

**LUCIANA MARA SILVA**

**REPOSITÓRIO INSTITUCIONAL COMO SISTEMA TÉCNICO-  
SOCIAL: COMPOSIÇÃO, AMBIENTE E ESTRUTURA**

Dissertação apresentada ao programa de Pós-Graduação em Ciência da Informação, do Centro de Ciências da Educação, da Universidade Federal de Santa Catarina, como requisito parcial para obtenção do título de Mestre em Ciência da Informação, área de concentração: Gestão da Informação, linha de pesquisa: Fluxos de Informação.

Orientador: Professor Dr. Vinícius Medina Kern

Florianópolis – SC  
2014

Catálogo na fonte por: Carolini da Rocha – CRB: 14/1215.  
Revisão língua portuguesa: Heloisa Costa

Silva, Luciana Mara, 1975-;

Repositório institucional como sistema técnico-social:  
composição, ambiente e estrutura / Luciana Mara Silva. – 2014.  
149 f. : il. color. ; 30 cm

Dissertação (mestrado) – Universidade Federal de Santa Catarina,  
Programa de Pós-graduação em Ciências da Informação.  
Orientação: Prof. Vinicius Medina Kern.

1. Repositórios institucionais. 2. Comunicação científica. 3.  
Acesso aberto. 4. Sistemismo. 5. Sistemas de Informação. I. Kern,  
Vinicius Medina. II. Universidade Federal de Santa Catarina. III.  
Título.

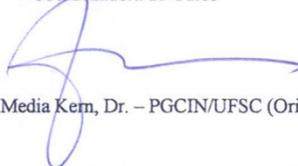
**LUCIANA MARA SILVA**

**REPOSITÓRIO INSTITUCIONAL COMO SISTEMA TÉCNICO-  
SOCIAL: COMPOSIÇÃO, AMBIENTE E ESTRUTURA**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação do Centro de Ciências da Educação da Universidade Federal de Santa Catarina em cumprimento ao requisito parcial para a obtenção do título de Mestre em Ciência da Informação.

**APROVADA PELA COMISSÃO EXAMINADORA  
EM FLORIANÓPOLIS, 28 DE FEVEREIRO DE 2014.**

Profª. Marisa Brascher Basilio Medeiros, Dra.  
Coordenadora do Curso



Prof. Vinicius Media Kern, Dr. – PGCIN/UFSC (Orientador)



Profª. Maria Cristina Soares Guimarães, Dra. – FIOCRUZ/ ICICT



Profª. Gleisy Regina Bóries Fachin, Dra. - CIN/UFSC



Profª. Magda Teixeira Chagas, Dra. - PGCIN/UFSC



Dedico este trabalho ao meu filho,  
Bruno Silva de Souza, minha eterna inspiração.



## AGRADECIMENTOS

A Deus, por me “carregar no colo” nos momentos em que mais precisei e me fazer acreditar que eu chegaria ao fim.

Aos meus familiares, em especial aos meus pais, que contribuíram para a minha formação ética, profissional e acadêmica.

Ao meu filho Bruno Silva de Souza, por compreender os momentos de ausência e me recompensar com seu eterno carinho.

À Universidade Federal de Santa Catarina (UFSC) e ao Programa de Pós-Graduação em Ciência da Informação (PGCIN), pela oportunidade.

Aos professores do Programa de Pós-Graduação em Ciência da Informação (PGCIN) pelo aprendizado profissional e acadêmico que as suas disciplinas me proporcionaram.

À turma 2011 do Mestrado em Ciência da Informação pela troca de informações, conhecimento, angústias e alegrias. Em especial a Eliane Mota Orel, Lidiana Sagaz Silva e Paula Balbis por compartilhar comigo os últimos momentos de alegrias e angústias.

Ao mestrando Robson Garcia Formoso, pela contribuição no grupo de pesquisa deste trabalho.

Aos membros da banca examinadora, professoras Dra. Gleisy Regina Bóries Fachin, Dra. Magda Teixeira Chagas e Dra. Maria Cristina Soares Guimarães pelo aceite e contribuições a esta pesquisa.

Ao meu orientador, Prof. Vinícius Medina Kern, pela dedicação, parceria, ensinamentos, compreensão, paciência nas horas de angústia, e por ser este competente educador.

Ao Diretor de Campus da Unisul Grande Florianópolis, Prof. Hercules Nunes de Araújo e a Gerente de Operações e Serviços Acadêmicos Ana Paula Rosso Dalla Rosa, pelo incentivo, apoio e confiança.

Aos colegas da biblioteca da Unisul, em especial a equipe da Grande Florianópolis, por compreender os momentos de ausência. Sem o apoio dessa equipe não seria possível minha dedicação ao mestrado. Obrigada pela parceria!

Às bibliotecárias da Unisul: em especial à Ana Cláudia, Carolini, Salete, Sibelly, Tatyane e Lilieudi, pela amizade, incentivo e por toda ajuda necessária nesse processo de aprendizado.

À bibliotecária Deisi Martignago, que partilhou comigo momentos de dificuldade, dando apoio e incentivo.

À Prof.<sup>a</sup> Maria Inês Castineira, pelo incentivo e eterna torcida para que eu realizasse esse sonho.

Ao Renan Cascaes e Adriano Pires pelo apoio na parte tecnológica e conversão de textos.

Às amigas: Fabiana, Janice, Maérli e Lisandra, por compreender os momentos ausentes. Em especial a Lisandra Medeiros que não me deixava desanimar nos momentos de grande angústia.

Enfim, agradeço a todos que de alguma maneira contribuíram para realização desse trabalho e que estiveram presentes na minha caminhada. Muito obrigada!

"A Era da Informação oferece muito à humanidade, e eu gostaria de pensar que nós nos elevaremos aos desafios que ela apresenta. Mas é vital lembrar que a informação - no sentido de dados brutos - não é conhecimento, que conhecimento não é sabedoria, e que sabedoria não é presciência. Mas a informação é o primeiro passo essencial para tudo isso."

(Arthur C. Clarke)



## RESUMO

SILVA, Luciana Mara. **Repositório institucional como sistema técnico-social:** composição, ambiente e estrutura. 2014. 149 f. Dissertação (Mestrado em Ciência da Informação). Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de Santa Catarina, Florianópolis, 2014.

Repositórios institucionais (RIs) são sistemas de informação técnico-sociais que emergem como estratégia do movimento de acesso livre para comunicação científica. O potencial dos RIs ainda é tema de intensa discussão na literatura da área, o que leva à pergunta desta pesquisa: como entender um RI de forma abrangente, não fragmentada? O objetivo foi o de construir uma descrição sistêmica de um RI, caracterizando seus elementos: componentes, itens do entorno e ligações estruturantes. Os procedimentos metodológicos tratam da abstração desses elementos a partir de revisão da literatura e entrevistas com peritos, orientadas pelo modelo de sistema composição-ambiente-estrutura-mecanismo (CESM) de Bunge. É uma pesquisa descritiva e exploratória com abordagem ontológica a partir da filosofia sistêmica de Mario Bunge. Como resultado, apresenta-se a síntese de um modelo de RI genérico, consolidado a partir de modelos parciais relativos aos textos fontes analisados e refinado a partir das entrevistas. O modelo CESM resultante postula-se como uma representação abrangente de RI, que pode favorecer novas pesquisas sobre mecanismos dos RIs – os processos que promovem ou obstruem seu funcionamento.

**Palavras-chave:** Repositórios institucionais. Comunicação científica. Acesso aberto. Sistemismo. Sistemas de informação.



## ABSTRACT

SILVA, Luciana Mara. Institutional repository as a technical-social system: composition, structure, and environment, 2014. 149 f. Dissertation (Masters in Information Science). Graduate Program in Information Science, Federal University of Santa Catarina, Florianópolis, 2014.

Institutional repositories (IR) are technical-social information systems that emerge as strategy of the open access movement to scientific communication. The potential of IR is under intense debate in the literature, which leads to our research question: how to understand an IR in a comprehensive, non-fragmented way? The objective is elaborate a systemic description of IR, characterizing its elements: components, environmental items, and structuring bonds. The methodological procedures deal with the abstraction of these elements from a literature review and interviews with experts, guided by Mario Bunge's composition-environment-structure-mechanism (CESM) system model. It is a descriptive and exploratory research with an ontologic approach from Bunge's systemic philosophy. The result is a generic IR system model, synthesized from partial models relative to the texts analyzed, and refined considering the interviews. The resulting CESM model, we claim, is a comprehensive representation of IR that should enlighten new research endeavors on IR mechanisms – the processes that promote or obstruct the functioning of IR.

**Keywords:** Institutional repositories. Open access. Systemism. Information systems.



## LISTA DE FIGURAS

Figura 1 – Representação gráfica da modelagem a partir do texto fonte (1)	62
Figura 2 – Representação gráfica da modelagem a partir do texto fonte (2)	64
Figura 3 – Representação gráfica da modelagem a partir do texto fonte (3)	65
Figura 4 – Representação gráfica da modelagem consolidada do sistema "repositório institucional"	69



## LISTA DE QUADROS

Quadro 1 – Canais formais e informais da comunicação científica.....	26
Quadro 2 – Principais marcos do movimento do acesso livre à informação .....	32
Quadro 3 – Três visões de mundo e suas características .....	46
Quadro 4 – Exemplos de modelos CESM .....	47
Quadro 5 – Grupo de pesquisadores .....	51
Quadro 6 – Fontes de dados para coleta .....	52
Quadro 7 – Legenda de cores para destaque nos textos fontes.....	53
Quadro 8 – Extrato do texto fonte 1, marcado com elementos de um repositório institucional: componentes (vermelho), itens do ambiente (verde) e ligações (azul claro).....	54
Quadro 9 – Análise comparativa dos pesquisadores (exemplo: componentes).....	55
Quadro 10 – Contexto do termo no texto fonte .....	55
Quadro 11 – Exemplo da síntese terminológica de um componente.....	57
Quadro 12 – Exemplo das três marcações .....	58
Quadro 13 – Elementos [e seus sinônimos e correlatos] do modelo de sistema “repositório institucional” do texto fonte (1) .....	61
Quadro 14 – Elementos [e seus sinônimos e correlatos] do modelo de sistema “repositório institucional” do texto fonte (2) .....	63
Quadro 15 – Elementos [e seus sinônimos e correlatos] do modelo de sistema “repositório institucional” do texto fonte (3) .....	64
Quadro 16 – Elementos [e seus sinônimos e correlatos] do modelo de sistema “repositório institucional” consolidado .....	67
Quadro 17 – Detalhamento das ligações modeladas no sistema "repositório institucional" .....	70



## **LISTA DE ABREVIATURAS E SIGLAS**

BOAI – Budapest Open Access Initiative

BRAPCI – Base de Dados de Artigos de Periódicos em Ciência da Informação

CESM – *Composition-Environment-Structure-Mechanism*

ENANCIB – Encontro Nacional de Pesquisa em Ciência da Informação

OAI-PMH – Open Archives Initiative - Protocol for Metadata Harvesting

OCLC – Online Computer Library Center

RI – Repositório Institucional

RIIs – Repositórios Institucionais

ROAR – Registry of Open Access Repositories

OPENDOAR – Directory of Open Access Repositories

SCOT – Social Construction of Technology

TICs – Tecnologias de Informação e Comunicação



## SUMÁRIO

<b>1 INTRODUÇÃO .....</b>	<b>23</b>
<b>2 OS CAMINHOS DA COMUNICAÇÃO CIENTÍFICA .....</b>	<b>25</b>
2.1 O IMPACTO DAS TECNOLOGIAS DE INFORMAÇÃO E COMUNICAÇÃO .....	28
2.2 INICIATIVAS DE ARQUIVOS ABERTOS E MOVIMENTO DE ACESSO LIVRE .....	29
<b>3 REPOSITÓRIOS INSTITUCIONAIS .....</b>	<b>35</b>
3.1 FUNÇÃO E REPRESENTAÇÃO DOS REPOSITÓRIOS INSTITUCIONAIS NO SISTEMA DE COMUNICAÇÃO CIENTÍFICA	36
3.2 BARREIRAS E DESAFIOS NO CAMINHO DOS REPOSITÓRIOS INSTITUCIONAIS .....	38
3.3 A IMPORTÂNCIA DOS MOTORES DE BUSCA PARA VISIBILIDADE DOS REPOSITÓRIOS INSTITUCIONAIS .....	42
<b>4 A ABORDAGEM SISTÊMICA DE MARIO BUNGE .....</b>	<b>45</b>
4.1 SISTEMISMO COMO VISÃO DE MUNDO.....	45
4.2 POSTULADOS DO SISTEMISMO .....	46
4.3 O MODELO DE SISTEMA CESM .....	47
4.4 AS REGRAS GERAIS DA PESQUISA SISTEMISTA .....	48
<b>5 PROCEDIMENTOS METODOLÓGICOS.....</b>	<b>49</b>
5.1 CARACTERIZAÇÃO DA PESQUISA .....	49
5.2 PROCEDIMENTOS DE COLETA E ANÁLISE DE DADOS .....	50
<b>6 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS.....</b>	<b>61</b>
<b>7 CONSIDERAÇÕES FINAIS.....</b>	<b>77</b>
<b>REFERÊNCIAS.....</b>	<b>79</b>
<b>APÊNDICE .....</b>	<b>87</b>
<b>APÊNDICE A - Poster Enacib .....</b>	<b>89</b>
<b>ANEXOS .....</b>	<b>91</b>
<b>ANEXO A – Marcação e coleta do Texto 1 .....</b>	<b>93</b>
<b>ANEXO B – Marcação e coleta do texto 2.....</b>	<b>113</b>
<b>ANEXO C – Marcação e coleta do texto 3.....</b>	<b>135</b>



## 1 INTRODUÇÃO

Os Repositórios Institucionais (RIs) surgiram como uma das estratégias do Movimento do Acesso Livre para disseminar a produção científica publicada em revistas indexadas, por meio da inserção, guarda e disseminação da produção intelectual de uma instituição de pesquisa. “Emerge, assim, como uma peça fundamental no quebra-cabeça da iniquidade no acesso a informação científica” (GUIMARÃES; SILVA; NORONHA, 2009, p. 262).

Os RIs podem ser vistos como sistemas sociais e técnicos, nos quais há intervenção humana e tecnológica para pleno funcionamento do sistema. Poucos estudos evidenciam a importância dos elementos do relacionamento técnico-social. As iniciativas para viabilização e êxito dos repositórios ultrapassam questões tecnológicas. Existe a necessidade de um forte trabalho com a colaboração das pessoas para o povoamento sustentável de suas coleções por meio do autoarquivamento (GUEDON, 2009).

A concepção e implantação dos RIs no Brasil, com vistas ao compartilhamento do conhecimento em acesso livre, tende a acelerar em termos globais a visibilidade da pesquisa brasileira por meio da fluidez que o processo de comunicação científica assume na busca e recuperação da informação. Baseados na iniciativa dos arquivos abertos, “potencializam não só a organização e a armazenagem dos conteúdos, mas também sua recuperação de maneira mais rápida e mais ampla do que o modelo atual do processo de comunicação científica” (LEITE; COSTA, 2007, p. 217). Esta discussão será evidenciada na seção 2, que aborda os caminhos da comunicação científica, o impacto das tecnologias de informação e comunicação e as iniciativas de arquivos abertos e movimento de acesso livre.

Para elucidar o potencial do RI, a pesquisa abordará, na seção 3, a **função e representação dos RIs no sistema de comunicação científica**; com vistas ao promissor impacto do RI na comunicação científica; **as barreiras e desafios no caminho dos RIs**: dentre as barreiras destaca-se o deficiente esclarecimento da comunidade científica acerca de suas vantagens, objetivos e importância no sistema de comunicação científica; e a **importância dos motores de busca para visibilidade dos RIs**: que destaca a invisibilidade dos RIs para motores de busca a partir da adoção de padrões de interoperabilidade de metadados.

Diante do contexto apresentado, questionamentos, divergências e/ou lacunas, acima citados, emerge a necessidade de entender o RI como “um todo”, de maneira ampla: como entender um RI de forma abrangente e não fragmentada?

Esta pesquisa trata de elaborar essa visão a partir do modelo *Composition-Environment-Structure-Mechanism* (CESM) de Bunge (1997, 2003) e tem como objetivo construir uma descrição sistêmica de um RI caracterizando seus elementos: componentes, itens do entorno e ligações estruturantes. Os objetivos específicos são: (1) identificar os termos que descrevem elementos de um sistema de RI a partir de fontes de informação selecionadas; (2) sintetizar os elementos abstraídos de cada fonte em uma modelagem parcial de RI; (3) sintetizar os modelos parciais em um modelo genérico de sistema de RI.

A abordagem teórico-metodológica é baseada no sistemismo de Bunge (1997, 2003), apresentado na seção 4. A pesquisa sistemista parte da descrição de sistemas identificando seus componentes, itens do ambiente e ligações estruturantes. Trata-se de uma pesquisa descritiva e exploratória. Os procedimentos metodológicos da pesquisa, expostos na seção 5, incluem as seguintes etapas: (1) seleção de fontes abrangentes; (2) marcação dos componentes, itens do ambiente e ligações reconhecidos em cada fonte pelo grupo de pesquisadores; (3) discussão em grupo dos elementos marcados pelos diferentes pesquisadores; (4) síntese e consolidação de um modelo de RI por cada texto fonte; (5) discussão da síntese com peritos na área; (6) síntese de um único modelo abrangendo a análise, discussão e consolidação de todas as fontes analisadas. Essa descrição é necessária para as próximas etapas de uma pesquisa sistêmica sobre RI: a conjectura e teste dos mecanismos que operam a emergência, crescimento, manutenção e morte de repositórios institucionais. A seção 6 apresenta os resultados da pesquisa e sua análise. As considerações finais são dadas na seção 7, complementadas por referências e anexos que incluem os textos fontes com a marcação realizada.

## 2 OS CAMINHOS DA COMUNICAÇÃO CIENTÍFICA

Para o avanço da ciência, as pesquisas, as informações, ideias e descobertas precisam ser registradas, estabelecendo meios que garantam o acesso e o reconhecimento do conhecimento produzido. “O conhecimento científico é produto das atividades relacionadas com a investigação científica, e esta última tem no processo de comunicação um elemento inerente a sua própria natureza” (LEITE, 2006, p. 67).

A comunicação científica busca a divulgação dos resultados de pesquisas, permitindo a transferência dos conhecimentos gerados a partir da investigação científica (CRESPO; RODRIGUES, 2011). Para Targino (2000, p. 54) “é a comunicação científica que favorece ao produto (produção científica) e aos produtores (pesquisadores) a necessária visibilidade e possível credibilidade no meio social em que produto e produtores se inserem”.

Na visão de Weitzel (2006b, p. 52):

a produção científica pode ser entendida como um recurso imprescindível para promover o desenvolvimento da ciência. Sua origem remonta à constituição e consolidação de todo um sistema de comunicação científica no mundo e sua história está vinculada à história da própria ciência, de forma que sua estrutura foi acompanhada pela especialização dos saberes e pela autonomização do campo científico ao longo dos últimos quatro séculos.

Para Le Coadic (2004, p. 31), “as comunidades científicas são, sobretudo, redes de organizações e relações sociais formais e informais que desempenham várias funções. Uma das funções é a de comunicação”. Por meio dos canais de comunicação formal (escrito) e informal (oral), é possível a divulgação de pesquisas e o intercâmbio de informações entre pesquisadores e/ou profissionais da área (LE COADIC, 2004).

No entender de Leite (2006, p. 67),

[...] comunicar o conhecimento científico significa compartilhar o conhecimento produzido por meio da investigação científica. Pressupõe um fluxo de conhecimento (tácito e explícito) entre uma fonte

e um receptor por meio de um determinado canal. Nesse sentido, os canais de comunicação são os mecanismos pelo qual o conhecimento produzido é veiculado formal ou informalmente.

Para melhor elucidar os canais de comunicação formal e informal, representa-se no Quadro 1 suas principais características e diferenças.

Quadro 1 - Canais formais e informais da comunicação científica

<b>Canais Formais</b>	<b>Canais Informais</b>
Público potencialmente grande	Privada, público restrito
Informação permanente armazenada e recuperável	Informação não armazenada e não recuperável
Informação relativamente antiga	Informação recente e atualizada
O fluxo da informação é orientado pelo usuário	O fluxo da informação é orientado pelo seu produtor
Disseminação uniforme	Direção do fluxo escolhida pelo produtor
Redundância moderada	Redundância às vezes muito importante
Pouco <i>feedback</i> para o autor, a interação é indireta	<i>Feedback</i> significativo para o autor, a interação é direta

Fonte: Adaptado de Meadows (1974, p. 93).

