



GUIA DE PRESERVAÇÃO DIGITAL DE PERIÓDICOS ELETRÔNICOS NA REDE CARINIANA

Instruções para Instituições usuárias, administradores e editores

EQUIPE IBICT

Brasília – DF
Novembro/2013

GUIA DE PRESERVAÇÃO DIGITAL DE PERIÓDICOS ELETRÔNICOS NA REDE CARINIANA

Instruções para Instituições usuárias, administradores e editores

Brasília – DF
Novembro/2013

Lista de Abreviaturas e Siglas

Archival Unit (AU)	Unidades de Arquivamento (UA). Uma coleção independente de conteúdo em uma caixa LOCKSS. Unidades de arquivamento são mantidas como um todo pelos processos do LOCKSS.
Arquivo PAC	Arquivo Proxy <i>Auto-Config</i> . O Arquivo de Configuração automática de Proxy define como os navegadores de internet irão automaticamente escolher o servidor proxy mais adequado.
Caixa LOCKSS	É um <i>software</i> rodando no servidor LOCKSS que armazena no disco os dados coletados. Às vezes referido como Cache LOCKSS ou Caixa LOCKSS.
Firewall	Elemento de segurança de uma rede de computadores que bloqueia o acesso não autorizado e permite comunicações autorizadas.
IP	<i>Internet Protocol</i> . O identificador numérico atribuído a cada computador ligado à internet.
LOCKSS	<i>Lot of Copies Keep Stuffs Safe</i> (Muitas Cópias Mantêm Coisas Seguras). Um <i>software</i> criado pela equipe LOCKSS da Universidade de Stanford, originalmente desenvolvido para a preservação de periódicos eletrônicos.
Manifest Page	Fornecer declaração de permissão para o LOCKSS rastrear e coletar uma Unidade de Arquivamento a partir de um conjunto de dados de um contribuinte.
OpenURL	Protocolo para a interoperabilidade entre uma fonte de informação e um componente de serviço, denominado como servidor de <i>links</i> contextuais, que oferece serviços específicos de informação.
PLN	Sigla para LOCKSS Private Network (Rede Privada LOCKSS). Trata-se de uma rede privada entre instituições a fim de preservar dados.

Sumário

1. Apresentação	9
2. Para quem?	9
3. Quais publicações participam?	9
4. Ações realizadas	10
5. Orientações	10
5.1 Como entrar	10
5.1.1 Identificação de conteúdo	11
5.1.2 Requisitos técnicos	12
5.1.2.1 Manifesto LOCKSS	12
5.1.2.2 Arquivos suplementares	14
5.1.2.3 Hyperlinks	14
5.2 Como permanecer	15
5.2.1 Versão do OJS na base URL	15
5.3 Como acessar	16
5.3.1 Proxy	16
5.3.2 OpenURL	17
6. Referências	18

1. Apresentação

O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), na condição de uma unidade de pesquisa integrante da estrutura do Ministério da Ciência Tecnologia e Inovação (MCTI), tem como missão promover a competência, o desenvolvimento de recursos e a infraestrutura de informação em ciência e tecnologia para a produção, socialização e integração do conhecimento científico-tecnológico.

A arquitetura técnica da Rede CARINIANA se baseou no modelo de preservação digital distribuída LOCKSS. Criado em 1998, o *software* LOCKSS foi produto de um projeto da Universidade de Stanford com o objetivo de auxiliar os bibliotecários na preservação dos documentos publicados na internet visando, principalmente, a continuidade de acesso a esses materiais (LOCKSS, 2008b).

O modelo de preservação digital distribuída consiste na distribuição de cópias de documentos digitais em locais geograficamente dispersos, algo que assegura a preservação, desde que sejam elaborados acordos e políticas entre as instituições parceiras.

Este guia tem por objetivo orientar as instituições usuárias, administradores de portais e editores de periódicos em relação aos procedimentos que comprometem a preservação digital dos portais de periódicos do Sistema Eletrônico de Editoração de Revistas – SEER/OJS, participantes da Rede CARINIANA.

