

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

GT-8 – Informação e Tecnologia

ETAPAS PARA ELABORAÇÃO DE HEMEROTECAS DIGITAIS UTILIZANDO O OPEN HARVESTER SYSTEMS: O CASO DO HARVESTER IN LIBRARY AND INFORMATION SCIENCE

Pedro Ivo Silveira Andretta (Universidade Federal de Rondônia - UNIR)

Marcos Leandro Freitas Hübner (Universidade Federal de Rondônia - UNIR)

Wanderson Cássio Oliveira Araújo (Universidade Federal de Ceará - UFC)

Joliza Chagas Fernandes (Universidade Federal de Rondônia - UNIR)

Adriana Bruna Silva Albuquerque (Universidade Federal de Rondônia - UNIR)

STEPS FORWARDS IN THE PREPARATION OF DIGITAL NEWSPAPER LIBRARIES USING THE OPEN HARVESTER SYSTEMS: THE CASE OF HARVESTER IN LIBRARY AND INFORMATION SCIENCE

Modalidade da Apresentação: Pôster

Resumo: Este relato tem como tema a elaboração de hemerotecas digitais a partir do uso do Open Harvester Systems. O objetivo da pesquisa é apresentar as etapas empregadas na elaboração de uma hemeroteca digital, utilizando a tecnologia Harvester para a coleta e difusão dos registros de periódicos de acesso livre global em Ciência da Informação, com base na experiência com o Harvester in Library and Information Science (HiLIS). Em nossos apontamentos teóricos, abordamos conceitos sobre Open Archives Initiative, Protocol for Metadata Harvesting e Open Harvester Systems. Como metodologia, valemo-nos previamente de uma abordagem qualitativa, exploratória e bibliográfica e da pesquisa-ação para a implantação do HiLIS. Como resultados, apresentamos seis etapas: Instalação do OHS em rede; Elaboração de uma Política de Desenvolvimento de Coleções; Seleção de Fontes de Informação; Configurações para entrada de dados, agrupamento de metadados e categorizações; Entrada de registros de metadados e Customização da Plataforma. Em nossas considerações finais, indicamos ajustes na plataforma e também a possibilidade de utilizar os dados armazenados para lançar métricas sobre a rede de atores e interesses da Ciência da Informação, em perspectivas sincrônica e diacrônica.

Palavras-chave: Acesso Livre. Open Harvester System. Hemeroteca Digital – Procedimentos.

Abstract: The theme of this report is the preparation of digital newspaper libraries using the Open Harvester Systems. This study aims to present the steps used in preparation of a digital newspaper library, counting on the Harvester technology for the collection and diffusion of registers of the global open access periodicals in Information Science, using the experience from the Harvester in Library and

Information Science (HiLIS). In our theoretical observations we approached concepts of Open Archives Initiative, Protocol for Metadata Harvesting and Open Harvester Systems. Previously, we used a qualitative, exploratory, bibliographical and action-research methodology to implement the HiLIS. As a result, we presented 6 steps: the OHS Network Installation, preparation of a Collection Development Policy, Selection of Information Sources, Configurations for the data entry, cluster of metadata and categorizations, entry of metadata registers and Platform Customization. In our final considerations, we indicated adjustments in the platform and the possibility of using the stored data to release metrics about the Information Science actors network and interests, in synchronic and diachronic perspectives.

Keywords: Open Access. Open Harvester System. Digital Newspaper Library – Procedures.]

1 INTRODUÇÃO

O Acesso Livre – ou Acesso Aberto – tem proporcionado avanços para a comunidade científica, dinamizando as práticas de produção e circulação de materiais acadêmico-científicos. Junto a esse movimento, vemos a consolidação de sistemas eletrônicos para implementação e manutenção de periódicos e repositórios, protocolos para intercâmbio de registros e, ainda, bases de dados capazes de coletar, armazenar e recuperar conteúdos de diversas fontes de informação. São exemplos desse último caso os sistemas Harvester.

Os sistemas Harvester proporcionam uma difusão mais ampla dos conteúdos científicos selecionados, além de reduzir o tempo de levantamento bibliográfico ao pesquisador. São exemplos desses sistemas o GeoColeta¹, REA Paraná², Kollektin³, U.N.C. Library⁴, AsiaJOL⁵, além de outros próprios para a Ciência da Informação, como a Base Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI)⁶ e o Repertório da Produção Periódica Brasileira de Ciência da Informação (RPPBCI)⁷, os quais indexam artigos de publicações nacionais.