No entendimento de Meadows (1999), o processo de acumulação do conhecimento envolve trocas de informação, sendo necessária a divulgação desse conhecimento de forma ‘durável e acessível’. Sob esse olhar, surge a necessidade de se criar uma estrutura de comunicação, ou seja, um fluxo de comunicação que possibilite mais eficiência na transferência da informação em ampla escala. O periódico científico surge por volta do século XVII como marco nesse modelo de comunicação científica<sup>1</sup>, “representa um dos veículos que compõem o produto final, formal, consolidado da disseminação de resultados de pesquisas realizadas por estudiosos de todo o mundo” (BAPTISTA et al., 2007, p. 3).

---

<sup>1</sup> Duas publicações foram marcos desse período: Le Journal des Savants, que refletia as atividades nas áreas de ciências e humanidades e, o Philosophical Transactions of the Royal Society of London, que publicava as comunicações originadas em sessões públicas, equivalente hoje aos anais de congresso.

O periódico científico desempenha papel fundamental no processo da comunicação científica por se constituir na principal via de veiculação de novos conhecimentos e possibilitar a sua difusão para determinado público. Com efeito, sob a influência dos denominados colégios invisíveis, a disseminação de resultados de pesquisas e as discussões de temas científicos assumem, por intermédio deles, um formato mais dinâmico. Afinal, trata-se de um canal de divulgação que engloba diversas autorias, é publicado em intervalos determinados de tempo e apresenta, de forma condensada, os conhecimentos recém-gerados, dando-lhes visibilidade no meio acadêmico e científico (BARBALHO, 2005, p. 126).

Com o avanço da tecnologia, a caracterização dos canais de comunicação (formal e informal) e seus fluxos informacionais, está se tornando cada vez mais indistinta, ou seja, por meio dos canais eletrônicos de informação, a comunicação passa a ser, ao mesmo tempo, formal e informal. Bauman (2001), na obra “Modernidade Líquida”, traduz o que se vive hoje por meio da metáfora do estado líquido (estado temporal), que muda constantemente a sua forma, e sua mobilidade remete à ideia de leveza, de fluidez. Na comunicação científica, é essa mobilidade que ultrapassa barreiras, impacta o acesso, o compartilhamento e a disseminação do conhecimento. Possibilita a construção de novas formas, espaços colaborativos em rede, interação e participação comunitária.

As perspectivas da comunicação científica em espaços colaborativos em rede estão alinhadas às considerações de Oliveira e Noronha (2005, p. 88) ao mencionar em que a “comunicação científica como processo fundamental no desenvolvimento científico se beneficiou com o surgimento das comunicações em rede, pela agilização da comunicação, compartilhamento de informações e interligação de pesquisadores e instituições”.

Nessa discussão cabe a reflexão acerca das implicações da velocidade do conhecimento, com o impacto das tecnologias de informação e comunicação, nessa nova concepção de rede.

## 2.1 O IMPACTO DAS TECNOLOGIAS DE INFORMAÇÃO E COMUNICAÇÃO

O acesso à informação tem sido cada vez mais instantâneo em razão das Tecnologias de Informação e Comunicação (TICs). “Assim, o processo de comunicação formal, e, de modo mais intenso o periódico científico, passa por muitas transformações tecnológicas” (MORENO; LEITE; MÁRDERO ARELLANO, 2006, p. 92). “A internet oferece todas as possibilidades com as quais nem mesmo os mais entusiastas escritores de ficção ou os utópicos sonhadores de Alexandria poderiam sonhar” (MOREIRA, 2005, p. 60). Ribeiro e Vidotti (2009, p. 106) argumentam que, por meio da facilidade para criação e publicação em meio digital, a “*World Wide Web* transformou usuários comuns e instituições comerciais e educacionais, entre outros, em produtores e ‘alimentadores’ dos ambientes informacionais digitais”. A internet possibilita que a informação chegue ao receptor em tempo real, que continua a processar em tempo real, é um meio de comunicação contínuo que possibilita a interação entre as pessoas, no qual um ou mais usuários podem ter acesso, fazendo perguntas, debatendo, circulando em rede novas informações (CASTELLS, 2003).

Para Oliveira e Noronha (2005), a comunicação científica foi afetada por essas transformações: há maior fluidez no desenvolvimento das pesquisas e otimização na troca de informações, colocando em contato pesquisadores localizados em diferentes partes do mundo.

O surgimento e desenvolvimento das novas tecnologias de comunicação e informação, em especial a Internet, têm modificado o processo de comunicação, tanto a informal quanto a formal, estabelecendo uma nova categoria na comunicação científica: a **comunicação eletrônica** (OLIVEIRA; NORONHA, 2005, p. 78, grifo nosso).

Para Alves (2011) as mudanças de suporte encontram-se cada vez mais dinâmicas e interativas, O autor destaca como exemplos de suportes e ambientes de comunicação científica: os espaços colaborativos acadêmicos, os blogs científicos, os colégios invisíveis eletrônicos, as redes sociais. Tais suportes configuram-se como um novo campo de debates em tempo real, devido às facilidades da comunicação eletrônica. Processos de produção, uso e disseminação da informação,

inerentes à comunicação científica, estão hoje interligados, formando uma grande rede, que Musso (2004, p. 31) define como “uma estrutura de interconexão instável, composta de elementos em interação, e cuja variabilidade obedece a alguma regra de funcionamento”.

Com base na Online Computer Library Center (OCLC), Weitzel (2006a) procurou descrever o processo de comunicação científica eletrônica, representando como elementos do fluxo dos produtos científicos: os repositórios digitais (institucional, especializados ou geral); os serviços agregadores (serviços comerciais ou ofertados pela comunidade ou financiado por agências de fomento); os cientistas, docentes e discentes em processo de produção, ensino e pesquisa.

Weitzel (2006b) afirma que, com a aplicação das TIC e a consolidação de algumas iniciativas, principalmente as Iniciativas de Arquivos Abertos e o Movimento de Acesso Livre, este cenário está passando por transformações e está em pleno estágio de reorganização dos processos e produtos da comunicação científica. Ao refletir sobre os movimentos de acesso livre, Baptista e outros (2007, p. 2) afirmam que “os benefícios de tal movimento são, entre outros, a maior visibilidade das pesquisas e sua utilização pelo maior número possível de interessados, o que promove, em última instância, o desenvolvimento da ciência”.

Para elucidar mais a discussão acerca desse “novo padrão” da comunicação científica para o desenvolvimento da ciência, cabe a explanação da seção 2.2, que discorre sobre as iniciativas de arquivos abertos e o movimento de acesso livre.

## 2.2 INICIATIVAS DE ARQUIVOS ABERTOS E MOVIMENTO DE ACESSO LIVRE

A produção e comunicação científica estão ligadas à disseminação dos resultados de pesquisa e trocas de informações entre os pares e, objetiva além da publicação, a troca de ideias e registro do conhecimento (ALVES, 2011). A informação deve “permanecer em movimento”, para ir ao encontro à necessidade do usuário/pesquisador. Meadows (1999, p. 211) reforça que “o requisito básico é que a informação seja fornecida no momento e na forma que torne mais provável sua absorção pelo pesquisador”. Diante do grande universo de conhecimento produzido e não disponível a todos, a comunidade científica criou um movimento que luta pelo acesso livre à informação.

Leite (2006, p. 84, grifo nosso) infere que o

**movimento mundial de acesso livre à informação científica**, fundamentado no modelo de arquivos abertos, tem provocado mudanças paradigmáticas no processo de comunicação científica, e, desde então **configura-se como a principal inovação no contexto da comunicação entre os cientistas**.

Para avançar nessa contextualização é necessário citar dois marcos de crucial importância, nesse novo cenário de comunicação científica: o manifesto da Budapest Open Access Initiative (BOAI)<sup>2</sup>, em 1999, baseado no uso do protocolo Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH)<sup>3</sup> e a reunião de Bethesda, em 2003. O objetivo da BOAI foi “acelerar o progresso do esforço internacional de tornar artigos de pesquisa em todos os campos acadêmicos livremente disponíveis na Internet” (COSTA, 2006, p. 41). Por **acesso aberto ou livre acesso** a BOAI define:

a disponibilidade livre na Internet pública, permitindo a qualquer utilizador ler, baixar, copiar, distribuir, imprimir, pesquisar ou acessar o link para os textos completos destes artigos, rastreá-los para indexação, utilizá-los como dados para *software*, ou utilizá-los para qualquer outro propósito legal, sem barreiras financeiras, legais ou técnicas além das próprias do acesso a Internet. A única restrição sobre a reprodução e distribuição, e a única função dos direitos de autor neste domínio, deve ser o controle do autor sobre a integridade de seu trabalho e o direito de ser devidamente reconhecidos e citados (BUDAPEST OPEN ACCESS INITIATIVE, 2001, tradução nossa).

A partir da definição de acesso aberto ou acesso livre, diferentes conceitos de termos correlatos, sinônimos são encontrados na literatura.

---

<sup>2</sup> A BOAI foi lançada em 2001, defende a disponibilização em acesso livre das pesquisas científicas, sem custos sem barreiras, baseada no uso do protocolo OAI-PMH. <http://www.soros.org/openaccess>.

<sup>3</sup> O protocolo OAI foi desenvolvido pela *Open Archives Initiative*, define um mecanismo para coleta de registros de metadados em repositórios.

Entre as diferentes abordagens, partilha-se da definição de Suber (2003 apud BAPTISTA et al, 2007, p. 5) quando menciona que:

[...] acesso livre diz respeito à acessibilidade ampla e irrestrita a conteúdos disponíveis em formato digital, no sentido em que remove barreiras de preço e de permissão, tornando a literatura científica disponível com o mínimo de restrições de uso.

A BOAI define duas estratégias de acesso livre pautadas e baseadas no uso do protocolo OAI-PMH: a via dourada e a via verde. A via dourada “compreende os periódicos científicos eletrônicos, cujo acesso aberto aos seus conteúdos é garantido pelos próprios editores” (COSTA, 2006, p. 41). Harnad (2007) alerta que a via dourada ainda é muito lenta, apontando entraves para autores instituições que não puderem arcar com custos das publicações. A linha verde compreende o autoarquivamento e/ou autodepósito em repositórios “por parte dos autores, de artigos científicos já publicados ou aceitos para publicação em um periódico referendado” (COSTA, 2006, p. 41)”. Harnad (2007) defende a ideia de autoarquivamento afirmando que “é parte das tarefas de todos os pesquisadores em todo o mundo [...] no entanto, apenas 15% dos pesquisadores autoarquivam espontaneamente”. Leite e Costa (2007) apontam que o aumento da visibilidade da produção científica e o autoarquivamento por parte dos próprios pesquisadores infere no modo como os indivíduos de uma comunidade científica criam, compartilham e utilizam o conhecimento.

A reunião de Bethesda, em 2003, apresenta definições do que são publicações de acesso aberto, atribuindo duas principais condições: a) autores que asseguram a todo usuário direito livre aos seus trabalhos; b) versão completa do trabalho publicado em um repositório *online*. Neste contexto, é relevante resgatar a proposição de arquivos abertos por meio do uso do protocolo OAI-PMH, que significa interoperável, a interoperabilidade dos conteúdos armazenados em arquivos abertos. “A interoperabilidade é um fator primordial, pois possibilita a busca em um só local e a reunião de conteúdos dos diferentes tipos de publicações online produzidas por diferentes *softwares* sem conflitos” (WEITZEL, 2006b, p. 61).

Outras manifestações e manifestos importantes ocorreram nacionalmente e internacionalmente a favor do acesso livre à informação científica, conforme exposto no Quadro 2.

Quadro 2 - Principais marcos do movimento do acesso livre à informação  
(continua)

<b>Período</b>	<b>Movimento</b>
2001	Carta aberta da Public Library of Science (PLoS)
14/02/2002	Iniciativa de Budapeste para o Acesso Aberto
30/10/2002	Carta ECHO
27/08/2003	Association of Learned and Professional Society Publishers (ALPSP)
22/10/2003	Declaração de Berlim sobre o Livre Acesso ao Conhecimento
11/2003	Declaração de Princípios do WellcomeTrust em apoio à edição em livre acesso
4/12/2003	Posicionamento do Inter Academy Panel sobre o acesso à informação científica
5/12/2003	Declaração do International Federation of Libraries Association (IFLA) sobre o livre acesso à literatura científica e aos documentos da pesquisa
12/12/2003	Declaração de Princípios da Cúpula Mundial sobre a Sociedade da Informação (SMSI)
15/01/2004	Declaração de Valparaíso
30/01/2004	Declaração da Organization for Economic Cooperation and Development (OECD) sobre o acesso aos dados da pesquisa financiada por fundos públicos
16/03/2004	Princípios de Washington D. C. para o Livre Acesso à Ciência
30/07/2004	Publicação do relatório do comitê do Parlamento Britânico sobre edição científica
13/09/2005	Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica
26/09/2005	“Declaração de Salvador sobre Acesso Aberto: A Perspectiva dos Países em Desenvolvimento” Declaração de Salvador – Compromisso com a Equidade
12/2005	Carta de São Paulo
05/2006	Declaração de Florianópolis

2007	Projeto Lei PL 1120/2007 - Dispõe sobre o processo de disseminação da produção técnico-científica pelas instituições de ensino superior no Brasil e dá outras providências
02/2013	Proposta de política mandatória de acesso livre a ser aplicada no âmbito da avaliação nacional do sistema científico inglês (Research Excellence Framework – REF)

Fonte: Kuramoto, 2006, p. 97. Com modificações.

Seguindo características das iniciativas acima mencionadas, o movimento de acesso livre propõe maximizar o acesso aos resultados de pesquisa, “maximiza e acelera o impacto das pesquisas e, conseqüentemente, sua produtividade, progresso e recompensas” (BRODY; HARNAD, 2004 apud LEITE; COSTA, 2007, p. 212). Segundo Harnad (2007), quanto mais livre for o acesso aos conteúdos, maior a visibilidade da produção dos pesquisadores. Na perspectiva do movimento de acesso aberto. Björk (2005 apud MULLER 2006, p. 32) classifica os canais de comunicação científica existentes:

- periódicos científicos eletrônicos com avaliação prévia pelos pares;
- servidores de *e-prints* para áreas específicas – repositórios para assuntos específicos;
- Repositórios Institucionais de universidades específicas;
- autoarquivamento em páginas pessoais dos autores.

Considerando os quatro tipos de canais para o acesso aberto descritos acima, Mueller (2006, p. 33) afirma que “a maior novidade e talvez o maior potencial para desenvolvimento estariam nos Repositórios Institucionais, que de certa forma disciplinam e sistematizam a atividade do arquivamento individual”. Leite e Costa (2007) consideram os Repositórios Institucionais um caminho que emerge como uma poderosa iniciativa para o armazenamento, busca e disseminação da comunicação científica em acesso livre. A próxima seção discorre sobre a temática RI, objeto de estudo dessa pesquisa.



### 3 REPOSITÓRIOS INSTITUCIONAIS

Historicamente, em 1991 surgiu o ArXiv, primeiro repositório digital de pré-prints, por iniciativa do físico Paul Ginsparg, num laboratório de física em Los Alamos, Novo México, EUA. A partir desta iniciativa, “pesquisadores passaram a criar arquivos eletrônicos de pré-prints e postprints como alternativa para publicação direta de seus trabalhos em texto completo” (MARCONDES et al., 2005, p. 43). Daí em diante, o crescimento dos repositórios tem sido rápido, significativo e propulsor para o avanço da ciência (ROSA; GOMES, 2010a).

Como fonte de pesquisa que evidencia a expansão dos repositórios em nível global menciona-se o Open Doar<sup>4</sup> (Directory of Open Acces Repositories) e o Registro de Repositórios de Acesso Livre (ROAR)<sup>5</sup>. O Roar é um diretório mantido pela Universidade de Southampton, no Reino Unido, que objetiva reunir informações sobre os repositórios de acesso aberto e sua evolução a nível nacional e internacional. Em pesquisa realizada em dezembro de 2013 constatou-se o registro de 3.565 repositórios nomeados repositórios institucionais. O Open Doar é um diretório mantido pelo Sherpa Services que reúne a coleção, em nível mundial, dos repositórios, categorizando informações para facilitar a análise dos mesmos. Em pesquisa realizada em dezembro de 2013 constatou-se o registro de 2.545 repositórios nomeados repositórios institucionais. É importante mencionar que ambos os diretórios são alimentados com cadastro voluntário, logo, nem todos os repositórios nomeados como repositórios institucionais assim se caracterizam, pois os números incluem repositórios temáticos, bibliotecas digitais, de teses e dissertações, entre outros. Para especificação é necessário uma análise documental acerca dos dados cadastrados, sendo que essa especificidade não é objeto de estudo dessa pesquisa.

Os Repositórios Institucionais caracterizam-se por estar em um ambiente eletrônico em que está a produção científica da instituição/organização, armazenada em formato digital, permitindo a busca e a recuperação de informação para seu posterior uso tanto no âmbito nacional quanto internacional. Para Rosa e Gomes (2010a), um Repositório Institucional (RI) é um arquivo digital que contém a produção intelectual criada pelos acadêmicos, pesquisadores e

---

<sup>4</sup> <http://www.openoar.org/>

<sup>5</sup> <http://roar.eprints.org/>

professores de uma instituição, sendo acessível a utilizadores finais (internos e/ou externos) à instituição, com poucas ou nenhuma barreira de acesso. Fato é que a literatura aponta inúmeros conceitos de RI, nesse contexto,

quer seja como resposta a escalada nos custos das assinaturas de periódicos científicos; como novo “serviço de informação”; como um elemento adicional no sistema de comunicação científica; como estratégia para conferir visibilidade institucional, **os RIs podem ser a resposta!** (GUIMARÃES, SILVA, NORONHA, 2009, p. 263, grifo nosso).

Nesta pesquisa adotou-se o conceito de Guimarães, Silva e Noronha (2009, p. 263) que corrobora a visão do RI como um sistema técnico-social:

[...] muito antes de ser mais um dispositivo ou infraestrutura tecnológica a favor da ampliação da circulação de informação científica, um RI é um complexo sócio técnico, onde estão envolvidos um conjunto de atores com visões, práticas e perspectivas diversas, cujas relações são tecidas a luz de micro e macropolíticas, locais e globais.

Para por em prática todo potencial dos RIs ainda há muito a ser investigado e explorado, assim, destaca-se nas próximas seções a representação dos RIs no sistema de comunicação científica, as barreiras e desafios no caminho dos RIs e a importância dos motores de busca para visibilidade dos RIs.

### 3.1 FUNÇÃO E REPRESENTAÇÃO DOS REPOSITÓRIOS INSTITUCIONAIS NO SISTEMA DE COMUNICAÇÃO CIENTÍFICA

A via verde (*green road*) é um dos caminhos que conduz o acesso aberto à informação científica. Engloba os repositórios temáticos (produção científica de uma área) ou institucionais (produção científica de uma instituição). A redação desta seção aborda uma breve discussão acerca da função e representação dos RIs, no contexto da comunicação científica, considerando-os como perspectiva na expansão da

disseminação da produção intelectual em acesso aberto das instituições de pesquisas.

Para Leite e Costa (2007), os RIs visam melhorar o processo de comunicação científica a partir de mecanismos que aumentam tanto a eficácia da preservação da produção intelectual (pesquisadores e instituições), quanto a visibilidade de ambos.

O papel que Repositórios Institucionais representam, de fato, em **duas questões fundamentais. Primeiro**, na melhoria do ensino, do aprendizado e da pesquisa. Em outras palavras, na melhoria do saber e da comunicação científica. **Segundo**, no potencial que encerram como instrumentos de gestão do conhecimento produzido, disseminado e utilizado nas e pelas universidades (LEITE; COSTA 2007, p. 213 grifo nosso).

Guimarães, Silva e Noronha (2009, p. 268) apontam duas principais razões que concebem os RIs:

[...] por um lado, para tencionar (se não competir com) o modelo tradicional de publicações na ciência, e reduzir o poder de monopólio dos periódicos [vinculado ao movimento acesso livre]; por outro lado para complementar (se não refundar) o sistema de comunicação acadêmica [gestão de conteúdos/serviço de informação de uma instituição de pesquisa].