2. Para quem?

Este guia está direcionado aos seguintes destinatários:

- **Instituição usuária** – mantém o portal de periódicos sob a sua responsabilidade. A instituição deverá ser informada sobre quaisquer mudanças ocorridas no portal de periódicos OJS/SEER;
- **Administrador do portal** – além de monitorar a rede local, contata a Rede CARINIANA caso seja detectado algum problema técnico em alguma das revistas do portal;
- **Editor** – responsável pelo periódico, e quem se comunica com o administrador do portal para informar sobre os volumes publicados, bem como sobre qualquer problema referente às edições.

3. Quais publicações participam?

A Rede CARINIANA tem como um dos seus objetivos é preservar as revistas eletrônicas que utilizam a plataforma OJS/SEER no Brasil. No ano de 2013 foram incluídos pela Coordenação da Rede todos os títulos cadastrados na página “Portais no SEER” do IBICT. O resultado dessa ação foi o registro dos volumes publicados até 2012 dos periódicos que constam na lista oficial da CARINIANA. A lista contém apenas os títulos que não apresentaram inconsistências nos dados publicados na internet. Caso a sua publicação não conste na lista oficial, recomendamos seguir as orientações contidas neste guia.

4. Ações realizadas

Durante o ano de 2013, a equipe do IBICT enviou listas de periódicos das instituições parceiras e que utilizam o Sistema Eletrônico de Editoração de Revistas (SEER) para a equipe do LOCKSS da Universidade de Stanford, a fim de obter subsídios para o tratamento adequado tocante à forma, à formatação e à configuração dos periódicos eletrônicos para sua preservação. A equipe do LOCKSS relatou diversos problemas que impediam que alguns periódicos eletrônicos fossem preservados na Rede. Nessa fase piloto, a equipe IBICT solicitou aos administradores dos portais de periódicos eletrônicos das instituições parceiras que atendessem as orientações a fim de solucionar os problemas detectados.

Esse trabalho gerou gargalos nos processos técnicos e, por este motivo, tornou-se necessário criar um documento que auxiliasse os administradores e editores a buscarem tais resoluções de forma mais autônoma.

5. Orientações

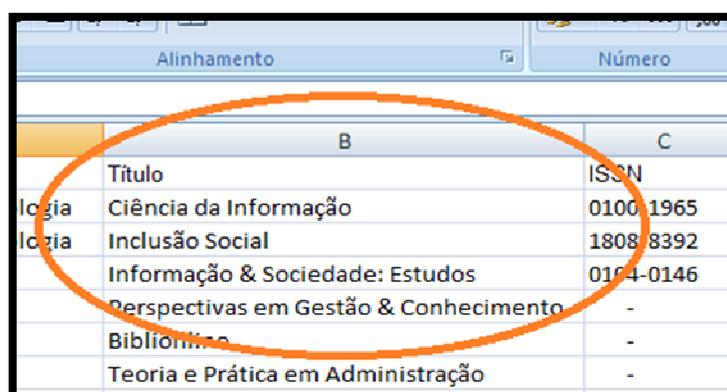
As ações aqui expostas deverão ser seguidas pelo Administrador do portal ou pelo Editor da revista a fim de padronizar as informações requeridas pela Rede CARINIANA para registro, permanência e recuperação dos conteúdos preservados.

5.1 Como entrar

Para que uma instituição forneça à Rede CARINIANA a lista de periódicos, o contato principal deve ser feito por intermédio do administrador com a Coordenação da Rede via correio eletrônico **cariniana@ibict.br**. O preenchimento incorreto dos dados resultará na não preservação dos artigos. Portanto, para submeter uma lista de periódicos para preservação é necessário seguir algumas orientações.