Nesse contexto, o presente relato tem como objetivo apresentar as etapas empregadas na elaboração de uma hemeroteca digital, valendo-se da tecnologia Harvester para a coleta e difusão dos registros de periódicos de acesso livre global em Ciência da Informação, a saber o Harvester in Library and Information Science (HiLIS)⁸. Destacamos, assim, a contribuição deste

1 GeoColeta. Disponível em: <<http://igeonidd.nuvem.ufrgs.br/geocoleta/home/index.php/index>>.

2 REA Parana. Disponível em: <<http://igeonidd.nuvem.ufrgs.br/geocoleta/home/index.php/index>>.

3 Kollektion. Disponível em: <<http://philo.at/kollektion/index.php/index>>.

4 U.N.C Library. Disponível em: <<http://bibliotecas.unc.edu.ar/index.php/>>.

5 Asian Journals Online. Disponível em: <<http://www.journal.acs.org.au/index.php/>>.

6 Base Referencial de Artigos de Periódicos em Ciência da Informação. Disponível em: <<http://www.brapci.ufpr.br/>>.

7 Repertório da Produção Periódica Brasileira de Ciência da Informação. Disponível em: <<http://bdpipe2.sibi.usp.br/metabuscaci/>>.

8 Harvester in Library and Information Science (HiLIS). Disponível em: <<http://www.lab.biblioteconomia.unir.br/hilis/>>.

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

trabalho em duas frentes: seu escopo e seu valor instrumental. A primeira relaciona-se diretamente com a possibilidade de melhorias às áreas de Arquivologia, Biblioteconomia e Documentação. Já a segunda, por fazer uso de uma ferramenta própria para coletar, organizar, recuperar e acessar a informação bibliográfica, é, em justa medida, do interesse dos profissionais da informação.

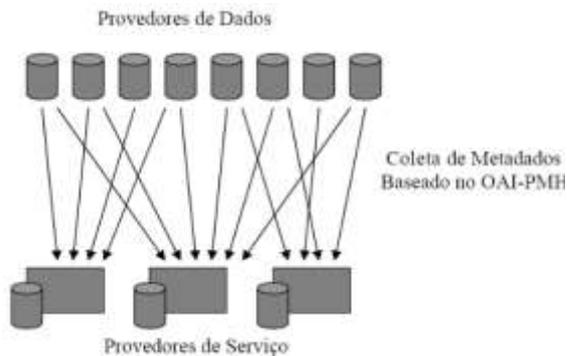
Nas próximas páginas, exporemos conceitos relativos ao Open Archives Initiative, ao Protocol for Metadata Harvesting e ao Open Harvester Systems para, em seguida, apresentar a metodologia e as etapas essenciais da elaboração de hemerotecas digitais com o Open Harvester Systems. É destaque no trabalho, além das etapas listadas, a Correspondência de metadados entre os formatos Dublin Core, MARC, MODS e ETD-MS, utilizados no Harvester in Library and Information Science – HiLIS e, por fim, as considerações para aplicações futuras.

2 OPEN ARCHIVES INITIATIVE, PROTOCOL FOR METADATA HARVESTING E OPEN HARVESTER SYSTEMS

A Open Archives Initiative (OAI) é uma organização centrada na Universidade de Cornell, nos Estados Unidos, patrocinada por diversas fundações voltadas a ciência e tecnologia, que desenvolve e promove padrões de interoperabilidade para repositórios com vistas a facilitar a disseminação eficiente de conteúdos e a comunicação científica. É também de sua responsabilidade a elaboração e popularização do Protocol For Metadata Harvesting (PMH). O OAI-PMH oferece interoperabilidade e extensibilidade que possibilitam compartilhar metadados dos repositórios para aplicações externas interessadas na coleta de dados.

O OAI-PMH fornece um quadro de interoperabilidade dividida em duas classes de participantes: provedores de dados e provedores de serviços. Os provedores de dados auxiliam na administração do sistema que suporta o protocolo de comunicação OAI-PMH; já os provedores de serviços utilizam os metadados colhidos por meio do protocolo de comunicação.

Figura 1: Diversos Provedores de Serviço realizando a coleta de metadados em vários Provedores de Dados



Fonte: Oliveira; Carvalho, 2009, p. 9.

Figura 2: Fluxo de informação nos Provedores de Serviços



Fonte: Oliveira; Carvalho, 2009, p. 9.