Inferre-se ser de fundamental importância ter bem claro os conceitos de RI, seu propósito em detrimento a expectativa do usuário para sua criação, povoamento e promoção do seu uso. Rosa e Gomes (2010b) alertam que este processo, para ser bem sucedido, passa pela definição de políticas institucionais para funcionamento do RI e por opções organizacionais de diferente natureza. As políticas de funcionamento de um RI

[...] são fundamentais para seu o estabelecimento como serviço de informação e reconhecimento por parte da comunidade [...] e deve contribuir para integrar o repositório na estratégia e no ambiente

de informação da instituição; apresentar uma visão clara dos principais atores envolvidos no contexto do repositório; satisfazer as necessidades da comunidade; atrair usuários; estabelecer responsabilidades, prerrogativas, direitos e deveres; povoar o repositório; torná-lo juridicamente viável; manter relações externas com as agências de fomento, editores e sociedades científicas; manter relações internas com administradores acadêmicos, pesquisadores e equipes de gestão da informação na instituição; preservação digital de longo prazo; gerenciar riscos; facilitar o trabalho da equipe gestora do repositório (LEITE, 2009, p. 71).

Nesta seção procurou-se evidenciar que independentemente das concepções de criação e uso dos RIs, sua representação na comunicação científica é fato, evidenciada por diferentes autores aqui citados. Contudo, o estado atual, o papel e a estratégia dos repositórios institucionais são temas de intensa discussão acadêmica e perpassam por questões relacionadas às discussões das subseções que seguem.

### 3.2 BARREIRAS E DESAFIOS NO CAMINHO DOS REPOSITÓRIOS INSTITUCIONAIS

Esta seção cita barreiras e desafios que os RIs precisam vencer para êxito do povoamento de suas coleções como: o deficiente esclarecimento da comunidade científica acerca de suas vantagens, objetivos e importância no sistema de comunicação científica; divergências sobre como é e como deve funcionar o RI e a complexidade das interações entre os atores envolvidos:

#### **a) o deficiente esclarecimento da comunidade científica acerca de suas vantagens, objetivos e importância no sistema de comunicação científica**

Estudos como o de Marques e Maio (2010) relatam que para o fortalecimento dos RIs e sua contribuição no progresso da sociedade e da ciência, estudos devem ser aprofundados para a sua compreensão como estratégia de promoção das pesquisas, visibilidade dos autores e das Instituições. A temática RI está em desenvolvimento e emerge num cenário mutante do fluxo da comunicação científica. “A comunidade

científica ainda não está suficientemente esclarecida de suas vantagens, sua função e da importância do seu engajamento para que essas iniciativas ocupem seu espaço” (WEITZEL, 2006b, p. 66). Os repositórios institucionais

valorizam não só os estoques de conhecimento registrado, mas também agregam valor ao conhecimento não registrado na medida em que a interação entre indivíduos, o compartilhamento de experiências e as trocas de conhecimento também são consideradas (LEITE, 2006, p. 210).

Na obra “como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto”, Leite (2009) menciona a necessidade de se fazer um *marketing* e estratégias de povoamento junto aos atores envolvidos no processo de implantação e desenvolvimento dos RIs.

Questões políticas também permeiam a ascendência dos RIs. Cita-se como exemplo o *case* de sucesso dos RIs nos EUA, que se baseiam no movimento que as agências de fomento de ordem pública e privada assumiram a favor do acesso aberto a informação científica (GUEDON, 2009).

Diante das lacunas na literatura como enfoque nesta discussão, emerge a necessidade de criar mecanismos para divulgar as vantagens e objetivos dos RI para comunidade científica. Infere-se que uma visão sistêmica do RI, objeto de estudo dessa pesquisa, venha a contribuir com pesquisas na área.

## **b) divergências sobre como é e como deve funcionar o RI**

Estudos apontam a falta de esclarecimentos quanto ao propósito de um RI, tendo em vista a necessidade da instituição de pesquisa e expectativas dos seus usuários. Essa definição deve nortear o povoamento do RI, estabelecendo os tipos de documentos armazenados. Guimarães, Silva e Noronha (2009) estabeleceram duas principais estratégias na concepção dos RIs: (1) como um serviço de informação (ambiente eletrônico em que está a produção científica da instituição/organização, armazenada em formato digital); (2) e /ou apoio ao movimento de acesso livre (Via Verde).

Na construção do RI aliado ao movimento de acesso livre, Leite e Costa (2007, p. 217) relatam que os RIs dispõem de uma série de características que “potencializam não só a organização e a armazenagem dos conteúdos, mas também sua recuperação de maneira mais rápida e mais ampla do que o modelo atual do processo de comunicação científica”. Para Guedon (2001), entre os maiores desafios desse novo modelo está a discussão entre o movimento de acesso livre e os editores comerciais e, a descoberta de um mecanismo universal de métrica para atribuir valor aos repositórios digitais, mecanismo que hoje já é reconhecido na via dourada (periódicos científicos). “O impacto da pesquisa é, especialmente nos dias atuais, a moeda mais valiosa no mercado de ações da pesquisa, e o acesso aberto aumenta de modo exponencial esses indicadores” (COSTA, 2008, p. 223).

Na concepção do RI como serviço de informação de uma instituição de pesquisa, o RI estaria “[...] a serviço de uma estratégia de gestão institucional da pesquisa, como um guia para avaliação e, principalmente, para estimular e facilitar o reuso dos conteúdos, acelerando o ciclo da pesquisa” (GUIMARÃES; SILVA; NORONHA, 2009, p. 270). Nessa perspectiva, Lynch (2003) infere que o RI pode ser visto como um conjunto de serviços que uma universidade oferta para sua comunidade, possibilitando o gerenciamento e a disseminação dos conteúdos digitais criados pelos seus membros. É crescente o número de universidades que criam seus RIs, segundo Hess e Ostram (2007), muitos incentivos existem para as Instituições apresentarem suas pesquisas, criarem seus repositórios, de imediato cita-se a alta visibilidade, uso e impacto por meio de citações.

Leite (2006, p. 88) complementa que o RI

[...] tem o potencial de servir como indicadores tangíveis da qualidade de uma universidade e de demonstrar a relevância científica, social e econômica de suas atividades de pesquisa, aumentando a visibilidade, o status e o valor público da instituição.

Diante das diferentes percepções sobre o que é e como deve funcionar um RI, avança-se para discussão da interação entre os atores envolvidos no processo de implantação, gestão, uso e povoamento do RI.

### **c) complexidade das interações entre os atores envolvidos.**

O pleno funcionamento do RI ultrapassa questões tecnológicas. Está também atrelado a uma rede complexa de atores envolvidos e/ou grupos sociais relevantes com atividades específicas no âmbito do RI. Para Rosa e Gomes (2010b, p.195)

o desenvolvimento de RI envolve um conjunto de atores que podemos enquadrar em várias categorias que não se excluem mutuamente: os criadores/gestores do repositório, os depositantes e os usuários finais. Em alguns contextos, dependendo das dinâmicas e opções institucionais, podem ser considerados outros intervenientes com funções específicas.

Guedon (2009) enfatiza que os RIs devem ser examinados à luz da abordagem teórico metodológica da Construção Social das Tecnologias (*Social Construction of Technology – SCOT*) com subsídios de ferramentas para melhor compreendê-lo como um sistema técnico social que emerge numa rede de tecnologias e pessoas. Nessa abordagem, destaca-se os grupos sociais relevantes, ou os diferentes atores (*stakeholders*) envolvidos que

[...] possuem um lugar de destaque no processo, e para ele trazem diferentes recursos (conhecimento, poder, financiamento, dentre outros), diferentes visões e expectativas sob a forma e sentido futuro dos artefatos (GUIMARÃES; SILVA; NORONHA, 2009, p. 266).

Os grupos sociais relevantes possuem recursos e interesses distintos sob os RIs. É preciso identificar quais são esses grupos, como eles se relacionam, os recursos e interesse que possuem para estabelecer estratégias de desenvolvimento e implantação dos RIs (GUIMARÃES; SILVA; NORONHA, 2009, p. 267). Os autores apontam como potenciais grupos sociais relevantes “[...] os pesquisadores, as instituições de pesquisa, os editores científicos (comerciais e universitários), as bibliotecas e os profissionais de informação, as agencias de fomento, os usuários, o Legislativo e outras organizações,

governamentais ou não” (GUIMARÃES; SILVA; NORONHA, 2009, p. 266).

A relação entre diferentes grupos sociais e atores envolvidos é complexa, com interesses e recursos diferentes e estas relações podem encaminhar o RI ao sucesso ou a sua falência. “Nesse sentido, qualquer visão simplificadora do relacionamento entre tecnologias, espaços de práticas constituídos e mudanças sociais é, no mínimo, ingenuidade” (GUIMARÃES; SILVA; NORONHA, 2009, p. 264).

Para Hess e Ostrom (2007), a medida do sucesso dos RIs será como as pessoas se comportam em respostas a essas ações e como essas respostas determinam resultados positivos para comunicação científica. Logo, identificar suas relações é determinante para o pleno funcionamento do RI. Nesse contexto, a proposta de pesquisa em descrever os componentes, itens do ambiente e estrutura de um RI, vem ao encontro com a necessidade de observar as ligações entres os grupos sociais relevantes.

### 3.3 A IMPORTÂNCIA DOS MOTORES DE BUSCA PARA VISIBILIDADE DOS REPOSITÓRIOS INSTITUCIONAIS

Iniciativas mundiais para construção de Repositórios Institucionais que utilizam uma arquitetura baseada no modelo de arquivos abertos, já mencionados na seção 2.2, concebem um “modelo de interoperabilidade de dados”, permitindo assim a integração de registros disponibilizados em repositórios nacionais e internacionais.

A interoperabilidade assume papel fundamental na representação da informação em rede, ou seja, no **acesso à informação**, proporcionando assim, “maior visibilidade e impacto mundial, maior espaço para a competitividade em níveis de igualdade, cooperação internacional facilitada, acesso, disseminação da informação” (LAASER, RODRIGUES, FACHIN, 2009, p. 7). Infere-se que a visibilidade se efetiva quando a pesquisa está disponível e acessível, sendo recuperada, lida ou citada, favorecendo a promoção da informação científica. Nessa perspectiva, presume-se que a maximização da visibilidade dos RIs será, também alcançada, por meio dos provedores de serviço que são “as instituições que fornecerem serviços de informação com valor agregado, por meio da coleta dos metadados de diversas instituições e de sua integração em um grande repositório central” (KURAMOTO, 2010).

Entre os provedores de serviços destacam-se o Google e Scholar Google pela popularidade que possuem junto à comunidade acadêmica,

que os utilizam como fonte de pesquisa na realização de suas produções. O Scholar Google difere do Google, pois limita seus resultados em documentos acadêmico-científicos. Dr. Anurag Acharya, engenheiro fundador do Scholar Google, afirma que o objetivo é oferecer a lista mais abrangente dos trabalhos de pesquisa disponíveis na *Web*, e que o Scholar Google limita seus resultados em documentos revisados, tais como: teses, livros, resumos e técnicos relatórios, entre outros documentos (ASSIS, 2005).

Em pesquisa realizada em 2001, nos EUA, com a amostragem de oito repositórios, foi constatado que quatro desses RIs analisados apresentou um índice abaixo de 50% na indexação de seus documentos no Scholar Google. Os autores relataram em sua pesquisa que a adoção do padrão Dublin Core como padrão de metadados (opção comum em RI) implica em restringir sua visibilidade, já que indexadores como o Scholar Google têm dificuldade na indexação de RIs que usam esse padrão. O esquema Dublin Core compromete a indexação de artigos pelo Scholar Google porque não inclui campos de metadados compatíveis com as especificações que o Scholar Google estabelece, explicam os autores Arlitsch e O'Brien (2012).

Se a pesquisa e sua comunicação são “inseparáveis”, como sustenta Meadows (1999), é válido questionar o sentido de empreender um sistema de informação como um RI que dificulta a recuperação por seus potenciais usuários via *web*. Arlitsch e O'Brien (2012) alertam que os pesquisadores que contribuem para povoamento dos RIs podem perder o interesse se suas publicações não puderem ser localizadas (e citadas) nos motores de busca academicamente orientados, como Scholar Google. O RI tem potencial para aumentar as taxas de citação do autor, e por sua vez afetar o ranking das universidades, mas segundo Arlitsch e O'Brien (2012), esse potencial pode ser seriamente comprometido se o conteúdo do RI é invisível para os pesquisadores que utilizam o Scholar Google.

Nesse sentido, Carvalho (2009) ressalta a importância de estudos que busquem a interoperabilidade entre os vários sistemas de produção e disponibilização de conteúdos digitais, para aproveitamento dos esforços e recursos já alocados na organização da produção científica nacional de acesso aberto.

Na seção 3 buscou-se a fundamentação teórica do objeto de estudo dessa pesquisa que são os RIs. Avança-se para a próxima seção que vai esclarecer ao leitor a abordagem teórico-metodológica desta pesquisa, baseada no sistemismo de Bunge.



## 4 A ABORDAGEM SISTÊMICA DE MARIO BUNGE

Sistemas de informação muito complexos geralmente são vistos de forma fragmentada, com funcionamento pouco compreendido até por seus protagonistas (KERN, 2011). Bunge (2003) define sistema como um objeto estruturado, composto por partes interligadas, com características sistêmicas (emergentes) que seus componentes não têm.

Tudo é sistema: “toda coisa, seja concreta ou abstrata, é um sistema ou um componente ou potencial componente de sistema” (BUNGE, 2003, p. 114). O estudo de sistemas, dessa forma, deve buscar a compreensão do todo e de suas relações com os indivíduos ou partes que o compõem:

A visão de que o conceito de sistema é central para a ciência e a tecnologia social e, com efeito, para todas as ciências e tecnologias, pode ser chamada sistemismo. O sistemismo tem dois componentes: ontológico e metodológico. O sistemismo ontológico afirma que o mundo é um sistema antes que uma coleção de indivíduos ou um bloco sólido. Seu parceiro metodológico é a visão de que os todos (*wholes*) são melhor entendidos por meio da análise (*top-down*) e indivíduos por meio da síntese (*bottom-up*) (BUNGE, 2006, p. 12-13, tradução nossa).

Na sequência, serão abordados conceitos do sistemismo de Bunge, que embasa teoricamente esta pesquisa: o sistemismo como visão de mundo, os postulados do sistemismo, o modelo de sistema composição-ambiente-estrutura-mecanismo (CESM) e as regras da pesquisa sistemista.

### 4.1 SISTEMISMO COMO VISÃO DE MUNDO

O sistemismo é a ontologia realista científica de Mario Bunge, que refuta e conjuga simultaneamente o individualismo e o holismo filosóficos. Refutam-se a negativa individualista a reconhecer a existência de entidades supraindividuais e a negativa holista a analisar sistemas a partir de seus indivíduos e suas interações entre si e com o ambiente. Conjugam-se a síntese holista, caracterizada no

reconhecimento da existência do todo, e a análise individualista, típica das ciências naturais.

O Quadro 3 representa as três ontologias (individualismo, holismo e sistemismo) e suas respectivas descrições:

Quadro 3 - Três visões de mundo e suas características

Ontologia	Epistemologia	Metodologia	Slogan	Autores e idéias (exemplos)
Individualismo	Racionalismo ou empirismo	Análise: micro-redução	<i>Divide et impera.</i>	Cartesianismo; Adam Smith – o interesse individual é que move a economia e a sociedade; Karl Popper – o conhecimento humano resulta de conhecimentos individuais.
Holismo	Intuicionismo	Síntese: macro-redução	<i>Conjuga et impera.</i>	Parmênides (“tudo é um”); Bertalanffy e Laszlo, autores de teorias de sistemas; Luhmann – eliminação do indivíduo e reificação das instituições; gestaltismo.
Sistemismo	Realismo científico	Análise e síntese	<i>Divide et conjuga.</i>	Newton, a quem couber reconhecer que o sistema solar é mais do que um agregado de corpos celestes; Tocqueville, analista histórico com visão sistêmica.

Fonte: Kern (2011) baseado em Bunge (2003, p. 126).

O sistemismo é uma visão de mundo que pode ser utilizada como base nos estudos empíricos de qualquer sistema, incluindo sistemas técnicos e sociais. É o caso dos RIs, foco desta pesquisa: seu funcionamento depende de um sistema técnico que interage com um sistema social.

## 4.2 POSTULADOS DO SISTEMISMO

Segundo Bunge (2000) o sistemismo fundamenta-se em cinco postulados:

- a) toda coisa, seja concreta ou abstrata, é um sistema ou um componente ou potencial componente de sistema;
- b) sistemas têm características sistêmicas (emergentes) que seus componentes não têm;
- c) todos os problemas deveriam ser abordados de forma sistêmica em vez de em forma fragmentada;
- d) todas as ideias deveriam ser unidas em sistemas (preferencialmente teorias);
- e) o teste de qualquer coisa, seja ideia, método ou artefato, supõe a validade de outros itens que são tomados como pontos de referência (*benchmark*) provisoriamente.
- f) Esses postulados sustentam a visão de mundo sistêmica de Bunge. Fornecem os fundamentos ontológicos e também apontam diretrizes metodológicas, mas não são suficientes para modelar

sistemas. A modelagem de sistemas precisa do modelo de sistema CESM, descrito na próxima subseção.

### 4.3 O MODELO DE SISTEMA CESM

Para Bunge (2003), um sistema concreto pode ser descrito segundo o modelo CESM (*Composition, Environment, Structure, Mechanism*). Conforme Bunge (1997):

- a) **Composição** é a coleção de partes ou elementos componentes;
- b) **Entorno ou Ambiente** é a coleção de itens externos que atuam ou sofrem ação por algum componente;
- c) **Estrutura** é a coleção de ligações entre componentes e entre esses, e, itens do ambiente. As ligações entre componentes formam a endoestrutura e as ligações entre componentes e itens do ambiente formam a exoestrutura;
- d) **Mecanismo** é a coleção de processos que geram a novidade qualitativa, isto é, que promovem ou obstruem as transformações, causando a emergência, crescimento, manutenção ou o desmantelamento do sistema ou de alguma de suas propriedades.

A composição é atômica. Cada componente é um indivíduo ou átomo do sistema. (BUNGE, 1979). Os itens do ambiente podem ser buscados entre as origens dos fatos sociais que desencadeiam mudanças em sistemas – para Bunge (2003), essas origens podem ser ambientais, biopsicológicas, econômicas, políticas e culturais. Cada componente ou item do ambiente tem pelo menos uma ligação e a coleção de ligações forma a estrutura do sistema. O mecanismo não é diretamente detectável e precisa ser conjecturado. O Quadro 4 exemplifica o modelo CESM aplicado a sistemas existentes.

Quadro 4 – Exemplos de modelos CESM

Sistema	C Composição	E Ambiente	S Estrutura	M Mecanismo
Organismo vivo	Micro- e mesossistemas físicos e químicos (esp. água, carboidratos, lípidos, proteínas e ácidos nucleicos).	Meio rico em nutrientes e fluxos de energia, com variáveis (pressão, temperatura etc.) restritas a pequenos intervalos.	Ligações diretas e indiretas, físicas ou químicas, covalentes e não-covalentes que mantêm unidos os componentes, mais as ligações químicas, físicas e biológicas com o ambiente.	Processos que mantêm o organismo vivo, e g., síntese de algumas moléculas, captura e armazenagem de energia, transporte, rearranjo, montagem e desmontagem de componentes que acompanham o metabolismo.
Empresa	Os funcionários ( <i>personnel</i> ) e a gerência.	Mercado e governo.	As relações de trabalho entre membros da empresa e entre esses e o ambiente.	As atividades que resultam em produtos da empresa.
Comunidade linguística	As pessoas que falam a mesma língua.	A cultura na qual a linguagem é usada.	A coleção de relações de comunicação linguística.	A produção, transmissão e recepção de símbolos.
Sistema solar	Sol, planetas e asteroides.	Demais corpos celestes.	Forças gravitacionais.	Translação dos componentes em órbitas que permitem certa continuidade do sistema (sem dispersão ou colapso) pela inércia.

Fonte: Kern (2011) a partir de Bunge (2003).

#### 4.4 AS REGRAS GERAIS DA PESQUISA SISTEMISTA

A abordagem de problemas recomendada pelo sistemismo pode ser sumariada em sete regras metodológicas gerais (BUNGE, 1997):

- a) colocar todo fato social em seu contexto mais amplo (ou sistema);
- b) dividir cada sistema em sua composição, ambiente e estrutura;
- c) distinguir os vários níveis de sistema e exibir suas relações;
- d) procurar ou conjecturar os mecanismos;
- e) verificar a hipótese ou teoria mecanísmica manipulando experimentalmente as variáveis referidas;
- f) preferir hipóteses, teorias e explicações mecanísmicas (dinâmicas) às fenomenológicas (cinemáticas) e aos modelos de equilíbrio e descrições de dados;
- g) em caso do mau funcionamento do sistema, examinar todas as quatro fontes possíveis – C, E, S, M – e tentar reparar o sistema alterando alguma ou todas as fontes.