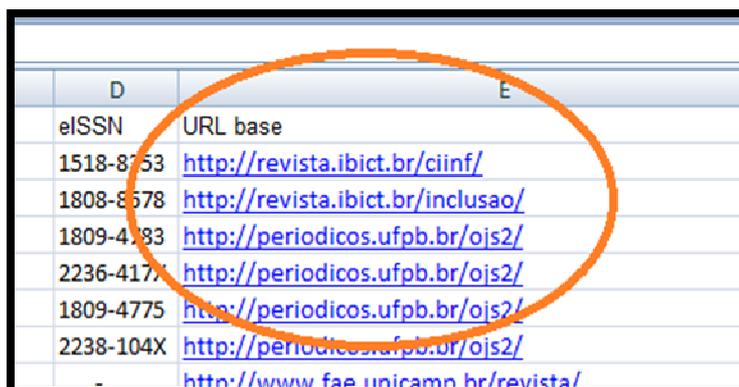
Para envio da lista de periódicos ou planilha de títulos (construída no Excel) tornou-se necessário o preenchimento destes campos: editora/Instituição, título, ISSN e eISSN, URL base, Identificador do periódico e anos disponíveis, conforme exemplificados nas figuras a seguir:

Campo Título



Alinhamento	Número	
B	C	
Título	ISSN	
logia	Ciência da Informação	0100-1965
logia	Inclusão Social	1808-8392
	Informação & Sociedade: Estudos	0104-0146
	Perspectivas em Gestão & Conhecimento	-
	Bibliônica	-
	Teoria e Prática em Administração	-

Campo URL base



D	E
eISSN	URL base
1518-8753	http://revista.ibict.br/ciinf/
1808-8678	http://revista.ibict.br/inclusao/
1809-4183	http://periodicos.ufpb.br/ojs2/
2236-4177	http://periodicos.ufpb.br/ojs2/
1809-4775	http://periodicos.ufpb.br/ojs2/
2238-104X	http://periodicos.ufpb.br/ojs2/
-	http://www.fae.unicamp.br/revista/

5.1.1 Identificação de conteúdo

EDITORIA/INSTITUIÇÃO

Deve ser informado o nome da instituição, departamento ou agência responsável pela publicação. Deve ser escrito por extenso.

TÍTULO DO PERIÓDICO

Corresponde ao nome da publicação e também deve ser redigido por extenso.

ISSN

Denomina-se o identificador de publicações seriadas no formato impresso. Há um hífen após os 4 (quatro) primeiros dígitos.

eISSN

É o identificador de publicações seriadas no formato eletrônico. Quanto ao preenchimento do campo, deve haver um hífen após os 4 (quatro) primeiros números.

* Pode ocorrer do ISSN e/ou eISSN não serem localizados. Neste caso, acrescenta-se um hífen "-" no campo correspondente.

URL BASE

É o endereço que aparece antes do índice '.php' nas URLs da publicação. O preenchimento deste campo deve ser realizado considerando os seguintes critérios:

- Eliminar a parte do *index.php* de todas as URLs;
- Os endereços devem terminar em /(barra);
- O endereço deve ser iniciado por HTTP

IDENTIFICADOR DO PERIÓDICO

É um código geralmente curto que está localizado imediatamente após o 'índice.php' nas URLs da publicação. É utilizado para distinguir os periódicos hospedados na mesma URL base.

ANOS DISPONÍVEIS

Correspondem aos períodos ou anos separados por vírgula e/ou hífen. É importante considerar os intervalos entre um período e outro.

Ex: 2000-2013;

2000-2003, 2005, 2007-2013

NOTAS E COMENTÁRIOS

Trata-se de um campo opcional onde se pode escrever alguma informação sobre a publicação que seja relevante.

Obs. Para que os periódicos possam ser, de fato, preservados é importante que todos estes campos sejam preenchidos. A falta de um dos dados impede que a lista seja enviada para a preservação. No caso dos ISSN e do eISSN, a publicação poderá se preservada caso seja informado pelo menos um dos dois.

5.1.2 Requisitos técnicos

Nesta seção serão abordados procedimentos e orientações técnicas de como preparar um periódico eletrônico para preservação na Rede.

