O desvio padrão de 2,26 foi calculado para a média encontrada de 22,5 Megabytes. Desta forma podemos calcular o crescimento da rede para o total de volumes preservados juntamente com os volumes em tratamento.

$$X = (2708 + 6477) \cdot 22,5 = 206662,5$$

Segundo análise, a rede precisaria de pouco mais de 206 Gigabytes para armazenar esse conteúdo de 9185 volumes em 974 títulos.

Resultados



Hoje a rede preserva 372 títulos de periódicos, no total de 2708 volumes e trata 602 títulos com 6477 volumes. A rede tem como objetivo preservar todos 1286 títulos do Portal SEER e periódicos de instituições parceiras. Paralelamente, a rede pretende preservar todos os 533 títulos do Portal Livro Aberto e já iniciou o tratamento de 242 livros enquanto o desenvolvimento da ferramenta de integração LOCKSS D-SPACE é finalizada para iniciar os processos de coleta e preservação.

Considerações Finais

No intuito de alcançar os seus objetivos e preservar novos tipos de conteúdo, a iniciativa em preservação digital distribuída necessita de pesquisa constante, pois é uma nova área da ciência a ser desbravada.

Referências

- FERREIRA, Miguel. Introdução à preservação digital: conceitos, idéias e actuais consensos. Guimarães, Portugal: Escola de Engenharia da Universidade do Minho, 2006. 88 p. Disponível em: <<http://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>>. Acesso em: 13 mar. 2014.
- LOTS OF COPIES KEEP STUFF SAFE. About LOCKSS. In: LOCKSS. Stanford, CA: 2008a. Disponível em: <http://www.lockss.org/lockss/About_LOCKSS>. Acesso em: 20 fev. 2014.
- MÁRDERO ARELLANO, Miguel Angel. Critérios para a preservação digital da informação científica. 2008. 354 f. Tese (doutorado) - Universidade de Brasília, Departamento de Ciência da Informação, 2008. Disponível em: <http://bdtd.bce.unb.br/tesesimplificado/tde_busca/arquivo.php?codArquivo=4547>. Acesso em: 03 abr. 2014.
- TANENBAUM, Andrew S. Redes de computadores. 4ª. Ed. Elsevier: Editora Campus, 2003.
- SKINNER, K., SCHULTZ, M. A Guide to Distributed Digital Preservation. Atlanta, GA: Enducopia Institute, 2010. Disponível em: <<http://www.metaarchive.org/GDDP>>. Acesso em: 19 mar 2014.