Bunge não detalha procedimentos metodológicos específicos. Nesta pesquisa, conforme o capítulo de procedimentos metodológicos a seguir, são aplicadas as duas primeiras regras: (1) identificar ou reconhecer o sistema em questão; (2) descrever os elementos detectáveis do RI, sendo delineado um novo método de trabalho para modelagem proposta. O mecanismo, por ser usualmente oculto e requerer conjectura e teste, foi excluído do escopo da pesquisa.

## 5 PROCEDIMENTOS METODOLÓGICOS

A metodologia proposta para a realização da pesquisa compreende um conjunto de etapas e procedimentos a serem definidos para atender aos objetivos do estudo. Minayo (2007, p. 14) aponta que “a metodologia inclui simultaneamente a teoria de abordagem (o método), os instrumentos de operacionalização (as técnicas) e a criatividade do pesquisador [...]”.

Na sequência, são abordados os itens de caracterização da pesquisa (classificação), as técnicas escolhidas para sua realização, seu universo, assim como o detalhamento dos procedimentos metodológicos que permitiram o alcance dos objetivos propostos.

### 5.1 CARACTERIZAÇÃO DA PESQUISA

A classificação dos tipos de pesquisa se dá à luz do estabelecimento de um critério. Quanto ao critério dos objetivos de estudo propostos, esta pesquisa caracteriza-se como descritiva e exploratória.

Descritiva porque procura observar, registrar, classificar e interpretar as variáveis. Busca descobrir com a máxima precisão possível a frequência com que determinado fenômeno ocorre, sua natureza e característica (CERVO; BERVIAN; SILVA, 2007), ao passo que descreve a composição, ambiente e estrutura dos RIs.

É exploratória porque busca maior familiaridade com o objeto de estudo. As pesquisas exploratórias, segundo Gil (2002), são realizadas em áreas nas quais o conhecimento é mais escasso. Objetivam desenvolver, esclarecer e modificar conceitos e ideias para que novos estudos tenham maior precisão na formulação e problemas e hipóteses. No âmbito da pesquisa sistemista, utilizando aqui o modelo CESH, poucos registros são encontrados na literatura.

Quanto aos procedimentos utilizados para coleta de dados, caracteriza-se como bibliográfica por meio da análise e abstração de dados do material-fonte. O universo da pesquisa é a literatura científica sobre RI com amostragem intencional/opportunística de três textos abrangentes e críticos sobre RI, juntamente com duas entrevistas de peritos da área, conforme detalhamento da seção a seguir. Considera-se que as opções de coleta de dados adotadas nessa pesquisa tendem a ser referência para as próximas pesquisas de descrição sistemista, pois na

literatura há poucos textos abrangentes que resultem em modelagem rigorosa e rastreável.

A abordagem da pesquisa é ontológica, recorre à filosofia, busca interpretar e descrever o mundo (de forma não subjetiva) a partir de um modelo de sistema que vem da filosofia sistemista (que é realista, não subjetivista) de Mario Bunge.

## 5.2 PROCEDIMENTOS DE COLETA E ANÁLISE DE DADOS

Nesta pesquisa, descreve-se um RI como sistema social e técnico, abrangendo os dois primeiros itens das regras gerais da pesquisa sistemista de Bunge (1997), citadas na seção 4.5: (1) colocar todo fato social em seu contexto mais amplo (ou sistema), (2) dividir cada sistema em sua composição, ambiente e estrutura. Bunge não dá diretrizes detalhadas de como elaborar a descrição de componentes, itens do ambiente e ligações do sistema. Por isso, foi necessário criar diretrizes nesta pesquisa.

A abordagem metodológica é a abstração de componentes, itens do ambiente e ligações estruturantes entre componentes e desses com o ambiente. Trata-se de uma pesquisa aplicada que requer conhecimento do modelo CESM (composição, ambiente e estrutura) e a abstração dos elementos do sistema a partir da observação e análise de material fonte. Exclui-se da modelagem o mecanismo para se ater aos aspectos observáveis do sistema, uma vez que o mecanismo costuma ser oculto, requerendo conjectura e teste (BUNGE, 1997, 2003).

Considerou-se que a coleta de dados pode ser bibliográfica, documental, por meio de observação, levantamento ou entrevista, ou uma combinação dessas. Quaisquer que sejam as fontes que expressem os elementos (composição, ambiente e estrutura do RI), parecem relevantes para uma análise sistêmica. Optou-se nesta pesquisa pela coleta de dados na literatura científica porque os elementos estão explícitos, a fonte é objeto de análise rigorosa e a coleta é rastreável. Posteriormente, foi ampliada a coleta de dados a partir de entrevistas com especialistas sobre os atores e fatores importantes para os RIs, com posterior discussão e refinamento do modelo no grupo de pesquisa, pois a literatura fornece bons elementos, mas, não os esgota facilmente.

As etapas da pesquisa incluem (1) a seleção de fontes abrangentes, (2) a marcação dos componentes, itens do ambiente e ligações reconhecidos em cada fonte pelo grupo de pesquisadores nomeados no Quadro 5, (3) discussão em grupo dos elementos marcados pelos diferentes pesquisadores, (4) a síntese e consolidação de um

modelo de RI por cada texto fonte, (5) discussão da síntese com peritos na área, e finalmente (6) síntese de um único modelo abrangendo a análise, discussão e consolidação de todas as fontes analisadas.

Quadro 5 - Grupo de pesquisadores

<b>Pesquisador</b>	<b>Vínculo</b>
Luciana Mara Silva	Mestranda do PGCIN/UFSC e bibliotecária da UNISUL
Robson Garcia Formoso	Mestrando do PPGECC/UFSC e professor da Faculdade SENAI
Vinícius Medina Kern	Professor do PGCIN/UFSC e do PPGECC/UFSC

Fonte: Elaborado pela Autora, 2013.

Na sequência, apresenta-se cada uma dessas etapas com o detalhamento de sua execução. Conforme já mencionado, Bunge não detalha os procedimentos para pesquisa sistêmica, julga-se então que a partir das ações executadas, cria-se uma nova metodologia de trabalho para descrição sistêmica de sistemas de informação. Nesse contexto, mesmo não sendo o objeto de estudo dessa pesquisa, os procedimentos são apresentados em grande nível de detalhamento, dada a novidade metodológica.

### **1 - Seleção de fontes abrangentes**

No início da pesquisa, acreditava-se ser possível fazer uma revisão sistemática da literatura, com base na pergunta “quais são os componentes, itens do ambiente e ligações estruturantes em um repositório institucional?” Foram utilizadas para o levantamento bibliográfico as bases de dados: Base de Dados de Artigos de Periódicos em Ciência da Informação (BRAPCI), *Web Of Science*, *Scopus* e *Scielo*, por serem bases relevantes para área da Ciência da Informação e por disponibilizarem artigos em texto completo. Foram realizadas as buscas por meio das palavras-chave: “repositório institucional”, “repositórios institucionais”, “*institutional repository*”, “*institutional repositories*”. Livros e relatórios de instituições de pesquisa também foram consultados. Nessa fase da pesquisa foram levantados aproximadamente 100 títulos. Essa etapa foi realizada no segundo semestre de 2012, contudo novas pesquisas foram realizadas para leitura de artigos mais recentes que complementaram a revisão da literatura. O período

delimitado para pesquisa foram os últimos 10 anos, data marco do movimento de acesso livre.

Contudo, reconheceu-se ser um procedimento inviável, pois requereria competência em quantidade e qualidade de recursos distantes do que se tinha disponível. Detectada a inviabilidade da revisão sistemática, optou-se então pela coleta de dados em textos selecionados oportunisticamente na literatura científica sobre RI, com a intenção de obter uma visão abrangente, cobrindo as possíveis origens dos fatos sociais citados por Bunge (2003): ambiental, biopsicológica, econômica, política e cultural. Restringiu-se a coleta na literatura a três textos fontes que se considerou permitir identificar os componentes e itens do ambiente essenciais. Adicionalmente, foram entrevistados dois peritos em diversos aspectos dos RIs, quanto às relações e interações existentes entre os componentes e, entre esses e o ambiente. As fontes eleitas são apresentadas no Quadro 6.

Quadro 6 - Fontes de dados para coleta

Textos Fontes	Peritos na área entrevistados
<p>(1) <b>RI é a resposta, mas qual é a pergunta?:</b> primeiras anotações para implementação de repositório institucional (GUIMARÃES; SILVA; NORONHA, 2009).</p> <p>(2) <b>Invisible institutional repositories:</b> addressing the low indexing ratios of IRs in Google (ARLITSCH; O'BRIEN, 2012).</p> <p>(3) <b>It's a repository, it's a depository, it's an archive...:</b> open access, digital collections and value (GUEDON, 2009).</p>	<p>(1) <b>Eloy Antônio Santos Cordeiro Rodrigues</b><sup>6</sup> – Universidade do Minho</p> <p>(2) <b>Hélio Kuramoto</b><sup>7</sup> – IBICT/ UFMG</p>

Fonte: Elaborado pela Autora, 2013.

<sup>6</sup> Diretor dos Serviços de Documentação da Universidade do Minho. Em 2003, liderou a criação do RepositoriUM – o repositório institucional da Universidade do Minho. Atua na promoção e divulgação do acesso livre ao conhecimento e a criação de repositórios institucionais em Portugal.

<sup>7</sup> Doutor em Ciências da Informação e da Comunicação. Atualmente é Tecnologista Sênior HIII do Instituto Brasileiro de Informações em Ciência e Tecnologia (IBICT) e docente na UFMG. Atua na área de Ciência da Informação com ênfase nos temas: organização e tratamento da informação; comunicação científica; e tecnologias da informação e da comunicação. Nos últimos 10 anos, este pesquisador tem se dedicado à causa do Acesso Livre à Informação Científica.

O primeiro texto fonte foi eleito por ser considerado um artigo de grande abrangência teórico-conceitual e teor crítico: “RI é a resposta, mas qual é a pergunta?” (GUIMARÃES; SILVA; NORONHA, 2009). A análise desse primeiro texto, por Silva, Formoso e Kern (2013), serviu como teste da metodologia adotada e teve seus resultados apresentados no XIV Encontro Nacional de Pesquisa em Ciência da Informação<sup>8</sup> (Apêndice A).

O segundo texto fonte foi adicionado a partir da busca sistemática já realizada, pois aborda um assunto de extrema importância para o funcionamento do RI: a visibilidade na Web. O terceiro texto fonte que complementa a amostra foi eleito por ser um texto com abrangência na área e com autoridade reconhecida na literatura científica para o assunto RI.

## **2 - Marcação dos componentes, itens do ambiente e ligações reconhecidos**

Nessa etapa buscou-se atender ao primeiro objetivo da pesquisa: identificar os termos que descrevem elementos de um sistema de RI (componentes, itens do entorno/ambiente e ligações estruturantes). Esses termos são em geral substantivos que representam atores com algum papel importante no sistema. Nem todos os substantivos representam atores, mas todo autor está associado a um ou mais termos.

Convencionou-se uma forma de marcação do texto fonte com cores no texto para uma melhor visualização dos elementos marcados e posterior análise, discussão e consolidação do modelo proposto. Foi elaborada então uma legenda de cores conforme exemplifica Quadro 7.

Quadro 7 - Legenda de cores para destaque nos textos fontes

	Destaque no texto para os componentes
	Destaque no texto para os itens do ambiente
	Destaque no texto para as ligações estabelecidas entre os componentes e/ou destes com itens do ambiente.

Fonte: Elaborado pela Autora, 2013.

Para facilitar a marcação dos elementos por cores no texto, os arquivos em PDF dos textos fontes foram convertidos para Word. Essa etapa foi concebida a partir de um trabalho em grupo: três pesquisadores

<sup>8</sup> Encontro realizado em Florianópolis entre os dias 29 de outubro a 1 de novembro de 2013. <http://enancib.sites.ufsc.br/index.php/enancib2013/XIVenancib>

na análise do texto fonte (1) e dois pesquisadores na análise dos textos fontes (2-3). Cada um dos pesquisadores fez a leitura do texto com a abstração dos elementos conforme metodologia proposta e ilustração do Quadro 8.

Ficou determinado que os elementos marcados que se repetissem no texto, não seriam destacados em duplicidade.

Quadro 8 - Extrato do texto fonte 1, marcado com elementos de um repositório institucional: componentes (vermelho), itens do ambiente (verde) e ligações (azul claro)

Essa visão, também mais alinhada com o **Movimento do Livre Acesso**, mas não restritiva aos **artigos** **revisados pelos pares**, equaciona aumento de visibilidade com livre **acesso**, e aponta para o caráter mandatório de **auto-arquivamento** que emana das **instituições**. Aqui, o RI estaria também a serviço de *uma estratégia de gestão institucional da pesquisa, como um guia para avaliação e, principalmente,*

Fonte: Elaborado pela Autora, 2013.

A marcação exemplificada no Quadro 8 pode apresentar diferenças entre os pesquisadores, que devem debater a pertinência de suas concepções e consensuar uma marcação do grupo. O processo de consolidação é descrito a seguir.

### **3 - Discussão em grupo dos elementos marcados pelos diferentes pesquisadores**

Houve diferenças nas marcações dos textos fonte a partir das dificuldades encontradas na análise e abstração dos elementos do texto pelos pesquisadores. Foram realizados encontros presenciais para discussão e acordo da marcação de cada texto fonte analisado.

As marcações individuais dos pesquisadores participantes da pesquisa foram reunidas para que se possa discutir em grupo as concepções de modelagem e consensuar um modelo parcial derivado da fonte coletada.

Para facilitar a análise e discussão das marcações simultâneas e/ou divergências foi organizado um quadro com os itens destacados por cada um dos pesquisadores em ordem alfabética para cada uma das partes: componentes, ambientes e ligações. O Quadro 9 exemplifica a técnica adotada.

Elementos marcados por ambos os pesquisadores eram consolidados e elementos divergentes discutidos para posterior consolidação. Para a discussão de elementos divergentes, recorria-se ao texto fonte para saber o contexto em que o termo estava inserido, conforme exemplifica o Quadro 10.

Quadro 9 - Análise comparativa dos pesquisadores (exemplo: componentes)

<b>COMPONENTE</b>	<b>Pesquisador 1</b>	<b>Pesquisador 2</b>	<b>Pesquisador 3</b>
	Artigo científico Artigos revisados Conteúdos intelectuais Copyright Metadados Pesquisadores Profissionais de informação e das bibliotecas Publicação científica	Artigo científico Artigos revisados pelos pares Pesquisador individual Preprints Profissionais de informação e das bibliotecas	Artigo científico Artigos revisados pelos pares Conteúdos intelectuais Hardware Pesquisadores Preprints Profissionais de informação Softwares Usuário

Fonte: Elaborado pela Autora, 2013.

Quadro 10 - Contexto do termo no texto fonte

(continua)

<b>Termo</b>	<b>Trecho no texto</b>	<b>Componente ?</b>	<b>Item do ambiente?</b>	<b>Ligação ?</b>
Artigos científicos	De todas as lições, a mais importante, e também a mais óbvia: a plena realização do potencial dos RIs, enquanto ampliação do acesso a artigos científicos revisados por pares, depende fundamentalmente de sua integração/harmonização com os mecanismos de avaliação científica em curso.	Documento		
Preprints	Um RI típico possui, em média, uma coleção de cerca de 3.800 objetos digitais, de tipologia	Documento		

	diversificada, de <i>preprints</i> a <i>postprints</i> , relatórios técnicos, material multimídia, dentre outros.			
Instituição de pesquisa	[...] No segundo caso, o RI apresenta-se mais como estratégia de gestão de conteúdos intelectuais de uma instituição de pesquisa.		Instituição	

Fonte: Elaborado pela Autora, 2013.

O resultado dessa marcação em cada texto fonte selecionado está representado nos Anexos A, B e C, respectivamente.

#### **4 - Síntese e consolidação de um modelo de RI para cada texto fonte**

A partir da consolidação dos elementos abstraídos de cada texto fonte explicitados no passo 3, mais uma rodada de discussão foi realizada com os pesquisadores. Essa etapa de síntese é necessária porque os textos citam elementos do modelo, mas, também, seus sinônimos, instâncias, coletivos, partes, atributos e propriedades – o que torna desafiadora a tarefa de marcar os elementos. A abstração requer essa etapa final de refinamento para atender ao segundo objetivo específico da pesquisa: sintetizar a composição, ambiente e estrutura abstraída de cada um dos textos analisados. O Quadro 11 exemplifica a síntese de uma terminologia. Esta etapa foi realizada com todos os elementos destacados no texto fonte.

Os modelos de RI representando cada texto fonte são apresentados na seção 6: apresentação de discussão dos resultados.

Quadro 11 - Exemplo da síntese terminológica de um componente

Elemento do modelo	Termos correspondentes no texto a partir da consolidação dos pesquisadores	Síntese da terminologia
Componentes	Artigo científico, artigo, pré-print, publicação acadêmica, publicação científica, coleção digital, produção científica da instituição, literatura cinzenta, documentos revisados por pares, <i>paper</i> acadêmico, publicações, produção intelectual, PDF documentos, trabalhos acadêmicos, objetos digitais, teses, <i>cases</i> , artigo de revista, livros, capítulo de livros, trabalho apresentado em conferência, dissertação.	Documento

Fonte: Elaborado pela Autora, 2013.

### 5 - Discussão da síntese com peritos na área

A partir do resultado da modelagem de cada texto fonte baseados na literatura, levantou-se os elementos essenciais de um RI, porém, restaram lacunas - a literatura explicita os elementos, mas é limitada ao mencionar as relações entre eles. Assim, buscou-se conhecer o entendimento de peritos na área, para explicitar as relações, conhecer sua concepção sobre os elementos essenciais que compõem e que formam o ambiente (ou entorno) de RI, bem como as relações que se estabelecem entre esses elementos.

O contato para solicitação da entrevista com os peritos foi realizado por e-mail. A entrevista foi agendada via Skype. Antes da entrevista foi enviado por e-mail o pôster apresentado no ENANCIB (Apêndice A), para contextualizar os entrevistados na pesquisa, assim como, as questões abaixo relacionadas que nortearam a entrevista:

- a) considera que os componentes e itens do ambiente levantados em nosso modelo são essenciais?
- b) postula a existência de algum outro elemento adicional ou alternativo aos presentes modelos?
- c) quais considera que são as ligações, as relações existentes entre os componentes e itens do ambiente do repositório institucional?

- d) em especial, interessa identificar/explicitar as ligações que são mobilizadas nos processos que levam ao sucesso ou insucesso dos RIs;
- e) de que forma observa as políticas de informação dos RIs/mandatos, relevantes para estreitar a ligação pesquisador-instituição-usuário?

A discussão dos resultados compartilhou a percepção dos entrevistados referente ao modelo de RI apresentado. As falas foram evidenciadas no texto com a menção de citação ao entrevistado 1 e/ou 2, conforme contexto apresentado. O Entrevistado 1, foi o Prof. Eloy António Santos Cordeiro Rodrigues e o entrevistado 2, foi o Prof. Hélio Kuramoto. As entrevistas foram realizadas em 4 de dezembro e 11 de dezembro, respectivamente.

### **6 - Síntese de um único modelo abrangendo a análise, discussão e consolidação de todas as fontes analisadas**

Finalmente chega-se à última etapa em cumprimento ao terceiro objetivo específico da pesquisa: sintetizar os modelos de RIs abstraídos para cada fonte em um modelo genérico de sistema de RI.

Com os três modelos parciais consolidados e entrevistas realizadas foi realizada a análise comparativa de cada texto fonte (Quadro 12), identificando as semelhanças, divergências e lacunas. Retornou-se a alguns conceitos advindos do texto para discutir e consolidar a modelagem final apresentada. Nessa etapa a contribuição dos entrevistados e pesquisadores do grupo foi fundamental para apresentar as ligações internas do sistema (entre componentes) e externas (componente-item do ambiente) que a literatura científica contempla. As entrevistas em especial proporcionaram a leitura da prática *versus* teoria na concepção dos repositórios, elementos e ligações importantes que nem sempre aparecem na literatura da área.