5.1.2.1 Manifesto LOCKSS

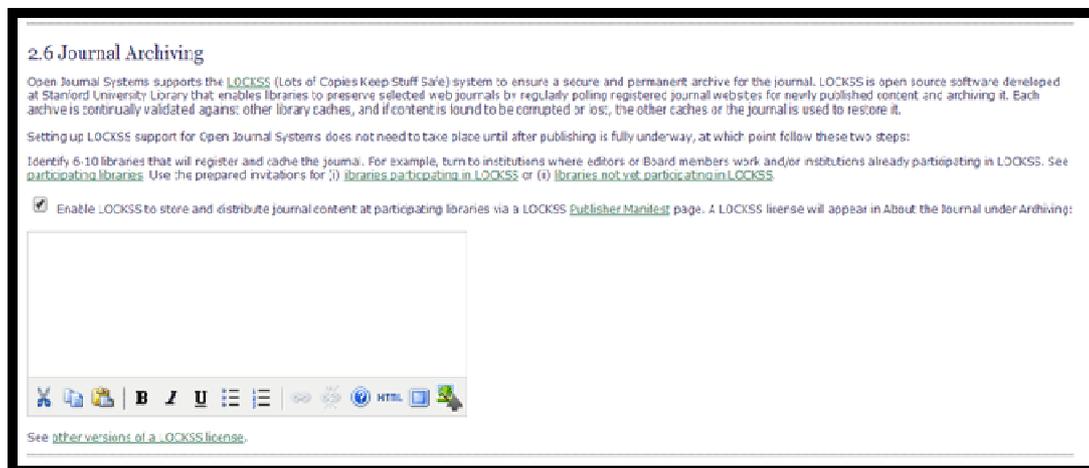
Para que um periódico seja preservado na Rede pelo sistema LOCKSS, é necessário adicionar um manifesto à revista, declarando conformidade com os processos de coleta e preservação adotados pelo *software*.

O sistema OJS/SEER é integrado ao sistema LOCKSS desde a versão 1.8. Para incluir este manifesto, o administrador deverá acessar a página de configuração (*setup*) do periódico desejado. A página, como no exemplo abaixo, será exibida:

Five Steps to a Journal Web Site

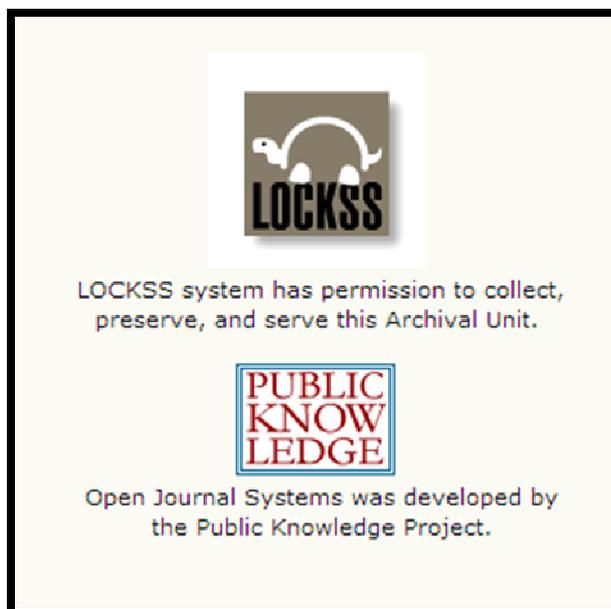
1. [Details](#)
Name of journal, ISSN, contacts, sponsors, and search engines.
2. [Policies](#)
Focus, peer review, sections, privacy, security, and additional about items.
3. [Submissions](#)
Author guidelines, copyright, and indexing (including registration).
4. [Management](#)
Access and security, scheduling, announcements, copyediting, layout, and proofreading.
5. [The Look](#)
Homepage header, content, journal header, footer, navigation bar, and style sheet.

Após o administrador clicar na segunda opção, “Políticas” (*Policies*), a tela apresentará várias opções. Na seção 2.6, como na imagem a seguir, será possível habilitar o manifesto LOCKSS (*LOCKSS MANIFEST*).



Uma vez habilitado, este manifesto poderá ser encontrado acessando o *link* da revista e adicionando “/gateway/lockss” ao final da URL.

Ex: <http://revista.instituicaouniversidade.br/engecom/index.php/ciinf/gateway/lockss>



Caso seja apresentada uma tela contendo as informações mostradas na última imagem, o manifesto foi devidamente habilitado para o periódico selecionado.