Os provedores de serviços, por meio de um sistema harvester, recolhem e utilizam os metadados de diversos repositórios e provedores de dados, desde que estes estejam em conformidade com o padrão OAI-PMH, de modo a prover condições de pesquisa aos usuários em uma única interface de busca. Madalli (2015) reitera que “a colheita de metadados depende principalmente da utilização das normas por provedores de dados. No entanto, por várias razões, gestores de repositórios, ao preencher as coleções, muitas vezes não seguem padrões globais de metadados” (MADALLI, 2015, p. 93, tradução nossa).

O Open Harvester Systems (OHS) é uma colheitadeira de metadados de código aberto e agregador desenvolvido pelo Public Knowledge Project (PKP). Este software utiliza o protocolo OAI-PMH, o que viabiliza a coleta dos registros bibliográficos em diferentes tipos de fontes e, assim, garante que repositórios de conteúdo semelhante possam ser interoperáveis.

3 METODOLOGIA

Para elaborarmos as etapas seguintes, empreendemos a princípio uma pesquisa que permitiu a implantação e implementação do HiLIS. Tal pesquisa se caracterizou, do ponto de vista de sua abordagem, como qualitativa; do ponto de vista do objetivo, como exploratória; e do ponto de vista dos procedimentos técnicos, como bibliográfica, a partir das considerações de Fernando e Hettiarachchi (2013) e do Public Knowledge Project (2014), além de utilizar também o método de pesquisa-ação.

4 RESULTADOS

As etapas para a elaboração de hemerotecas digitais utilizando o HiLIS foram:

Etapa 1ª – Instalação do OHS em rede. Nessa etapa, copiamos todos os arquivos do pacote OHS para o servidor da Universidade Federal de Rondônia por meio do *File Transfer Protocol* (FTP). Em seguida alteramos o arquivo “config.inc.php” com as informações do servidor e do banco de dados.

Etapa 2ª – Elaboração de uma Política de Desenvolvimento de Coleções. Para tanto, observamos as recomendações e diretrizes de Vergueiro (2010) e Weitzel (2013), a fim de elaborarmos uma Política de Desenvolvimento de Coleções própria.

Etapa 3ª – Seleção de Fontes de Informação. Nessa etapa, identificamos periódicos de Acesso Livre que utilizam⁹ o protocolo OAI-PMH. Em seguida, mapeamos os periódicos da área de Ciência da Informação ou de áreas transversais, por meio da Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), do Directory of Open Access Journals (DOAJ), da Scientific Electronic Library Online (SciELO), do Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal (Latindex) e da Library and Information Science Abstracts (LISA), além de páginas web especializadas em Biblioteconomia.

Etapa 4ª – Configurações para entrada de dados, agrupamento de metadados e categorizações (Browse). Nessa etapa, no sistema OHS, seguimos para a página *Crosswalks*, na qual discriminamos as correspondências entre os diferentes formatos de metadados, apoiados nos estudos empreendidos por Alves e Souza (2007), pela Library of Congress (2012) e no site The J. Paul Getty Trust¹⁰. Esse trabalho resultou na elaboração do Apêndice A - Correspondência de metadados entre os formatos Dublin Core, MARC, MODS e ETD-MS utilizando o Harvester in Library and Information Science (HiLIS). Na página *Sort Orders*, repetimos algumas correspondências entre os diferentes formatos de metadados para casos específicos, com o objetivo de estabelecer cinco categorias de busca Browse, a saber: *Title, Date, Publisher, Subject* e *Author*. Uma vez estabelecidas as configurações, pudemos iniciar a entrada de arquivos.

Etapa 5ª – Entrada de registros de metadados. Nessa etapa, retomamos a listagem de periódicos selecionados na Etapa 1. No sistema OHS, seguimos para a página *Add Archive*, na

9 Para isso substituímos a expressão “index” ao final do URL do periódico e adicionamos a combinação “/oai”.

10 The J. Paul Getty Trust. Disponível em: <http://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html>.

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

qual preenchemos os campos: *Title*, *URL*, *Type*, *OAI base URL*, *Index Method* e *Metadata Format* de cada provedor de dados, isto é, de cada periódico selecionado. Ainda no OHS, na página *Manage Archive*, escolhemos individualmente os conjuntos de dados que seriam importados de cada provedor de dados. Todos os conjuntos de dados dos periódicos foram importados para compor o HiLIS.