Quadro 12 - Exemplo das três marcações

(continua)

	<b>TEXTO FONTE (1)</b>	<b>TEXTO FONTE (2)</b>	<b>TEXTO FONTE (3)</b>
COMPONE NTE	<b>Documento</b> [artigo científico, artigos revisados por pares, pré- print, postprints, publicação	<b>Documento</b> [publicação acadêmica, <i>paper</i> acadêmico, publicações, produção	<b>Documento</b> [publicação científica, publicação acadêmica, coleção digital,

	acadêmica, e-prints, teses, dissertações eletrônicas, coleções especiais digitalizadas, materiais multimídia, materiais de cursos, relatórios técnicos], <b>Pesquisador</b> [comunidade científica] (depositante), pesquisadores (depositante), <b>Staff do RI</b> [profissionais da informação, profissionais de TI]	intelectual, PDF documentos, trabalhos acadêmicos, resultado de investigação, objetos digitais, teses, dissertações, <i>cases</i> , artigo de revista, pré-prints, livros, capítulo de livros, trabalho apresentado em conferência, <b>Pesquisador</b> [corpo docente, discente e técnico da Instituição], <b>Staff do RI</b> [equipe multidisciplinar], <b>software gestor</b> (Dspace, Digital Commons, E-Prints, IR+, ContentDM, DigiTool, ArXiv).	produção científica da instituição, literatura cinzenta, artigos, dissertação, artigos revisados por pares, resultados de pesquisa], <b>Pesquisador</b> [corpo docente], <b>Staff do RI</b> [categorias de pessoas envolvidas]
--	---	--	--

Fonte: Elaborado pela Autora, 2013.

Com os dados coletados e o modelo elaborado, passa-se para a seção 6, com a apresentação e discussão dos resultados dessa pesquisa.



## 6 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Seguindo as etapas detalhadas na seção anterior, chegou-se em uma modelagem para cada texto fonte e um modelo genérico para representação do RI. Os resultados serão representados da seguinte forma: (1) exposição de um quadro com todos os elementos destacados e consolidados de cada texto fonte; (2) representação gráfica dos elementos e respectivas ligações de cada texto fonte; (3) apresentação dos principais dados coletados nas entrevistas; (4) apresentação do modelo genérico proposto seguido da discussão dos resultados.

A modelagem do texto fonte (1) “RI é a resposta, mas qual é a pergunta?”, de Guimarães, Silva e Noronha (2009) está representada no Quadro 13, e as ligações entre os seus componentes e/ou entre estes e os itens do ambiente representada na Figura 1.

Quadro 13 - Elementos [e seus sinônimos e correlatos] do modelo de sistema “repositório institucional” do texto fonte (1)

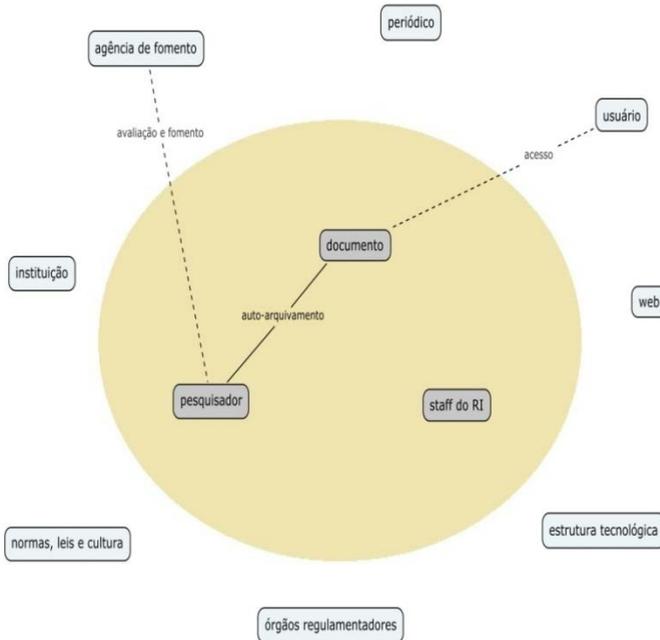
(continua)

<b>Componentes</b>	<b>Documento</b> [artigo científico, artigos revisado por pares, <i>pré-print</i> , <i>postprints</i> . publicação acadêmica, <i>e-prints</i> , teses, dissertações eletrônicas, coleções especiais digitalizadas, materiais multimídia, materiais de cursos, relatórios técnicos], <b>pesquisador</b> [comunidade científica (depositante), pesquisador (depositante)], <b>staff do RI</b> [profissionais da informação, profissionais de TI].
<b>Itens do entorno (ambiente)</b>	<b>Instituição</b> [bibliotecas, guia para avaliação, instituição de pesquisa, mecanismos de avaliação, micro-macro políticas locais e globais, política de desenvolvimento do RI, políticas infraestruturais, universidade], <b>usuários</b> [membros da sua comunidade (comunidade acadêmica), comunidade científica, pesquisadores interessados], <b>periódicos</b> [editoras científicas comerciais, políticas dos editores comerciais], <b>normas, leis e cultura</b> [caráter mandatório de depósito, <i>copyright</i> , movimento de acesso livre, padrão OAI, propriedade intelectual], <b>órgãos reguladores</b> [micro-macro políticas locais e globais, organizações governamentais], <b>estrutura tecnológica</b> [ <i>hardware</i> , infraestrutura tecnológica, metadados, <i>software</i> ], <b>web</b> [internet]. <b>Agência de fomento</b> .
<b>Ligações</b>	<b>Autoarquivamento</b> (pesquisador-documento), <b>ligações</b>

<b>(e elementos ligados)</b>	<b>de avaliação e fomento</b> [recompensa dos pesquisadores, avaliação científica] (pesquisador-agência de fomento), <b>acesso</b> [acesso aos conteúdos] (usuário-documento).
------------------------------	--

Fonte: Elaborado pela Autora, 2013.

Figura 1- Representação gráfica da modelagem a partir do texto fonte (1)



Fonte: Elaborado pela Autora, 2013.

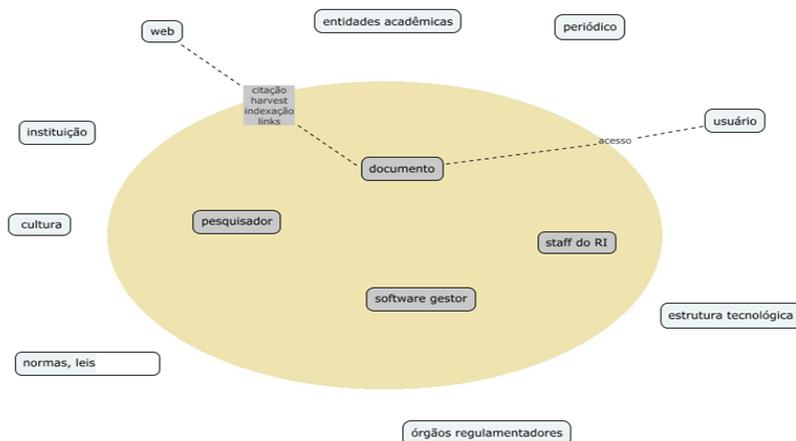
A modelagem do texto fonte (2) “*Invisible institutional repositories: addressing the low indexing ratios of RIs in Google*”, de autoria de Arlitsch e O'Brien (2012), é apresentada no Quadro 14, com representação gráfica na Figura 2.

Quadro 14 - Elementos [e seus sinônimos e correlatos] do modelo de sistema “repositório institucional” do texto fonte (2)

<b>Componentes</b>	<b>Documento</b> [publicação acadêmica, <i>paper</i> acadêmico, publicações, produção intelectual, PDF documentos, trabalhos acadêmicos, resultado de investigação, objetos digitais, teses, dissertações, disser case, artigo de revista, pré-prints, livros, capítulo de livros, trabalho apresentado em conferência, dissertação], <b>pesquisador</b> [corpo docente, discente e técnico da Instituição], <b>staff do RI</b> [equipe multidisciplinar], <b>software gestor</b> [Dspace, Digital Commons, EPrints, IR+, ContentDM, DigiTool, ArXiv].
<b>Itens do entorno ou ambiente</b>	<b>Instituição</b> [instituições acadêmicas, bibliotecas acadêmicas, bibliotecas, universidades], <b>usuários</b> [corpo docente ativo, estudantes universitários, público, usuário, pesquisadores acadêmicos], <b>normas e leis</b> [copyright, Dublin core, metadados, OAI-PMH], <b>web</b> [ <i>crawlers</i> , provedor de dados, google, scholar google, Open doar, sites de pesquisa orientada, motor de busca, <i>websites</i> , yahoo], <b>estrutura tecnológica</b> [tecnologia/internet], <b>órgãos regulamentadores</b> [OCLC, ranking universitário], <b>cultura</b> [movimento acesso aberto, acesso aberto, consentimento do autor], <b>periódicos</b> [repositórios de editora, indústria editorial] <b>entidades acadêmicas</b> [sociedades acadêmicas].
<b>Ligações (e elementos ligados)</b>	<b>Acesso/web</b> [citação das publicações], Harvest, indexação, citação do repositório, recuperação da informação, <i>links</i> , métricas] ( <i>web</i> -publicação), <b>Acesso/documento</b> [acesso aos conteúdos] (usuário-documento).

Fonte: Elaborado pela Autora, 2013.

Figura 2 - Representação gráfica da modelagem a partir do texto fonte (2)



Fonte: Elaborado pela Autora, 2014.

A modelagem do texto fonte (3) “*It’s a repository, it’s a depository, it’s anarchiv...: open access, digital collections and value*”, de autoria de Guedon (2009), está representada no Quadro 15 e, as ligações entre os seus componentes e/ou entre estes e os itens do ambiente representada na Figura 3.

Quadro15 - Elementos [e seus sinônimos e correlatos] do modelo de sistema “repositório institucional” do texto fonte (3)

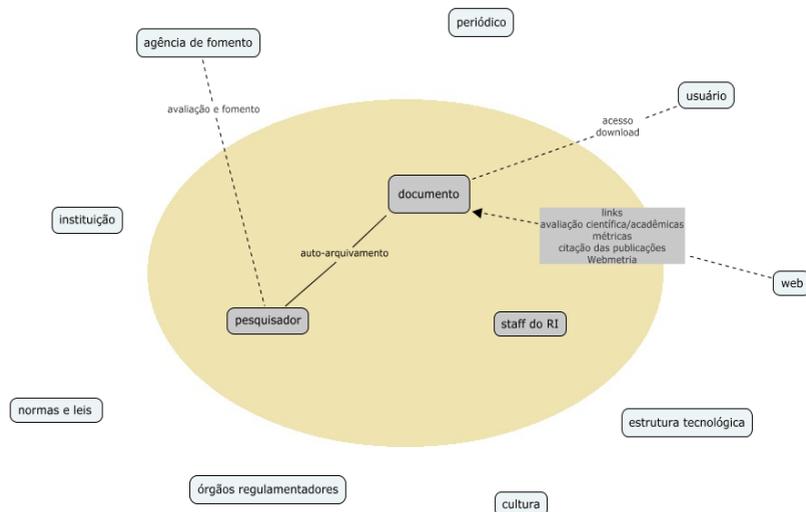
(continua)

<b>Componentes</b>	<b>Documento</b> [publicação científica, publicação acadêmica, coleção digital, produção científica da instituição, literatura cinzenta, artigos, dissertação, artigos revisados por pares, resultados de pesquisa], <b>pesquisador</b> [corpo docente], <b>staff do RI</b> [categorias de pessoas envolvidas, bibliotecários].
<b>Itens do entorno ou ambiente</b>	<b>Instituição</b> [gestores universitários, bibliotecas, universidade], <b>entidades acadêmicas</b> [consórcios de bibliotecas e associações], <b>cultura</b> [acesso livre, recursos culturais], <b>órgãos regulamentadores</b> [recursos econômicos, recursos políticos, rankings da universidade], <b>periódicos</b> [gatekeeper das revistas, periódicos de prestígio, editoras comerciais, editoras universitárias, editores, editores chefe] <b>normas e leis</b>

	[política obrigatória, mandato de depósito, copyright, regras do repositório, Creative Commons, metadados, OAI-PMH], <b>usuários</b> [pesquisadores e estudantes], <b>agência de fomento</b> [agências de concessão de pesquisa públicas e privadas, agências de financiamento da educação, agências de financiamento], <b>estrutura tecnológica</b> [infraestrutura, internet], <b>web</b> [Opendoar, motores de busca, Scholar Google, Oaister].
<b>Ligações (e elementos ligados)</b>	<b>Autoarquivamento</b> (pesquisador-documento), <b>ligações de avaliação e fomento</b> (pesquisador-agência de fomento), <b>acesso documento</b> [acesso aos conteúdos] downloads, (usuário-publicação) <b>acesso web</b> [links, métricas, citação das publicações, <i>webometria</i> ] ( <i>web-publicação</i> ).

Fonte: Elaborado pela Autora, 2014.

Figura 3 - Representação gráfica da modelagem a partir do texto fonte (3)



Fonte: Elaboração da Autora, 2013.

Destaca-se, a seguir, os principais dados coletados das entrevistas realizadas com os peritos da área.

Quanto à modelagem do RI apresentada e ao questionamento se os componentes e itens do ambiente levantados no modelo são

essenciais, ambos os entrevistados afirmaram que sim. Neste aspecto, o entrevistado 1 apresentou uma consideração importante: o software enquanto “software gestor” considera um componente e não um item do ambiente, conforme estava apresentado na modelagem do texto fonte (1). Essa observação foi revista pelo grupo e consensuada no modelo genérico de RI apresentado na Figura 4.

No questionamento das ligações apresentadas e/ou essenciais que supostamente a literatura não apresenta, destaca-se os seguintes dados coletados das entrevistas:

- a) é preciso dar ênfase ao papel dos RIs na avaliação dos pesquisadores na universidade. Esta avaliação está representada pela ligação pesquisador depositante x instituição. (Entrevistado 1);
- b) políticas e mandatos dos RIs são determinantes para o seu sucesso (Entrevistados 1 e 2);
- c) a intervenção pública para o incentivo a construção e povoamento dos RI é essencial para o seu sucesso (Entrevistado 2);
- d) é preciso incentivar os pesquisadores para o autodepósito enquanto a visibilidade/impacto das suas produções. Estatísticas de uso do RI devem ser repassadas aos pesquisadores (Entrevistado 1);
- e) o autodepósito é fator determinante para o sucesso do RI (Entrevistado 1 e 2);
- f) o RI é do pesquisador, da Universidade “não é propriedade” da biblioteca. Logo, o envolvimento de todas as partes é fundamental para o povoamento do RI (entrevistado 1);
- g) é preciso que mudanças no atual modelo de negócio dos periódicos científicos ocorram, para que as manifestações a favor da informação científica em acesso aberto tenham êxito (entrevistado 2).

Quando questionados sobre a concepção dos repositórios a favor do movimento de acesso aberto a informação ou a serviço de informação de uma instituição de pesquisa, vale destacar:

- a) o RI tem o foco na produção científica com abertura para produção intelectual da universidade (entrevistado 2);
- b) as duas visões são compatíveis. A concepção do RI deve estar focada na produção científica e gradativamente pode acrescentar a produção intelectual da universidade – essa ação infere em maior credibilidade ao RI (entrevistado 1).

A partir das três modelagens apresentadas, textos fonte 1, 2 e 3, respectivamente, unificou-se os dados abstraídos das entrevistas

juntamente com o conhecimento adquirido dos pesquisadores, mediante estudos da literatura da área e chegou-se ao modelo genérico que segue (Quadro 16 e Figura 4), atendendo ao terceiro objetivo específico de “sintetizar os modelos parciais em um modelo genérico de sistema de RI”.

Quadro16 - Elementos [e seus sinônimos e correlatos] do modelo de sistema “repositório institucional” consolidado

(continua)

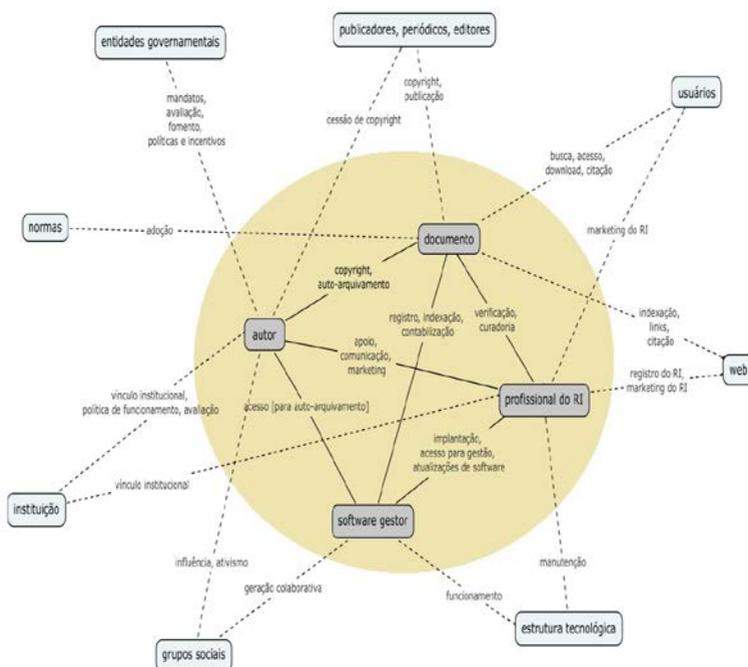
<p><b>Componentes</b></p>	<p><b>Documento</b> [artigo científico, artigo de revista, artigos revisado por pares, artigos, capítulo de livros, case, coleção digital, coleções especiais digitalizadas, dissertação, dissertações eletrônicas, dissertações, <i>e-prints</i>, literatura cinzenta, livros, materiais de cursos, materiais multimídia, objetos digitais, <i>paper</i> acadêmico, PDF documentos, postprints, pré-prints, produção científica da instituição, produção intelectual, publicação acadêmica, publicação científica, publicações, relatórios técnicos, resultado de investigação, resultados de pesquisa, teses, trabalho apresentado em conferência, trabalhos acadêmicos]; <b>autor</b> [comunidade científica (depositante), pesquisador (depositante), corpo docente, discente e técnico da Instituição, corpo docente]; <b>profissional do RI</b> [profissionais da informação, profissionais de TI, equipe multidisciplinar, categorias de pessoas envolvidas, bibliotecários] <b>software gestor</b> [Dspace, Digital Commons, EPrints, IR+, ContentDM, DigiTool, ArXiv].</p>
<p><b>Itens do entorno ou ambiente</b></p>	<p><b>Instituição</b> [bibliotecas acadêmicas, bibliotecas, gestores universitários, guia para avaliação, instituição de pesquisa, instituições acadêmicas, mecanismos de avaliação, micro-macro políticas locais e globais, política de desenvolvimento do RI, políticas infraestruturais, universidade]; <b>usuários</b> [comunidade científica, corpo docente ativo, estudantes universitários, membros da sua comunidade (comunidade acadêmica), pesquisadores acadêmicos, pesquisadores e estudantes, pesquisadores interessados, público, usuário]; <b>normas</b> [<i>copyright</i>, propriedade intelectual, creative commons, dublin core, mandato de depósito, metadados, OAI-PMH, padrão OAI, política obrigatória, regras do</p>

	<p>repositório]; <b>web</b> [crawlers, scholar google, google, internet, motor de busca, motores de busca, Oaister, Opendoar, provedor de dados, sites de pesquisa orientada, ranking universitário, rankings da universidade <i>websites</i>, yahoo]; <b>estrutura tecnológica</b> [<i>hardware</i>, infraestrutura tecnológica, infraestrutura, internet, metadados, <i>software</i>, tecnologia/internet]; <b>grupos sociais</b> [acesso aberto, acesso livre, caráter mandatário de depósito, consentimento do autor, movimento acesso aberto, movimento de acesso livre, recursos culturais, consórcios de bibliotecas e associações, sociedades acadêmicas, micro-macro políticas locais e globais, OCLC, recursos econômicos, recursos políticos]; <b>publicadores, periódicos e editores</b> [editoras científicas comerciais, editoras comerciais, editoras universitárias, editores chefe editores, <i>gatekeeper</i> das revistas, indústria editorial, periódicos de prestígio, políticas dos editores comerciais, repositórios de editora]; <b>entidades governamentais</b> [agências de concessão de pesquisa públicas e privadas, agências de financiamento, agências de financiamento da educação, organizações governamentais, agência de fomento]</p>
<p>Ligações (elementos ligados)</p>	<p><b>Acesso (busca/download)</b> [documento-usuários]; <b>acesso (citação-visibilidade/ indexação-crawling/haversting, contagens de acesso/links)</b> [documento-<i>web</i>]; <b>adoção/uso de padrões/metadados</b> [normas <math>x</math> documento] <b>apoio/mediação/comunicação e marketing</b> [profissional do RI-autor]; <b>autoarquivamento/copyright</b> [autor-documento]; <b>acesso para autoarquivamento</b> [autor-<i>software</i> gestor]; <b>cessão de copyright</b> [autor-publicadores, periódicos, editores]; <b>copyright/publicação</b> [documento-publicadores, periódicos, editores]; <b>funcionamento</b> [<i>software</i> gestor-estrutura tecnológica]; <b>geração colaborativa</b> [grupos sociais-<i>software</i> gestor]; <b>implantação, acesso para gestão e atualização de <i>software</i></b> [profissional do RI-<i>software</i> gestor]; <b>influência/ativismo</b> [grupos sociais-autor]; <b>manutenção</b> [profissional do RI-estrutura tecnológica]; <b>marketing do RI</b> [profissional do RI-usuários] [profissional do RI-<i>web</i>]; <b>mandatos/avaliação de pesquisa/fomentos/ políticas</b></p>

	<p><b>e incentivo</b> [entidades governamentais - autor]; <b>registro/contabilização</b> [<i>software</i> gestor-documento]; <b>verificação/curadoria</b> [profissional do RI – documento]; <b>vínculo institucional</b> [profissional do RI – instituição]; <b>vínculo</b> institucional/ política de funcionamento/ avaliação da produção [autor– instituição].</p>
--	---

Fonte: Elaboração da Autora, 2013.