5.1.2.2 Arquivos suplementares

O formato OJS (SEER) permite adicionar *links* para o conteúdo suplementar em periódicos. No entanto, é possível criar o *link* de referência e não inserir o conteúdo referenciado. Essa prática pode impedir que o sistema LOCKSS colete os dados de forma coerente. O servidor da revista/periódico enviará uma "*flag*" de erro para o "rastreador" do sistema, e este entenderá que a Unidade de arquivamento (periódico) não está completa.

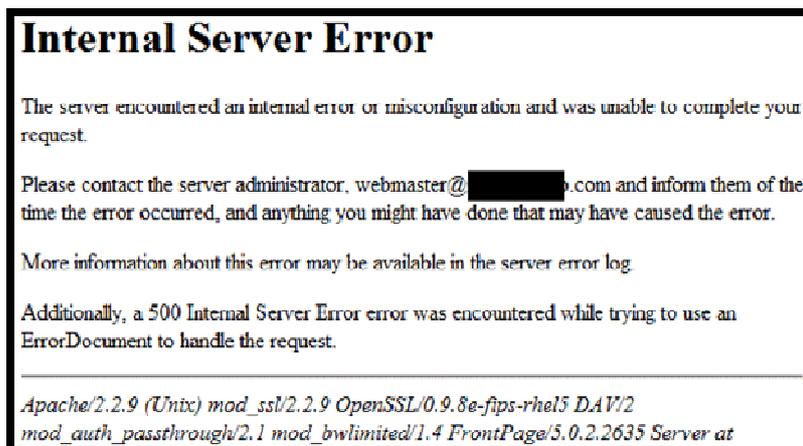
O administrador do portal deverá seguir as seguintes instruções para verificar a ocorrência deste problema:

Acesse os *links* dos arquivos suplementares, como no exemplo:

Ex: <http://universidade.revista.br/revistas/ojs/index.php/titutor/rt/suppFiles/2683/0>

Caso o servidor retorne o conteúdo desejado, o conteúdo foi referenciado corretamente.

Se a página de erro interno do servidor (500) for exibida, o conteúdo não foi devidamente referenciado, conforme mostra a seguinte figura:



Para a solução desta situação, ao inserir os *links* de referência ao conteúdo suplementar, certifique-se de que o conteúdo foi enviado. Caso não seja possível atualizar o periódico com seu conteúdo suplementar, desativar (remover) os *links*.

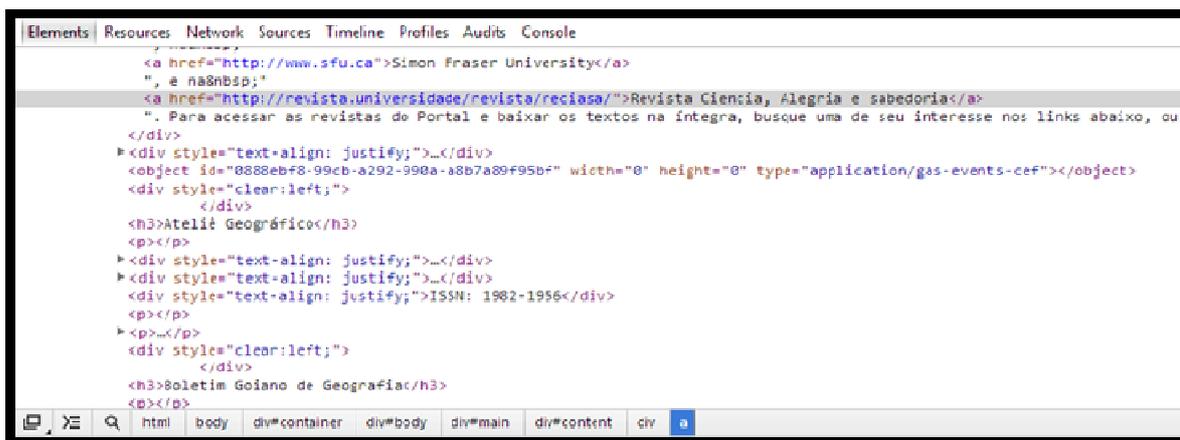
5.1.2.3 Hyperlinks

Um *hyperlink* (ou *link*) é uma referência aos dados que o leitor pode seguir diretamente, este *link* aponta para um documento inteiro ou para um elemento específico dentro de um documento. Ao se escrever o código do *link*, pode haver a supressão do protocolo http. Atualmente, isso não vem acarretando erros ou problemas para a leitura em navegadores atualizados. Porém, nas etapas de preservação, caso um periódico eletrônico referencie ele mesmo, o sistema de coleta entenderá essa referência como um documento