Etapa 6ª – Customização da Plataforma. No sistema OHS, seguimos para a página *Site Settings*, na qual preenchemos informações sobre Título, Página Inicial e Sobre e escolhemos o *Theme*. Em seguida, em *Layout*, ordenamos as caixas de recursos do sistema. Para o HiLIS optamos por criar um *Theme* minimalista.

5 CONSIDERAÇÕES FINAIS

Neste trabalho, apresentamos a elaboração de hemerotecas digitais com a utilização do OHS. Para tanto, valemo-nos de nossa experiência com o HiLIS e sintetizamos sua implantação em 6 etapas. Nas primeiras etapas, realizamos a instalação do sistema em rede, em conjunto com a elaboração da Política de Desenvolvimento de Coleções e a seleção de periódicos.

A configuração e caracterização do Harvester abordou aspectos da entrada de dados, agrupamento de metadados e categorizações, de modo que foi necessário estabelecer correspondências entre os elementos Dublin Core, MARC, MODS e ETD-MS. Concluídas as correspondências, demos início à entrada de arquivos, isto é, o processo de *harvesting*, ou coleta, dos registros dos periódicos. Nesse processo, notamos algumas falhas de prospecção, de modo que, dos 223 títulos selecionados, apenas 198 foram carregados pelo sistema. Por fim, seguimos para a customização da plataforma.

Conforme as caracterizações de Rowley (2012), o HiLIS se apresenta como uma base de dados referencial, e é particularmente uma base de dados bibliográficos. Vilan Filho e Burnier (1996) destacam que as bases de dados bibliográficas devem ter um objetivo claro. Nesses termos, em conjunto com uma Política de Coleções, elaboramos a Missão e o Objetivo do HiLIS. O Objetivo do HiLIS é facilitar a pesquisa e o acesso à produção científica para acadêmicos e profissionais, em especial, das áreas de Arquivologia, Biblioteconomia, Ciência da Informação, Documentação, Gestão do Conhecimento, Gestão da Informação e Museologia.

Destacamos como próximos passos: a tradução do sistema para português e espanhol e a resolução de eventuais problemas relativos aos metadados coletados. Acreditamos que os dados armazenados no HiLIS possam servir futuramente para lançar considerações métricas

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

sobre a rede de atores e interesses da Ciência da Informação, em perspectiva tanto sincrônica quanto diacrônica. Não descartamos, para uma experiência futura, migrar os dados e compreender os procedimentos para implementação em novas plataformas Harvester¹¹.

REFERÊNCIAS

ALVES, M. D. R.; SOUZA, M. I. F. Estudo de correspondência de elementos metadados: Dublin Core e Marc 21. **Revista Digital de Biblioteconomia & Ciência da Informação**, Campinas, v. 4, n. 2, p. 20-38, 2007. Disponível em: < <http://www.brapci.ufpr.br/brapci/v/7463> >. Acesso em: 27 mai. 2016.

FERNANDO, I. D. K. L; HETTIARACHCHI, N. Installation and Customization Experience of Metadata Harvester System: Case from University of Ruhuna. **Journal of the University Librarians Association of Sri Lanka**, Colombo, v.17, n. 1, January 2013. Disponível em: <<http://jula.sljol.info/articles/abstract/10.4038/jula.v17i1.6642/>>. Acesso em: 10 mai. 2016.

LIBRARY OF CONGRESS. MODS to Dublin Core Metadata Element Set Mapping Version 3. **Metadata Object Description Schema (MODS)**. August, 2012. Disponível em: < <http://www.loc.gov/standards/mods/mods-dcsimple.html> >. Acesso em: 06 jan. 2016.

MADALLI, D. P. Thematic harvesting of agricultural resources from generic repositories. **Information Processing in Agriculture**. Beijing Shi, v.2, n. 2, p. 93-100, September 2015, Disponível em: < <http://www.sciencedirect.com/science/article/pii/S2214317315000293/pdf?md5=aea4c553e633e6ae71fb3b87d00b34de&pid=1-s2.0-S2214317315000293-main.pdf>>. Acesso em: 04 jan. 2016.

OLIVEIRA, R. R.; CARVALHO, C. L. **Implementação de Interoperabilidade entre Repositórios Digitais por meio do Protocolo OAI-PMH**. Goiânia: Universidade Federal de Goiania. Instituto de Informática, 2009. (Relatório técnico 003-09). Disponível em < http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_003-09.pdf>. Acesso em: 30 jul. 2017.