Figura 4 - Representação gráfica da modelagem consolidada do sistema "repositório institucional"



Fonte: Elaboração da Autora, 2014.

As ligações expressas no Quadro 16 e na Figura 4 são detalhadas no Quadro 17, de modo a favorecer o entendimento preciso das ligações modeladas, haja vista que a denominação, nomenclatura utilizada para expressar as ligações podem não ser suficientes para esclarecer a natureza e característica de cada ligação.

Quadro 17 - Detalhamento das ligações modeladas no sistema "repositório institucional"

(continua)

<b>Elementos ligados</b>	<b>Ligações e comentários</b>
<b>Autor</b> x <b>Software gestor</b>	<b>Acesso [para o autoarquivamento]</b> Essa ligação possibilita que o autor tenha pleno exercício no ato do autoarquivamento do seu documento no <i>software</i> gestor.
<b>Normas</b> x <b>documentos</b>	<b>Adoção (uso de padrões/metadados)</b> Ligação estabelecida entre um documento e a norma que atende a padrões de metadados. Ligação para interoperabilidade de dados no compartilhamento dos documentos, busca e acesso da informação. Essa ligação é fundamental para representação dos RIs na <i>Web</i> , conforme discussão da seção 3.3.
<b>Profissional do RI</b> x <b>Autor</b>	<b>Apoio [mediação]</b> Estabelece a relação de dois atores envolvidos: o profissional do RI e o autor. O profissional do RI é o contato permanente com o pesquisador/depositante para o autodepósito de suas publicações no RI. <b>Comunicação e marketing</b> É a ligação que busca parceria dos autores para povoamento das coleções dos RI, por meio do esclarecimento a esse grupo da função e representação do RI na comunicação científica, discussão da seção 3.2.

<b>Elementos ligados</b>	<b>Ligações e comentários</b>
<p><b>Autor x Documento</b></p>	<p><b>Autoarquivamento</b>  A ligação de autoarquivamento é fundamental para o povoamento dos repositórios tendo como resultado a ampliação do acesso à informação (VEIGA; MACHADO; ALVES, 2014). O êxito desse processo depende do comprometimento dos pesquisadores, um dos atores envolvidos no processo, o prof. Eloy Rodrigues (Entrevistado 1) comenta que é preciso incentivar os pesquisadores para o autodepósito, esclarecer o docente para o impacto e visibilidade de produção via RI. Essa afirmação corrobora com Rosa e Gomes (2010b, p. 158), quando apontam que “o reconhecimento por parte das instituições do potencial dos RIs e a constatação da tendência para baixos valores de autoarquivamento, tem conduzido as instituições à definição e divulgação entre seus membros de políticas institucionais de arquivamento nos RIs, adequadas aos seus contextos específicos.” O prof. Hélio Kuramoto (Entrevistado 2) relata que políticas de informação para concepção e povoamento dos repositórios devem ser planejadas pelas instituições de pesquisa para manter os repositórios em pleno funcionamento e uso. O cenário apresentado implica a necessidade de reavaliação de políticas que permeiam a filosofia de acesso aberto e acesso livre para o fluxo da comunicação científica. Na perspectiva de Harnad (2007) emerge a adoção, em nível mundial pelas universidades, instituições de pesquisa e agências de fomento, de mandatos oficiais de autoarquivamento.</p> <p><b>Copyright</b>  Ligação de autoria e de direito de cópia entre um autor e o documento depositado no RI.</p>

Elementos ligados	Ligações e comentários
<p><b>Documento x Usuário</b></p>	<p><b>Acesso, busca e download</b>            Apresenta-se aqui três ligações correlatas que implicam no usuário ter acesso aos documentos armazenadas nos RI. Esse acesso inicia pela busca da informação no próprio RI e/ou em provedores de busca. Nesse caso, conforme discussão da seção 3.2, a visibilidade dos RIs pode estar sendo negligenciada por adoção de padrões de metadados que interferem na interoperabilidade de dados para pesquisa. Os usuários fazem suas buscas comumente por meio de provedores de serviços, logo, a invisibilidade dos RIs nesses provedores pode diminuir qualquer ação de <i>marketing</i> para divulgação e uso dos RIs, quiçá, garantir o seu povoamento.</p>
<p><b>Documento x Web</b></p>	<p><b>Indexação</b>            Ligação que promove a visibilidade dos documentos, dos autores, das instituições de pesquisa por meio da indexação das coleções dos RI e recuperação na <i>Web</i>, o que inclui os processos de <i>crawling</i> e <i>harvesting</i>.</p> <p><b>Links</b>            Ligação entre um documento depositado no RI e um lugar da <i>Web</i>. Inclui-se aqui as métricas que promovem o ranking dos repositórios pela contagem de acessos/<i>links</i>.</p> <p><b>Citação</b>            Essa ligação promove o impacto das pesquisas e reconhecimento dos autores, por meio da citação de suas produções armazenadas nos RIs. Inclui-se aqui também a literatura cinzenta por meio da citação dos documentos dos RIs pelos usuários a partir da busca, uso e reuso da informação. Costa (2006) faz a abordagem de um processo cíclico: maior acessibilidade é igual ao maior uso que significa maior fator de impacto, que, por sua vez, significa maiores recompensas, que provocam mais produção, a qual, estando mais acessível, provoca maior uso e assim sucessivamente.</p>
<p><b>Autor x publicadore s/ periódicos editores</b></p>	<p><b>Cessão de Copyright</b>            Autor cedendo os direitos de cópia para os editores, na qual a divulgação/veiculação da informação passa a ser definida pelos editores.</p>

Elementos ligados	Ligações e comentários
<b>Documento x publicadores/periódicos editores</b>	<p><b>Copyright publicação</b> Essa ligação interfere no modelo de negócios atual dos detentores da informação científica, os periódicos científicos. Estes, muitas vezes, não autorizam os autores a depositarem a cópia dos artigos publicados em suas revistas ou exigem um período de embargo para a sua divulgação. Enquanto não houver mudanças nesse modelo de negócio os RIs tendem a não alcançar sucesso na proposição da “linha verde” inerente ao movimento de acesso livre.</p>
<b>Documento x Profissional do RI</b>	<p><b>Verificação/curadoria</b> Essa ligação implica na organização dos metadados de documentos publicados. O profissional do RI faz o papel de curador, cuida dos seus dados.</p>
<b>Estrutura tecnológica x Software gestor</b>	<p><b>Funcionamento</b> Ligação que estabelece o funcionamento do RI por meio de <i>software</i>, <i>hardware</i>, acesso a Internet, etc. Apesar de ser tomada como garantida, a falha dessa ligação impede a existência do RI.</p>
<b>Software gestor x Grupos Sociais</b>	<p><b>Geração colaborativa</b> Ligação entre o <i>software</i> gestor do RI e as comunidades que promovem, realizam e mantém o desenvolvimento do <i>software</i> por meio da colaboração em rede, usualmente na modalidade “<i>software</i> livre”.</p>
<b>Profissional do RI x Software gestor</b>	<p><b>Implantação</b> Essa ligação estabelece o funcionamento prático do RI, por meio da implantação do <i>software</i> gestor.’ <b>Acesso para gestão</b> Estabelece o gerenciamento das coleções dos RIs (inclusão, acesso e/ou <i>download</i> dos documentos). <b>Atualizações de <i>software</i></b> Assegura o funcionamento técnico do RI por meio da instalação das atualizações do <i>software</i> gestor.</p>
<b>Autor x Grupos Sociais</b>	<p><b>Influência/ ativismo</b> Essa ligação está estreitamente ligada ao Movimento de Acesso Livre (linha verde). Faz a defensoria da informação científica circulando sem barreiras.</p>

Elementos ligados	Ligações e comentários
<b>Estrutura tecnológica</b> <b>x</b> <b>Profissional do RI</b>	<b>Manutenção</b> Ligação vinculada à manutenção e atualização constante do aparato tecnológico para manter o pleno funcionamento do RI.
<b>Usuários</b> <b>x</b> <b>Profissional do RI</b>	<b>Marketing do RI</b> Estabelece a propagação do RI, por meio de ações de <i>marketing</i> para seu uso e esclarecimento à sua comunidade quanto à função e representação dos RIs no processo de comunicação científica.
<b>Instituição</b> <b>x</b> <b>Autor</b>	<b>Vínculo Institucional</b> Estabelece o vínculo institucional do autor com a instituição de pesquisa. <b>Políticas de funcionamento</b> Estabelece a implantação de políticas de informação para funcionamento do RI. <b>Avaliação [da produção]</b> Estabelece uma ligação que precisa ser fortalecida ou, em muitos casos, implementada nas instituições. Está atrelada à motivação dos autores/pesquisadores no povoamento do RI. Cabe às instituições criarem mecanismos de avaliação para avaliar seus pesquisadores por meio da produção intelectual inserida nos RIs. O repositório da UMinho traz essa boa prática e apresenta resultados positivos no povoamento do RI. No âmbito da promoção do RI como propulsor do movimento de acesso livre, essa política institucional precisa ser fomentada por políticas públicas criadas para que os RIs passem a ser um indicador de avaliação da produção científica.

Elementos ligados	Ligações e comentários
<p style="text-align: center;"><b>Entidades governamentais x Autor</b></p>	<p><b>Mandatos</b> Vinculação do autor à obrigação de depositar. Ligação que estabelece os mandatos obrigatórios para autoarquivamento.</p> <p><b>Avaliação</b> Estabelece a avaliação da produção científica por meio da produção dos pesquisadores depositadas nos RIs. Sistemas de avaliação do pesquisador a partir do autoarquivamento de sua produção nos RIs, vinculados às instituições a qual pertencem é, talvez a mais importante motivação para que o pesquisador permaneça povoando os RI. O RI da Universidade do Minho, <i>case de sucesso</i>, adota esse tipo de avaliação. O entrevistado 1, mentor do RI Minho, enfatiza que “é preciso incentivar os pesquisadores para o autodepósito prospectando a visibilidade/impacto das suas produções. Estatísticas de uso do RI devem ser repassadas aos pesquisadores. Deve haver um equilíbrio entre a obrigatoriedade do autodepósito com os benefícios e/ou recompensas para o pesquisador”.</p> <p><b>Fomento</b> Essa ligação estabelece benefícios tanto para o pesquisador/ usuário quanto para a instituição de pesquisa que pode se beneficiar de recursos provenientes das agências de fomento.</p> <p><b>Políticas e incentivos</b> Estabelece a ligação de políticas que incentivem os autores para depósito de suas produções no RI. No âmbito das políticas públicas que colaboram para causa do movimento de acesso livre, ainda há muito a discutir. É preciso haver uma ponte entre os mecanismos de avaliação em curso e os RIs, para que as iniciativas de RI tenham relevância junto à comunidade científica (GUIMARÃES; SILVA; NORONHA, 2009).</p>
<p style="text-align: center;"><b>Web x Profissional do RI</b></p>	<p><b>Registro do RI/</b> Essa ligação estabelece o registro do RI em diretórios como Open Doar e Roar.</p> <p><b>Marketing do RI</b> Ligação que promove o RI no processo de comunicação científica na <i>Web</i>.</p>

<b>Elementos ligados</b>	<b>Ligações e comentários</b>
<b>Documento x Software gestor</b>	<p><b>Registro e contabilização</b></p> <p>Essa ligação estabelece as estatísticas dos RI, dados importantes para as instituições de pesquisa, autor/pesquisador, e quem sabe, futuramente dados essenciais para entidades governamentais na avaliação da pesquisa acadêmica/científica.</p>
<b>Profissional do RI x Instituição</b>	<p><b>Vínculo institucional</b></p> <p>Estabelece o vínculo institucional do profissional do RI com a instituição.</p>

Fonte: Elaboração da Autora, 2013.

O quadro acima explicitou cada uma das ligações mapeadas com o intuito de dar precisão aos elementos ligados apresentados no modelo genérico de RI. A próxima seção apresenta as considerações finais da pesquisa.

## 7 CONSIDERAÇÕES FINAIS

O objetivo geral da pesquisa foi alcançado por meio da descrição sistêmica de um RI, caracterizando seus elementos: componentes, itens do entorno e ligações estruturantes, resultando na modelagem genérica de um RI.

As principais contribuições da pesquisa foram: (1) uma visão sistêmica do RI: o modelo criado se propõe a descrever "tudo o que é importante" sobre RI; (2) os procedimentos metodológicos adotados para a elaboração de sistema (composição, ambiente e estrutura), a partir da literatura e entrevistas com peritos.

A visão sistêmica é uma descrição que deve favorecer o empreendimento de pesquisas explicativas com hipótese – de associação entre variáveis e também associação com interferência (causais). Foi ilustrado o impacto de análises sistêmicas por meio de uma rápida discussão sobre a visibilidade na *web*, ligação ambiental essencial, porém negligenciada em alguns repositórios, conforme a apresentação dos resultados de uma pesquisa já realizada por Arlitsch e O'Brien (2012), destacando o caráter essencial da **ligação de acesso** para o funcionamento dos RIs. A opção explícita de adoção do padrão Dublin Core pelo *software* gestor é uma decisão que, em princípio, não parece questionável, muito menos equivocada. No entanto, do ponto de vista do "acesso", e considerando a importância do acesso no funcionamento de um RI, a decisão passa a ser questionável em razão da invisibilidade dos RI, conforme a conclusão dos autores. Assim, em uma visão setorial, a adoção do padrão Dublin Core é acertada. Apenas uma visão sistêmica, que considera o todo, pode revelar que essa decisão, aparentemente simples e segura, pode interferir no sucesso do RI.

Conceber um RI em função do movimento de acesso livre (via verde), sem admitir o depósito de outros documentos além de artigos indexados, também parece incorrer numa visão fragmentada, não sistêmica. Isso pode levar à solução de problemas pontuais, mas não necessariamente apoia a comunicação científica da instituição. Como uma ação inovadora, o RI pode estar a serviço de uma instituição de pesquisa por meio do depósito de dados brutos de pesquisa (curadoria de dados). É função de uma universidade receber e preservar os dados das pesquisas desenvolvidas. Se o repositório "restringe" o autoarquivamento apenas aos artigos indexados em periódicos científicos, parece que há um distanciamento com uma das funcionalidades do RI que é a preservação da produção intelectual de

uma instituição de pesquisa. Nesse contexto, considera-se que as duas concepções de repositório citadas na pesquisa: (1) a favor do movimento de acesso livre; (2) como serviço de informação de uma instituição de pesquisa, devem, em algum momento se convergir, tornando-se uma via única no caminho da comunicação científica. Leite (2009, p. 30) corrobora essa visão quando afirma que “é mais seguro iniciar o projeto com uma política de conteúdos mais restritiva e, gradativamente, flexibilizá-la até o ponto que for conveniente à instituição [...]”.

Os procedimentos metodológicos são um detalhamento do que Bunge referiu como “dividir cada sistema em sua composição, ambiente e estrutura” (BUNGE, 1997, p. 458). A visão sistêmica concebida pela modelagem CESM (nessa pesquisa ainda sem incluir a modelagem de mecanismos), permitiu uma representação do RI que considera os atores e fatores relevantes para o RI, bem como as suas ligações.

Considera-se como limitações da pesquisa: (1) o modelo não pode ser melhor que o consenso dos peritos envolvidos na modelagem, bem como é afetado pela escolha de fontes para a modelagem, que também depende da competência dos modeladores; (2) a modelagem requer um(a) gerente de projeto, responsável por tomar decisões quando não há consenso. No que tange a falseabilidade da pesquisa, o fundamento teórico-metodológico nos dá elementos para criticar o modelo: (1) as definições de componente atômico, item do ambiente e ligação servem para verificar as concepções; (2) as possíveis origens dos fatos sociais permitem questionar a escolha dos itens ambientais concebidos e sugerem onde procurar eventuais itens ambientais faltantes no modelo.

Espera-se que essa contribuição estimule a elaboração de novas pesquisas sistêmicas em sistemas de informação, em geral, e em RI, em particular. Essas pesquisas devem tratar da conjectura e teste dos mecanismos que operam a emergência, crescimento, manutenção, decadência e morte dos RIs.

## REFERÊNCIAS

ALVES, L. Informação e os sistemas de comunicação científica na Ciência da Informação. **DataGramaZero**: revista de informação, v. 12, n. 3, jun. 2011. Disponível em:

<<http://www.brapci.ufpr.br/documento.php?=&0000010229&dd1=890f3>>. Acesso em: 12 set. 2011.

ARLITSCH, K.; O'BRIEN, P. S. Invisible institutional repositories: addressing the low indexing ratios of IRs in Google. **Library hi tech**, v. 30, n. 1, p. 60-81, 2012.

Assisi, F.C. Anurag Acharya helped Google's scholarly leap. **INDOLink: Science**

& Technology. 2005. Disponível em:

<<http://www.indolink.com/SciTech/fr010305-075445.php>>. Acesso em 12 jan.2014.

BAPTISTA, A. A. et al. Comunicação científica: o papel da *open archives iniciativa* no contexto do acesso livre. **Enc. bibli: r. eletr. de bibliotecon. ci. inf.**, Florianópolis, n. esp., 1º sem. 2007. Disponível em: <<http://www.periodicos.ufsc.br/index.php/eb/article/view/1518-.2007v12nesp1p1/435>>. Acesso em: 1 set. 2011.

BARBALHO, Célia Regina Simonetti. Periódicos científico: parâmetros para avaliação de qualidade. In: FERREIRA, Sueli Mara Soares Pinto; TARGINO, Maria das Graças (Org). **Preparação de revistas científicas teoria e prática**. São Paulo: Reichmann& Autores, 2005. p. 123-158.

BAUMAN, Z. **Modernidade líquida**. Rio de Janeiro: J. Zahar, 2001.

BUDAPEST OPEN ACCESS INITIATIVE. 2001. Disponível em:

<<http://www.budapestopenaccessinitiative.org/boai-10-recommendations>>. Acesso em 20 ago. 2013.

BUNGE, M. A systemic perspective on crime. In: WIKSTRÖM, Per-Olof H.; SAMPSON, R. J. **The explanation of crime: context**,

mechanisms, and development. Cambridge University Press, 2006, p. 8-30.

BUNGE, Mario. **A world of systems**. Dordrecht, NL: D. Reidel, 1979.

BUNGE, M. **Emergence and convergence**: qualitative novelty and the unity of knowledge. Toronto: University of Toronto, 2003.

BUNGE, M. Mechanism and explanation. **Philosophy of the social sciences**, v. 27, n. 4, p. 410-465, 1997.

BUNGE, Mario. Systemism: the alternative to individualism and holism. **Journal of Socio-Economics**, v. 29, n. 2, p. 147-157, 2000.

CARVALHO, M. C. R. de. Bibliotecas universitárias brasileiras e a implantação de Repositórios Institucionais. **RevIU**, v. 1, n. 0, jul./dez., 2009. Disponível em:  
<<http://www.siglinux.nce.ufrj.br/~gtbib/site/2009/06/implantacao-de-repositorios/>>. Acesso em: 07 set. 2011.

CASTELLS, M. **A sociedade em rede**. 7. ed. São Paulo: Paz e Terra, 2003.

CERVO, A. L.; BERVIAN, P. A.; SILVA, R. da. **Metodologia científica**. 6. ed. São Paulo: Pearson Prentice Hall, 2007.

COSTA, S. M. S. Abordagens, estratégias e ferramentas para o acesso aberto via periódicos e Repositórios Institucionais em instituições acadêmicas Brasileiras. **Liinc em Revista**, Rio de Janeiro, v. 4, n. 2, p. 218-232, set. 2008. Disponível em:  
<<http://www.brapci.ufpr.br/.php?dd0=0000005080&dd1=11726>>.  
Acesso em: 10 set. 2011.