completo, entrando em uma função recursiva e gerando, por sua vez, um loop infinito. Dessa forma, a ferramenta de preservação digital irá suspender os processos da revista. O administrador do portal deverá seguir as seguintes instruções para verificar a ocorrência deste problema:

Verifique se existe algum *link* do periódico referenciando o próprio periódico:

- O código: `` não contém o protocolo "http://", e irá gerar a função recursiva de coleta.
- Identificado o problema, insira o protocolo antes da URL. Desta forma, o sistema entenderá que não se trata de uma revista dentro da revista, e sim da mesma revista.



```
Elements Resources Network Sources Timeline Profiles Audits Console
<a href="http://www.sfu.ca">Simon Fraser University</a>
", é a&nbsp;";
<a href="http://revista.universidade/revista/reciasa/">Revista Ciência, Alegria e sabedoria</a>
". Para acessar as revistas do Portal e baixar os textos na íntegra, busque uma de seu interesse nos links abaixo, ou
</div>
<div style="text-align: justify;"></div>
<object id="0888ebf8-99cb-a292-990a-a8b7a89f95bf" width="0" height="0" type="application/gas-events-cef"></object>
<div style="clear:left;">
</div>
<h3>Ateliê Geográfico</h3>
<p></p>
<div style="text-align: justify;"></div>
<div style="text-align: justify;"></div>
<div style="text-align: justify;">ISSN: 1982-1956</div>
<p></p>
<p></p>
<div style="clear:left;">
</div>
<h3>Boletim Goiano de Geografia</h3>
<h3></h3>
</div>
```

- Código correto: ``

5.2 Como permanecer

Com a finalidade de permanecer com a sua lista de periódicos, a instituição deve informar ao administrador qualquer mudança ocorrida. Quando um novo volume já estiver disponível, este deverá ser informado. Por exemplo, as publicações foram preservadas até o ano de 2012, pois os volumes referentes ao ano de 2013 não estão completos até o momento.

5.2.1 Versão do OJS na base URL

Os periódicos eletrônicos contêm metadados que devem permanecer estáticos para que não ocorram inconsistências em sua indexação e preservação. Um dos metadados é a URL base. Caso essa base seja modificada, os sistemas de indexação e preservação irão suspender os serviços.

Dessa forma, é importante certificar que a URL base não mude, mesmo quando é inserida a versão do OJS no endereço:

Ex.: `http://revista.instituicao.br/revistas/ojs-2.2.2/jorkgo/`

Existem duas orientações distintas para estes casos:

1. Retirar a versão antes dos processos de preservação e indexação, como:
<http://revista.instituicao.br/revistas/ojs/jorkgo/>
2. Certificar de que, caso ocorra atualização da ferramenta (OJS 3.0, por exemplo), a base URL continua com a mesma estrutura.

5.3 Como acessar

Os periódicos eletrônicos preservados na Rede CARINIANA podem ser recuperados. Para entender como uma instituição pode recuperar seus documentos, leia os próximos subitens.

5.3.1 Proxy

Uma das maneiras de se recuperar os dados preservados é por meio de uma configuração proxy no servidor da revista. Dessa forma, o servidor irá atender as requisições e poderá recuperar os dados em uma caixa LOCKSS quando a base de dados original não estiver disponível.

O primeiro passo para configurar o proxy em um portal de periódicos é acionar um administrador LOCKSS da Rede CARINIANA solicitando o acesso ao conteúdo via proxy.