PUBLIC KNOWLEDGE PROJECT. **Harvester2**: technical reference. p. 1-50, 2014. Disponível em: <<https://pkp.sfu.ca/wp-content/uploads/2014/04/TechnicalReference.pdf>>. Acesso em: 15 out. 2015.

ROWLEY, J. **A biblioteca eletrônica**. Tradução Antônio Agenor Briquet de Lemos. 2 ed. Brasília: Briquet de Lemos, 2012.

VERGUEIRO, W. **Seleção de materiais de informação**. 3. ed. Brasília: Briquet de Lemos, 2010.

VILAN FILHO, J. L.; BURNIER, S. Aspectos relevantes para a construção e uso de bases de dados bibliográficos. **Revista de Biblioteconomia de Brasília**. Brasília, v. 18, n. 2, p. 167-176,

¹¹ Damos atenção especial ao Omeka e à seção *Ferramentas* do Portal Open Archives <<https://www.openarchives.org/pmh/tools/tools.php>>.

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

jul./dez. 1990. Disponível em <

<http://basessibi.c3sl.ufpr.br/brapci/index.php/article/download/17704>>. Acesso em: 30 jul. 2017.

WEITZEL, S. R. Elaboração de uma política de desenvolvimento de coleções em bibliotecas universitárias. 2. ed. Rio de Janeiro: Interciência, 2013.

APÊNDICE A

Correspondência de metadados entre os formatos Dublin Core, MARC, MODS e ETD-MS utilizados no Harvester in Library and Information Science – HiLIS.

DUBLIN CORE	MARC	MODS	ETD-MS
Abstract/Description	Abstract	Abstract	Abstract
Contributor	-	Name/NamePart	Contributor
Coverage	Geographic Area Code/ Time Period of Content/ Subject Added Entry - Topical Term/ Subject Added Entry - Geographic Name/ Added Entry - Hierarchical Place Name	Geographic/ Temporal/ Hierarchical Geographicq/ cartographics	Coverage
Creator	Personal Name/ Corporate Name/ Meeting Name/ Added Entry - Personal Name/ Added Entry - Corporate Name/ Added Entry - Meeting Name/ Added Entry - Uncontrolled Name.	Name/ Name Part/ Display Form/ Affiliation/ Name Description/ Role	Creator
Date	Publication, Distribution, Etc/ Reproduction Note.	Date Issued/ Date Created/ Date Created/ Date Valid/ Date Modified/ Copyright Date/ Other Date	Date
Description	Language Code/ General Note/ With Note/ Dissertation Note/ Formatted Contents Note/ Summary, etc.	Note12/ Table of Contents/ Abstract	Description
Format	Physical Description/ Physical Medium/ Reproduction Note/ Electronic Location and Access.	Form/ Internet Media Type/ Extent	Format
Identifier	ISBN/ ISSN/ Summary, etc./ Electronic Location and Access.	Identifier/ Location	Identifier
Language	-	Language Term	Language
Publisher	Publication, Distribution, Etc.	Publisher	Publisher
Rights	Restrictions on Access Note/ Terms Governing Use and Reproduction Note.	Access Condition	Rights
Source	Source Of Acquisition/ Cataloging Source	-	Source

¹² Nota: dos registros utilizados para a composição dos metadados de descrição, este foi o único que não pode ser selecionado em razão da notificação de erro de duplicata.

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

Subject	Library of Congress Call Number/ Universal Decimal Classification Number/ Dewey Decimal Classification Number/ Subject Added Entry - Topical Term/ Uncontrolled Index Term.	Classification/ Topic/ Occupation	Subject
Title	Language Code/ Uniform Title (130)/ Abbreviated Title/ Uniform Title (240)/ Translation of Title By Cataloging Agency/ Title Statement/ Varying Form of Title/ Series Statement - Added Entry Title/ Series Statement/ Language Note/ Added Entry - Uncontrolled Related/Analytical Title.	Title/ Subtitle/ Part Number/ Part Name/ Non-Sorting Title	Title
Type	Leader/ Control Number/ Control Number Identifier/ Latest Transaction/ Additional Material Characteristics/ Physical Description/ Fixed-Length Data Elements/ Control Number/ Patent Control Information/ National Biography Number/ National Bibliographic Agency Control Number/Copyright or Legal Deposit Number.	Type Of Resource/ Genre	Type

Fonte: Dados da pesquisa.