COSTA, S. M. S. Filosofia aberta, modelos de negócios e agências de fomento: elementos essenciais a uma discussão sobre o acesso aberto à informação científica. **Ci. inf.**, Brasília, v. 35, n. 2, ago. 2006  
. Disponível em: <<http://www.scielo.br/sciel>

o.php?script=sci\_arttext&pid=S0100-19652006000200005&lng=en&nrm=iso>. Acesso em: 28 maio 2010.

CRESPO, I. M.; RODRIGUES, A. V. F. Normas técnicas e comunicação científica: enfoque no meio acadêmico. **Rev. dig. bibli.ci. inf.**, Campinas, v. 9, n. 1, p. 36-55, jul./dez. 2011. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000010844&dd1=63e04>>. Acesso em: 12 set. 2011.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002. 171 p.

GUEDON, J. C. It's a repository, it's a depository, it's an archive...: open access, digital collections and value. **ARBOR ciencia, pensamiento y cultura**, v. 185, n. 737, p. 581-595, maio/jun. 2009. Disponível em: [arbor.revistas.csic.es/index.php/arbor/ar=ticle/download/315/316](http://arbor.revistas.csic.es/index.php/arbor/ar=ticle/download/315/316)>. Acesso em: 20 jun. 2013.

GUEDON, J. C. **Oldenburg's long shadow**: librarians, research scientists, publishers, and the control of scientific publishing, association of research libraries. 2001. Disponível em: <<http://www.arl.org/resources/pubs/138guedon.shtml>>. Acesso em: 30 jul. 2011.

GUIMARÃES, M. C. S.; SILVA, C. H.; NORONHA, I. H. RI é a resposta, mas qual é a pergunta? Primeiras anotações para a implementação de repositório institucional. In: SAYÃO, L. F. et al. (Org.) **Implantação e gestão de repositórios institucionais**: políticas, memória, livre acesso e preservação. Salvador: EDUFBA, 2009. p. 261-281.

HARNAD, S. Entrevista com Stevan Harnad. **Enc. bibli.:** r. eletr. bibliotecon.ci. inf., Florianópolis, n. esp., 1º sem. 2007. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2007v12nesp1pvi/518>>. Acesso em: 7 ago. 2011.

HESS, C.; OSTROM, E. A Framework for Analyzing the Knowledge Commons. In: \_\_\_\_\_. **Understanding Knowledge as a Commons**:

from theory to practice. Cambridge: MIT Press, 2007. Disponível em: <<http://surface.syr.edu/cgi/viewcontent.cgi?article=1020&context=sul>>. Acesso em: 7 ago. 2011. (verificar Link)

KERN, V. M. O sistemismo de Bunge: fundamentos, abordagem metodológica e aplicação a sistemas de informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 12., 2011, Brasília. **Anais...** Brasília: Thesaurus, 2011. p. 2693-2709.

KURAMOTO, H. Entrevista concedida via Skype em 11 dezembro de 2013. Entrevistadores: L. Silva e V. Kern. Florianópolis, 2013. .  
Entrevista concedida para pesquisa de mestrado de Luciana Mara Silva, PGCIN- UFSC.

KURAMOTO, H. Informação científica: proposta de um novo modelo para o Brasil. **Ci. Inf.**, Brasília, v. 35, n. 2, maio/ago. 2006. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/831/1742>>. Acesso em: 2 jan. 2014.

KURAMOTO, H. Acesso livre + portal de periódicos da Capes = maior acesso à informação científica. **J. cienc.**, Rio de Janeiro, 28 ago. 2010. Notícias. Não paginado. Disponível em: <<http://www.jornaldaciencia.org.br/Detalhe.jsp?id=55239>>. Acesso em: 2 set. 2010.

LAASER, W.; RODRIGUES, R. S.; FACHIN, G. R. B. Educação a distância e recursos abertos. **Rev. iberoam. educ.**, v. 49, n. 4, maio 2009. Disponível em: <<http://www.rioei.org/deloslectores/2879.pdf>>. Acesso em: 20 abr. 2010.

LE CODIC, Y. **A ciência da informação**. 2. ed. Brasília: Briquet de Lemos, 2004.

LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da informação científica brasileira**: repositórios institucionais de acesso aberto. Brasília: Ibict, 2009.

LEITE, F. C. L. **Gestão do conhecimento científico no contexto acadêmico**: proposta de um modelo conceitual. 2006. 159 f. Dissertação

(Mestrado em Ciência da informação)-Universidade de Brasília, 2006. Disponível em: <[http://repositorio.bce.unb.br/bitstream/10482/3975/1/2006\\_FernandoCesarlimaleite.pdf](http://repositorio.bce.unb.br/bitstream/10482/3975/1/2006_FernandoCesarlimaleite.pdf)>. Acesso em: 19 fev. 2010.

LEITE, F. C. L.; COSTA, S. M. S. Gestão do conhecimento científico: proposta de um modelo conceitual com base em processos de comunicação científica. **Ci. inf.**, Brasília, v. 36, n. 1, p.92-107, jan./abr. 2007. Disponível em: <<http://www.scielo.br/pdf/ci/v36n1/a07v36n1.pdf>>. Acesso em: 14 nov. 2013.

LYNCH, C. A. institutional repositories: essential infrastructure for scholarship in the digital age. **ARL Bimonthly Report**, p. 1-7, 2003. Disponível em: <<http://www.arl.org/newsltr/226/ir.html>>. Acesso em: 20 ago. 2008.

MARCONDES, C. H. et al. **Bibliotecas digitais: saberes e práticas**. Salvador: UFBA; Brasília: IBICT, 2005. Disponível em: <<http://livroaberto.ibict.br/bitstream/1/1013/1/Bibliotecas%20Digitais.pdf>>. Acesso em: 20 ago. 2013.

MARQUES, A. M. N.; MAIO, S. R. da S. **Repositórios institucionais**. Disponível em:<<http://repositoriosdigitais.web.simplesnet.pt/PDF%27S/Artigo%2020Repositorios%20Institucionais.pdf>>. Acesso em: 1 maio 2010.

MEADOWS, A J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.

MEADOWS, A. J. **Communication in Science**. London: Butterworths, 1974.

MINAYO, M.C.S. O desafio da pesquisa social. In: \_\_\_\_\_. **Pesquisa social: teoria, método e criatividade**. 26. ed. Petrópolis: Vozes, 2007.

MOREIRA, W. Os colégios virtuais e a nova configuração da comunicação científica. **Ci. inf.**, Brasília, v. 34, n. 1, p.57-63, jan./abr. 2005. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000003143&dd1=00ef9>>. Acesso em: 10 set. 2011.

MORENO, F. P.; LEITE, F. C. L.; MÁRDERO ARELLANO, M. A. Acesso livre a publicações e repositórios digitais em Ciência da Informação no Brasil. **Perspect. ciênc. inf.**, Belo Horizonte, v.11 n.1, p. 82-94, jan./abr. 2006. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000003780&dd1=7313d>>. Acesso em: 8 set. 2011.

MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ci. inf.**, Brasília, v. 35, n. 2, p. 27-38, maio/ago. 2006. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000003924&dd1=1f52d>>. Acesso em: 10 set. 2011.

MUSSO, P. A filosofia da rede. In: PARENTE, A. (Org.). **Tramas da rede: novas dimensões filosóficas, estéticas e políticas da comunicação**. Porto Alegre: Sulina, 2004. p.17-38.

OLIVEIRA, É. B. P. M. de; NORONHA, D. P. A comunicação científica e o meio digital. **Inf. & soc. est.**, João Pessoa, v. 15, n.1, p. 75-92, jan./jan. 2005. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000003114&dd1=48885>>. Acesso em: 12 set. 2011.

RIBEIRO, O. B.; VIDOTTI, S. A. B. G. Otimização do acesso à informação científica: discussão sobre a aplicação de elementos da arquitetura da informação em repositórios digitais. **Biblios**, Rio Grande, v. 23, n. 2, p. 105-116, 2009. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000008441&dd1=5667e>>. Acesso em: 2 set. 2011.

RODRIGUES, E. A. S. C. Entrevista concedida via Skype em 4 dezembro de 2013. Entrevistadores: L. Silva e V. Kern. Florianópolis, 2013. Entrevista concedida para pesquisa de mestrado de Luciana Mara Silva, PGCIN- UFSC.

ROSA, F. G.; GOMES, M. J. Coordenadores de comunidades de repositórios institucionais: o caso do repositório. **Enc. bibli.** r. eletr. de bibliotecon. ci. inf., Florianópolis, n. esp., 2º sem. 2010a. Disponível em: <<http://periodicos.ufsc.br/index.php/eb/article/download/16591/15765%E2%80%8E>>. Acesso em: 20 ago. 2013.

ROSA, F.; GOMES, M. J. Dos promotores aos utilizadores: estudos sobre o RepositórioUM. In: GOMES, M. J.; ROSA, F. (Org.). **Repositórios institucionais: democratizando o acesso ao conhecimento.** Salvador: EDUFBA, 2010b. p. 153-203. Disponível em: <[http://repositorium.sdum.uminho.pt/bitstream/1822/11417/1/RI\\_Maria%20Jo%C3%A3o%20Gomes%20%26%20FI%C3%A1via%20Rosa.pdf](http://repositorium.sdum.uminho.pt/bitstream/1822/11417/1/RI_Maria%20Jo%C3%A3o%20Gomes%20%26%20FI%C3%A1via%20Rosa.pdf)>. Acesso em: 2 set. 2013.

SILVA, L. M.; FORMOSO, R. G.; KERN, V. M. Repositório institucional como sistema técnico-social: composição, ambiente e estrutura. In: Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), XIV, 2013, Florianópolis-SC. **Anais...**, 2013. p. 1-7. Disponível em: <<http://www.enancib2013.ufsc.br/index.php/enancib2013/XIV/enancib/paper/view/387/295>>. Acesso em: 17/11/2013.

TARGINO, M. das G. Comunicação científica: uma revisão de seus elementos básicos. **Inf. & soc. est.**, João Pessoa, v. 10, n.2, p. 7-85, 2000. Disponível em: <<http://www.br.apci.ufpr.br/documento.php?dd0=0000001182&dd1=098ce>>. Acesso em: 14 nov. 2013.

VEIGA, V. S. de O.; MACHADO, R. R.; ALVES, A. da S. Avaliação de usabilidade em Repositórios Institucionais: o caso da Fiocruz. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 25., 2013, Florianópolis. **Anais eletrônicos...** Disponível em: <<http://portal.febab.org.br/anais/article/view/1326/1327>>. Acesso em: 20 ago. 2013.

WEITZEL, S. da R. Fluxo da Informação Científica. In: POBLACION, D. A.; WITTER, G. P.; SILVA, J. F. M. da. **Comunicação e produção**

**científica**: contexto, indicadores e avaliação. São Paulo: Angellara, 2006a. p. 83-114.

WEITZEL, S. da R. O papel dos Repositórios Institucionais e temáticos na estrutura da produção científica. **Em questão**, Porto Alegre, v. 12, n. 1, p. 51-71, jan./jun. 2006b.

## **APÊNDICE**



# APÊNDICE A - Poster Enacib



Luciana Mara Silva  
Bibliotecária da Uniar e mestra em PGR/UFSC,  
Robson Garcia Formoso  
Professor da Faculdade Senac e mestra do PPG/UEFSB,  
Vinicius Medina Kern  
Professor do PGR e do PPG/UEFSB.

XIV Encontro Nacional de Pesquisa em Ciência da Informação  
(ENACIB 2013)  
Informação e interação: Ampliando perspectivas para o  
desenvolvimento humano  
Florianópolis - SC, de 29 de outubro a 01 de novembro 2013

## REPOSITÓRIO INSTITUCIONAL COMO SISTEMA TÉCNICO-SOCIAL: COMPOSIÇÃO, AMBIENTE E ESTRUTURA

### INTRODUÇÃO

Repositórios institucionais (RI) sistemas de informação técnico-socials cuja frequente falha é atribuída a diversos motivos, incluindo a falta de recursos e de adesão por autores envolvidos.

**Barreiras e desafios aos RI:** ♦ desconhecimento da comunidade científica (MARQUES; MAIO, [2007]); WEITZEL, 2006, LEITE, 2009) ♦ divergências sobre como é e como deve funcionar o RI como serviço de informação (GUIMARÃES; SILVA; NORONHA, 2009, LEITE, 2009) ♦ complexidade das interações entre os autores envolvidos (GUIMARÃES; SILVA; NORONHA, 2009) ♦ invisibilidade dos RI para motores de busca (ARLITSCH; O'BRIEN, 2012).

Estudos como o de Guedon (2009) apontam que visitação e uso dos RI ultrapassa questões tecnológicas. Necessária visão ampla, colaboração de autores sociais envolvidos – por exemplo, via auto-arquivamento.

**Problema:** Como entender repositórios de forma abrangente, não-enviada?

**Objetivo:** elaborar uma descrição sistêmica de um repositório institucional, caracterizando seus componentes, itens do entorno e ligações estruturantes.

### FUNDAMENTOS E PROCEDIMENTOS METODOLÓGICOS

**Base teórico-metodológica:** sistemismo de Bunge (1997, 2003).

**Sistema:** objeto estruturado, i.e., composto por partes interligadas, com características sistêmicas (emergentes) que seus componentes não têm. Tudo é sistema: "Toda coisa, seja concreta ou abstrata, é um sistema ou um conjunto ou potencial componente de sistema" (BUNGE, 2003, p. 114).

**Sistema concreto** pode ser descrito segundo o modelo CESM: *Composition, Environment, Structure, Mechanism* (BUNGE, 2003). C, E, S; levantamento; M: conjectura e teste de hipótese.

**Abordagem metodológica** para descrição do sistema RI: abstração dos elementos C, E, S a partir de revisão da literatura orientada pelo modelo CESM.

**Seleção de fontes:** oportunística, buscando abranger as origens dos fatos sociais segundo Bunge: ambiental, biopsicológica, econômica, política e cultural.

**Procedimentos:** [1] marcação individual, por vários pesquisadores, o texto-fonte quanto aos componentes, itens do entorno e ligações estruturantes de um RI, [2] síntese de um modelo de RI parcial por cada

pesquisador, [3] sua consolidação por consenso, [4] síntese dos modelos parciais pelo consenso do grupo de pesquisadores. Nesta pesquisa, o grupo de pesquisadores é composto por seus 3 autores.

### RESULTADOS PRELIMINARES

A Figura 1 mostra uma marcação individual de componentes, itens do ambiente e ligações a partir do texto de Guimarães, Silva e Noronha (2009).

Figura 1: Extrato do texto-fonte marcado com elementos de um repositório institucional: componentes (vermelho), itens do ambiente (verde) e ligações (azul caso) [fonte: autores].

A Figura 2 mostra o resultado da consolidação do modelo parcial de RI segundo o texto de Guimarães, Silva e Noronha (2009), de acordo com o consenso do grupo de pesquisadores.

Figura 2: Elementos [e seus sinônimos e correlatos] do modelo parcial de sistema "repositório institucional" [fonte: autores].

A Figura 3 verte graficamente o quadro da Figura 2.

Figura 3: Representação gráfica do modelo parcial de sistema de RI apresentado na Figura 2 [fonte: autores].

Componentes: documento [artigo científico, artigo preprint, publicação acadêmica], pesquisador [comunidade acadêmica, local(s) acadêmico(s)], comunidade científica [online, atual] do RI [profissional da informação, profissional de TI]

Itens do ambiente: instituição [biblioteca, guia de avaliação, pesquisa, ambiente de trabalho], instituição de pesquisa, mecanismos de avaliação, micro-meso-políticas locais, política de informação do RI, política de infraestrutura da instituição, universidade, sistemas [normatização acadêmica, comunidade científica], periódicos [revistas científicas, comerciais], política das editoras [comerciais], normas, leis e cultura [caso], mandatório de depósito, copyright, movimento de acesso livre, padrão OAI, propriedade intelectual], órgãos regulamentadores [micro-meso-políticas globais, órgãos governamentais], estrutura Metodológica [hardware, infra-estrutura tecnológica, metadados, software], web [internet]

Ligações: auto-arquivamento [auto-depositos] [pesquisador-publicação], ligações de avaliação e promoção [revisão por pares] [pesquisador-agência de fomento], acesso [acesso aos conteúdos] [usuário-publicação]

Figura 2: Elementos [e seus sinônimos e correlatos] do modelo parcial de sistema "repositório institucional" [fonte: autores].

A Figura 3 verte graficamente o quadro da Figura 2.

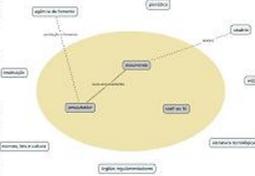


Figura 3: Representação gráfica do modelo parcial de sistema de RI apresentado na Figura 2 [fonte: autores].

### DISCUSSÃO

O modelo de RI proposto deve ser abrangente. O modelo parcial resultante ainda tem lacunas – há componente sem ligações com os demais e há itens do ambiente não ligados a componentes. As lacunas devem ser preenchidas a partir da análise dos textos complementares do corpus.

Quanto à relevância de uma descrição sistêmica, o próximo texto a ser analisado para avaliar elementos (ARLITSCH; O'BRIEN, 2012) deve permitir entender que uma decisão adequada desde um ponto de vista disciplinar, como a adoção do consagrado padrão de metadados Dublin Core, implica no sufocamento dos acessos ao repositório e, portanto, é uma decisão enviesada que não deve prevalecer.

### REFERÊNCIAS

- ARLITSCH, K.; O'BRIEN, P. S. Invisible institutional repositories: addressing the low indexing rates of Ina in Google. *Library Hi Tech*, v. 30, n. 1, p. 60-81, 2012.
- BUNGE, M. *Emergence and convergence: qualitative novelty and the unity of knowledge*. Toronto: University of Toronto, 2003. 310p.
- BUNGE, M. *Mechanism and explanation: Philosophy of the Social Sciences*, v. 27, n. 4, p. 410-465, 1997.
- GUIMARÃES, M. G. S.; SILVA, C. H.; NORONHA, L. H. RI e a resposta, mas qual a pergunta? Primeiros passos para a implementação de repositório institucional. In: SARAI, L. P. et al. (Org.) *Implementação e gestão de repositórios institucionais*. Política, memória, livre acesso e preservação. São Paulo: EDUEFA, 2009. p. 261-281.
- LEITE, F. C. L. *Como gerenciar e ampliar a visibilidade da informação científica brasileira*. Repatórios institucionais de acesso aberto. Brasília: Inet, 2009.
- MARQUES, Amélia Maria Nunes; MAIO, Silvia Raquel da Silva. *Repositórios institucionais*. [2007?]. Disponível em: <http://repositorios.ufrpe.br/web/impresst/2012/04/01/02/201208repositorio%20institucionais.pdf>. Acesso em: 1 maio 2010.
- WEITZEL, Geneva de Rittsch. O papel dos repositórios institucionais e temáticas na estrutura da produção científica. In: *Questão*. Porto Alegre, v. 12, n. 1, p. 51-71, jan./jun. 2006.

### ABSTRACT

Abstract: Institutional repositories are technical-social information systems whose failure is often attributed to several reasons, including the lack of resources and adherence by stakeholders. How to understand repositories is a comprehensive, unbiased manner? To answer this question, we describe institutional repositories as systems, characterizing their components, environmental items, and structural bonds. Our approach is the abstraction of those elements from a literature review guided by Bunge's compositional environment-structure-mechanism system model. The selection of sources is opportunistic, seeking to cover the origins of social facts according to Bunge: environmental, biopsychological, economic, political, and cultural. The procedure includes micro-meso-political policies, local information policy of the source itself, then addressing those markings to group in general a consensus mapping and its consolidation in a system model. Here we report on a pilot study of the abstraction process using a single source text. The resulting institutional repository model sought to be comprehensive enough to allow for enlightening repository management decisions – for instance, to give up on the adoption of the renowned Dublin Core metadata standard because it hampers the retrieval of repository documents, being therefore detrimental to scientific communication.