O administrador LOCKSS abrirá exceções nas regras de segurança para o endereço IP do servidor do portal, e irá fornecer o código proxy para a configuração, como no exemplo:

```
// PAC file
// Generated Thu May 02 18:23:04 BRT 2013 by LOCKSS cache lockss.ibict.br

// Generated from /info/ProxyInfo
function FindProxyForURL(url, host) {
  // Educação Volume 2, Educação Volume 4, (14 more)
  if (shExpMatch(url, "http://cascavel.ufsm.br/*"))
  { return "PROXY lockss.ibict.br:8080; DIRECT"; }

  // Perspectivas em Gestão & Conhecimento Volume 1, Perspectivas em Gestão & Conhecimento Volume 2
  if (shExpMatch(url, "http://periodicos.ufpb.br/*"))
  { return "PROXY lockss.ibict.br:8080; DIRECT"; }

  // IBICT ELN Plugin Registry
  if (shExpMatch(url, "http://props.lockss.org:8001/*"))
  { return "PROXY lockss.ibict.br:8080; DIRECT"; }

  // Ciência da Informação Volume 1, Ciência da Informação Volume 2, (39 more)
  if (shExpMatch(url, "http://revista.ibict.br/*"))
  { return "PROXY lockss.ibict.br:8080; DIRECT"; }

  // ETD - Educação Temática Digital Volume 1 (1999), ETD - Educação Temática Digital Volume 1-2 (2000), (36 more)
  if (shExpMatch(url, "http://www.fae.unicamp.br/*"))
  { return "PROXY lockss.ibict.br:8080; DIRECT"; }

  return "DIRECT";
}
```

Este é um arquivo PAC e poderá ser configurado no servidor do Portal de periódicos eletrônicos. A forma como fazê-lo dependerá da solução adotada pela instituição.

5.3.2 OpenURL

Semelhante à solução abordada no subitem anterior, *OpenURL* referencia um conteúdo mediante um endereço de internet (URL) com parâmetros, possibilitando inseri-lo diretamente na página do portal.

Será necessário contatar um administrador LOCKSS da Rede CARINIANA para que este libere o acesso ao conteúdo para o IP da instituição e forneça o endereço com parâmetros, como por exemplo:

Ex: <http://lockss.xyz.edu:8082/ServeContent?issn=1553-3832&volume=6&issue=2&spage=3>

Esse *link OpenURL* de exemplo referencia a uma página de um volume de um periódico preservado, e pode ser utilizado pelo portal para fornecer o conteúdo ao usuário final.

6. Referências

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **OJS em uma hora**: uma introdução ao Sistema Eletrônico de Editoração de Revistas SEER/OJS versão 2.1.1. Trad. Por Ramón Martins Sodoma da Fonseca. Brasília: IBICT, 2006. 144p. Disponível em: <http://pkp.sfu.ca/files/ojs_em_uma_hora.pdf>. Acesso em: 25 nov. 2013.

LOTS OF COPIES KEEP STUFF SAFE. About LOCKSS. In:_____. **LOCKSS**. [S.l.], 2008a. Disponível em: <http://www.lockss.org/lockss/About_LOCKSS>. Acesso em: 18 nov. 2013.

_____. Home. In:_____. **LOCKSS**. [S.l.], 2008b. Disponível em: <<http://www.lockss.org/lockss/Home>>. Acesso em: 19 nov. 2013.

_____. How it works. In:_____. **LOCKSS**. [S.l.], 2008c. Disponível em: <http://www.lockss.org/lockss/How_It_Works>. Acesso em: 18 nov. 2010.

_____. For librarians. In:_____. **LOCKSS**. [S.l.], 2008d. Disponível em: <http://www.lockss.org/lockss/For_Librarians>. Acesso em: 18 nov. 2010.

_____. For publishers. In:_____. **LOCKSS**. [S.l.], 2008e. Disponível em: <http://www.lockss.org/lockss/For_Publishers#Permission_to_Librarians>. Acesso em: 20 nov. 2013.

SKINNER, Katherine; SCHULTZ, Matt. **A guide to distributed digital preservation**. Atlanta, GA: Educopia Institute, Feb. 2010.