**Keywords:** Institutional repositories. Open access. Systemism. CESM system model. Information systems.

Patrocinadores:



Apoio:





## **ANEXOS**



## ANEXO A – Marcação e coleta do Texto 1

# RI é a resposta, mas qual é a pergunta?

Primeiras anotações para a implementação de Repositório Institucional

*Maria Cristina Soares Guimarães*

*Cicera Henrique da Silva*

*Ilima Horsth Noronha*

### LEGENDA

**Componentes em fundo vermelho** – coleção de elementos componentes  
**Itens ambientais em verde** – coleção de itens que não fazem parte do sistema mas atuam ou sofrem ação por algum componente  
**Ligações em azul claro** – coleção de ligações entre componentes e entre esses e itens do ambiente

### Introdução

No começo dos anos 1980, um artigo publicado por um influente pesquisador britânico da área de comunicação no *Journal of Information Science* (HALLORAN, 1983) teve como título a seguinte provocação: “Information may be the answer, but what is the question?”. Testemunha de tempos quando se forjou a expectativa que as tecnologias de informação e comunicação (TICs) seriam um instrumento poderoso e uma estratégia irresistível a favor da democracia e da ampliação da participação social nos processos de decisão do Estado, o autor se perguntava se não havia ali um otimismo excessivo. Ou, disponibilidade de informação é o que realmente importa? Ainda, disponibilidade de informação é condição necessária à democracia, mas seria suficiente? Os questionamentos do autor ecoam, ainda, aquelas perguntas clássicas, tão caras aos profissionais de informação, especialmente àqueles dedicados ao desenvolvimento de Sistemas de Recuperação de Informação (SRI): quem necessita de informação? Que tipo de informação diferentes grupos de usuários, ou toda a população, necessita? Quem decide o que é necessário para quem? Quem seleciona, organiza e provê acesso à informação? A que custo? Que uso é esperado ser feito da informação?

ri é a resposta, mas qual é a pergunta? primeiras anotações para a implementação... 1

Passadas quase três décadas, período esse de avanços extraordinários no âmbito das TICs, permanece (ou mesmo fortalece) a aposta que a mais ampla circulação de informação tem um papel fundamental no bem estar social das nações, especialmente dado o caráter central e produtivo do conhecimento científico e tecnológico. Acesso à informação tornou-se uma força propulsora do processo de inovação. A **Internet** e a descentralização do processo de produção e disseminação do conhecimento, junto a uma crescente prática de colaboração em rede são forças modeladoras da atual reconfiguração da prática da ciência, e de seus processos de comunicação e publicação. Nasce aqui uma oportunidade única para uma resposta construtiva às disfunções do mercado que operam no sistema de comunicação científica – o conhecimento como bem público é uma utopia possível (GUIMARÃES, 2009).

O uso intensivo das mídias eletrônicas tornam virtualmente impossível a limitação do acesso à informação, sob pena do uso de soluções jurídicas antipáticas e draconianas, que só fazem aprofundar o gap entre os *haves* e os *have not*. A resistência veio pelo **Movimento do Acesso Livre** à informação científica: “O propósito da informação é para ser compartilhada, assim como o propósito do pão é para ser comido”, (OPEN SOCIETY INSTITUTE, 2005, tradução nossa) .

Repositório Institucional (RI), sumariamente entendido como um locus online para coleta, preservação e disseminação da produção intelectual de uma instituição (de pesquisa), emerge, assim, como uma peça fundamental no quebra-cabeça da iniquidade no **acesso** à informação científica. O que começou no campo disciplinar, no seio da **comunidade científica**, como uma estratégia para intercâmbio de preprints, nos anos noventa do século passado, chega ao século XXI com ares de inevitabilidade: “[...] É improvável que qualquer instituição séria de pesquisa não vá ter seu próprio repositório digital até ao final desta década.”<sup>2</sup> (SWAN; CARR, 2008, p. 31).

Reverberando a provocação de Halloran (1983), e guardadas as devidas proporções, as discussões em torno do conceito, desenvolvimento

1 “The whole purpose of information is to be shared, as the purpose of bread is to be eaten.”

2 “It is unlike that any serious research-based institution will not have its own digital repository by the end of this decade.”

e implantação de RIs parecem seguir a mesma dinâmica – dispor de um RI é o que importa! Especialmente na última década, e particularmente no esteio do conjunto de declarações e manifestos, nacional e internacionais, a favor do Acesso Livre (KURAMOTO, 2006), os RIs emergem como resposta para um amplo conjunto de iniquidades no acesso à informação, nas suas mais diversas manifestações. Quer seja como resposta à escalada nos custos das assinaturas de periódicos científicos; como um novo “serviço de informação”; como um elemento adicional no sistema de comunicação científica; como uma estratégia para conferir visibilidade institucional, ou mesmo para a preservação digital da memória institucional, os RIs podem, potencialmente, ser a resposta! A despeito de tamanha onda de otimismo, o que a literatura na área aponta é que as experiências em curso testemunham que muito ainda há que ser investigado e colocado em prática para que todo o potencial dos RIs possa ser alcançado (SHEEREVES; CRAGIN, 2008).

Nesse sentido, o que talvez seja ainda pouco explorado, ainda que não esquecido, nas investigações em curso é que muito antes de ser mais um dispositivo ou **infraestrutura tecnológica** a favor da ampliação da circulação de informação científica, um RI é um complexo sociotécnico, onde estão envolvidos um conjunto de atores com visões, práticas e perspectivas diversas, cujas relações são tecidas à luz de **micro e macropolíticas, locais e globais**. Ainda que a importância dos aspectos técnicos seja inegável para sua concepção, desenvolvimento e implantação, eles não asseguram, ou sustentam, a apropriação e uso do RI pelos usuários reais e potenciais. Como insistentemente lembrado pelas análises empreendidas na área dos estudos sociais das tecnologias, mais particularmente, da construção social das tecnologias (PINCH; BIJKER, 1984): é somente pelo uso, e pela perspectiva do usuário, que uma tecnologia é absorvida no social e reconhecida em sua qualidade e seus méritos.

O texto que se segue objetiva ser uma breve reflexão sobre alguns pontos norteadores para o desenvolvimento e implantação de um RI. No limite entre a inexorabilidade de implantá-lo e a desqualificação de seu potencial no esforço de ampliação da circulação da informação científica, repousa o reconhecimento que um RI é muito mais que

uma contribuição ímpar das TICs em favor do Movimento do Livre Acesso. O "I", de institucional, que adjetiva o repositório (esse, sim, um objeto digital) coloca em foco uma instituição que é um complexo de relações sociais e de práticas singulares, embebidas na cultura e nas contingências políticas e de ordens outras, resultado de sua própria história, e cujas metas e objetivos mais amplos orientam um futuro almejado. Some-se a essa instituição um segundo adjetivo, "de pesquisa", onde domínios de conhecimento, tradições de pesquisa, culturas epistêmicas, práticas de comunicação e padrões de busca (*information seeking*) e uso de informação implicam (e estão implicadas) em relações idiossincráticas com o desenvolvimento, estabilização e uso de aparatos tecnológicos, e com a própria dinâmica da ciência.

Nesse sentido, qualquer visão simplificadora do relacionamento entre tecnologias, espaços de práticas constituídos e mudanças sociais é, no mínimo, ingenuidade. Perspectivas de análise construtivista e sociotécnica, especialmente no âmbito da "construção social das tecnologias" (PINCH; BIJKER, 1984) e da "informática social" (KLING, 2000; 2007), ressaltam a importância de um olhar cuidadoso por sobre os contextos institucional e cultural quando do desenvolvimento e implantação de um RI. Entram em campo, portanto, considerações políticas e todo um repertório de estratégias que devem ser levados em consideração *vis a vis* o problema a ser enfrentado.

Longe de ser um *tour de force* sobre o tema (esforço que, por certo, precisa ser empreendido), a discussão que se segue visa, antes, uma leitura transversal sobre alguns pontos discutidos na literatura sobre a natureza e função dos RIs, bem como suas estratégias de desenvolvimento e implantação, e como a perspectiva sociotécnica pode enriquecer o aprendizado esperado. Aqui, convivem esperanças e incertezas, como as duas faces de Juno: por um lado, lê-se o esforço e comprometimento genuíno de muitos (órgãos governamentais, organizações de classe, instituições de pesquisa, e pesquisadores individuais) para a promoção e implantação de RIs. Por outro lado, permanece a dúvida de como realizar a potência do acesso universal quando práticas estabelecidas, culturas, e interesses econômicos herdados de séculos ainda perduram no cerne da dinâmica científica.

Muito antes de procurar por respostas, o texto procura colocar algumas perguntas que, quiçá, possam estimular análises mais detalhadas sobre o tema, especialmente no contexto brasileiro.

### **a lente da construção social das tecnologias**

O campo de investigação "construção social das tecnologias" (Social Construction of Technology – SCOT) (PINCH; BIJKER, 1984) faz parte de um conjunto de abordagens teórico-metodológicas que se aglutinam sob o guarda-chuva do movimento de Social Shaping of Technology – SST (WILLIAM; EDGE, 1996), que advoga que uma tecnologia não é nem puramente tecnológica tampouco emana diretamente de uma estrutura ou demanda social. Ao contrário, ela é concebida, desenvolvida e alcança sua estabilização tecida em uma rede complexa e mestiça de atores, interesses, estratégias e atividades. O ponto central é entender que escolhas são feitas, e que qualquer perspectiva de imperativos ou determinismos (tecnológico, científico, político, econômico ou social) é impotente para descrever como as tecnologias chegaram ao que são em sua forma final. Trata-se, afinal, não de uma tecnologia, mas um dispositivo sociotécnico (GUIMARÃES, 1997).

De comum, essas perspectivas compartilham que:

- O processo de desenvolvimento das tecnologias é sempre contingente;
- As tecnologias nascem sob o conflito, a diferença e a resistência;
- As tecnologias estão implicadas nas estratégias dos atores, e só quando a concordância é possível é que as tecnologias ganham forma e sentido social mais amplo;
- Tanto as estratégias quanto suas consequências são um fenômeno emergente no sistema de relações entre os atores. Ou seja, elas nascem no curso da ação e, recursivamente, se influenciam.

Ao longo dos anos, a perspectiva teórica SCOT mostrou-se bastante produtiva para analisar o desenvolvimento e apropriação das TICs em

diferentes contextos sociais (WILLIAMS, 1997; WILSON; HOWCROFT, 2002), e foi tomada como base para o desenvolvimento do campo de investigação da "informática social", capitaneada por Kling (2000, 2007).

De forma sumária, e para o interesse da presente discussão, a abordagem SCOT defende que o processo de inovação que leva à apropriação e ao uso das tecnologias é complexo, coletivo, e submetido a intensas negociações de sentido, moldadas e informadas pelo contexto no qual os atores se posicionam. Ao longo do processo, mudam tanto o contexto e as relações entre os atores, quanto os artefatos tecnológicos. Quatro conceitos básicos orientam as análises na abordagem SCOT:

- Grupos sociais relevantes, ou, os diferentes atores (*stakeholders*) que possuem lugar de destaque no processo, e para ele trazem diferentes recursos (conhecimento, poder, financiamento, dentre outros), diferentes visões e expectativas sob a forma e sentido futuro dos artefatos;
- Flexibilidade interpretativa, ou, em função da participação de diferentes grupos sociais relevantes, o processo está aberto a várias interpretações de sentido e de uso;
- Quadro tecnológico, ou, um conjunto de conceitos, práticas e técnicas compartilhadas e utilizadas pelos grupos que informam os interesses comuns dos grupos;
- Fechamento ou estabilização, ou, quando o processo de negociação é finalizado pelo consenso e emerge dali um "modelo dominante".

Sob a ótica SCOT, um RI deve primeiramente ser situado em um contexto de mudança e aberto à inovação, como é o caso do setor de publicação acadêmica face às TICs e ao Movimento do Livre Acesso (BORGMAN, 2007; KING et al., 2006; BOHLIN, 2004). Essa configuração delinea o quadro tecnológico. Aqui, são potenciais grupos sociais relevantes, por exemplo, os pesquisadores, as instituições de pesquisa, os editores científicos (comerciais e universitários), as bibliotecas e os profissionais de informação, as agências de fomento os usuários, o Legislativo e outras organizações governamentais ou não. Ressalte-se que cada um deles traz para a discussão dos RIs

um conjunto de "recursos" e interesses, que também podem estar longe de representar um consenso interno no grupo. O que cada um expressa como uma macropolítica pode ser, de fato, um conjunto de micropolíticas, por vezes conflitantes. O que, teoricamente, os une em um "quadro tecnológico" é o interesse na ampliação do acesso à informação, ainda que para alguns isso signifique geração de riqueza e para outros continuidade ou atualidade de uma linha de pesquisa. Como condicionante, o fato do acesso não poder ser alcançado em detrimento da qualidade do conteúdo e da garantia da maior publicidade. Flexibilidade interpretativa é o que a literatura da área melhor retrata no momento, com várias perspectivas conceituais e modelos de desenvolvimento; e o fechamento e a estabilização dos RIs (p.e. um "modelo de negócio" ou um *design* dominante) é algo que aqueles que lutam bravamente para ganhar acesso à informação esperam que se realize em futuro próximo, e a seu favor.

Essa rápida apresentação é oportuna para que se tenha em mente que nas estratégias para desenvolvimento e implantação dos RIs é fundamental identificar os grupos sociais relevantes e os recursos e interesses que eles trazem para o processo; como eles se relacionam entre si, e com que agendas, implícitas e explícitas. Os alinhamentos de interesses podem unir um maior ou menor número de grupos sociais, em contraposição a outros, e antecipadamente orientar fortemente o curso da ação. Aqui entram também senão atores específicos (em geral, pesquisadores individuais com participação ativa) e políticas infraestruturais já estabelecidas, como aquelas relacionadas a *copyright* e propriedade intelectual.

Mais importante, e especialmente para os países em desenvolvimento, é fundamental reconhecer que as estratégias e recursos desses grupos sociais são contexto-dependentes, e que quaisquer experiências e *best practices* de outros países, especialmente os desenvolvidos, informam, mas não podem ser tomadas como "tamanho único". Por exemplo, ainda que a comunidade de físicos de alta energia possam guardar, e de fato o faz, uma série de similaridades em todos os países do mundo no que diz respeito a sua prática de produção, acesso e uso de informação, eletrônica ou não, elas expressam também particula-

ridades locais, especialmente aquelas do Sul, e mais particularmente quando tentam se alinhar com as do Norte. Ou seja, alguns campos disciplinares são mais ou menos “internacionais”, mas todos refletem características locais. O mesmo ocorre para o caso das editoras científicas, comerciais ou não (e mesmo se elas existirem no Sul). No que diz aos interesses e estratégias das universidades/instituições de pesquisa, o quadro ainda é mais complexo, não só no que diz respeito à cultura local mas também no que diz respeito à forma como sua produção intelectual, e seus grupos de pesquisa, são financiados e avaliados.

Esse último ponto talvez seja um dos mais importante para as análises sobre RIs empreendidas na perspectiva do Sul, e mais especificamente no caso brasileiro: o sistema de avaliação acadêmica praticado no país. Se, como no caso do Brasil, as macropolíticas que orientam o sistema de **recompensa dos pesquisadores** e premia as instituições de pesquisa insistem em um modelo de publicação acadêmica exógeno (ou internacional), é pouco provável que o papel, a motivação e o alcance do potencial dos RIs possam se realizar da mesma forma como no Norte.

No que se segue, a lente do Programa SCOT é usada para empreender uma rápida leitura de uma parte da literatura já disponível sobre RIs.

## natureza e função dos ris

São duas as principais razões que justificam o desenvolvimento e implantação dos RIs: por um lado, para tencionar (se não competir com) o modelo tradicional de publicação na ciência, e reduzir o poder de monopólio dos periódicos por outro lado, para complementar (se não re-fundar) o sistema de comunicação acadêmica. No primeiro caso, o conceito de RI está intrinsecamente vinculado ao Movimento do Acesso Livre, e seu conteúdo, 100% de artigos que passaram pelo crivo dos pares, **preprints** e/ou **postprints**. No segundo caso, o RI apresenta-se mais como estratégia de gestão de conteúdos intelectuais de uma **instituição de pesquisa**; consequentemente, seu conteúdo é bastante diversificado, e sem a necessária vinculação com os pares. O que os

une, teoricamente, é uma infraestrutura tecnológica com padrão OAI, que permite o autoarquivamento e que possibilita a preservação digital. O acesso livre é uma premissa, não um atributo: alguns RIs operam com perfis de acesso diferenciados, e em alguns casos restringem o acesso ao texto completo, provavelmente por questões de *copyright*.

Como consequência, e também pela sua recenticidade, o conceito ainda é reconhecido como escorregadio. Bailey Jr. (2008), em um recente artigo de revisão, aponta para a polissemia do termo. Para alguns, um RI é um conjunto de serviços que uma universidade/instituição oferece para os membros da sua comunidade com vistas ao gerenciamento e disseminação de seu conteúdo intelectual digital (LYNCH, 2003); para outros, uma base de dados web de material acadêmico institucionalmente definido (WARE, 2004), ou ainda, como sendo constituído de uma coleção de itens capturados para preservar o capital intelectual de uma ou várias instituições (CROWN, 2002).

Guédon (2009) faz uma leitura muito instrutiva dessa polissemia. Aqueles que defendem um RI como um “serviço de informação”, e colocam a ênfase no gerenciamento, organização e acesso, ecoam a partir de um contexto, de uma cultura e de uma prática de profissionais de informação, e buscam um caminho e um reposicionamento para as bibliotecas e seus profissionais em um ambiente crescentemente digital. A perspectiva de uma suposta neutralidade ao definir um RI como uma base de dados na web, com ênfase na interoperabilidade, vem dos esforços daqueles que operam nos limites entre as editoras científicas comerciais e as universidades, um “modelo de negócio” com séculos de sucesso. Por fim, aqueles que advogam que os RIs são uma estratégia para a reforma do sistema de publicação acadêmica estão profundamente envolvidos no Movimento do Livre Acesso, e mais particularmente, nas estratégias políticas para o acesso à produção científica oriunda das pesquisas em saúde financiadas pelo Estado, especialmente no contexto norte-americano.

Há ainda uma corrente que, sem necessariamente aderir a uma ou outra definição de RI, adotam uma perspectiva essencialmente funcional, e situam o RI como uma estratégia valiosa para dar visibilidade à produção acadêmica institucional, particularmente a partir

de um único ponto de acesso, o que contribuiria para aumentar o uso e impacto das pesquisas ali desenvolvidas (SWAN; CARR, 2008). Essa visão, também mais alinhada com o Movimento do Livre Acesso, mas não restritiva aos artigos revisados pelos pares, equaciona aumento de visibilidade com livre acesso, e aponta para o caráter mandatório de auto-arquivamento que emana das instituições. Aqui, o RI estaria também a serviço de uma estratégia de gestão institucional da pesquisa, como um **guia para avaliação** e, principalmente, para estimular e facilitar o re-uso dos conteúdos, acelerando o ciclo da pesquisa.

Uma visão representativa da própria comunidade científica é aquela defendida por Hamard (1990, 2006), um defensor incansável da extensão do sucesso do arXiv, de repositório temático de *preprints*, para repositório institucional. O foco aqui é no impacto e no uso da pesquisa, argumentando que muito se perde desse potencial dado o tempo e os custos da produção e distribuição. Ou seja, o depósito da versão final de um **artigo científico** em um RI (artigo esse já revisado por pares, e atendendo as questões de copyright e as **políticas dos editores comerciais**) é a forma mais rápida e efetiva de assegurar que um maior número de **pesquisadores interessados** possam lê-lo e citá-lo. Isso beneficiaria tanto o pesquisador individual como sua instituição, além de acelerar o ciclo da pesquisa. Essa visão é estritamente aderente ao Movimento do Livre Acesso e ao caráter mandatório do auto-arquivamento, e está em franca oposição àquelas estratégias de desenvolvimento de RIs que acolhem a diversidade de tipologia de conteúdos intelectuais produzidos pela instituição, o que, argumentam, seria muito mais a função de uma digital.

Independente das diferentes visões quanto à função a ser desempenhada pelo RI, mesmos os entusiastas do primeiro minuto reconhecem que, até o presente, esses não vêm sendo usados em todo o seu potencial, e que as instituições têm falhado em suas estratégias de desenvolvimento, implantação e sustentabilidade. Para aqueles que veem o RI como espaço para artigos científicos revisados por pares, o grande desafio é, sem dúvida, a questão do *copyright* e a tradição de um "modelo de negócio" de publicação científica que vem se mostrando exitoso ao longo dos últimos séculos. Nessa configuração, as editoras

científicas representam o grupo social relevante com mais recursos e têm, com cerca tranquilidade, mantido seu papel de destaque, especialmente nos países em desenvolvimento. Caberia perguntar, especificamente no caso brasileiro, que outras estratégias e alternativas outros grupos sociais relevantes, como universidades/ institutos de pesquisa e órgãos governamentais, poderiam trazer para tensionar o modelo vigente.

Quando pensado como um serviço de informação, o êxito de um RI deve ser lido na perspectiva interna das instituições, ou, no âmbito das micropolíticas e das estratégias de gestão dos conteúdos digitais (LYNCH, 2003). No âmbito das micropolíticas, a palavra-chave é sustentabilidade, por exemplo: de investimentos financeiros e do comprometimento institucional, e estímulo à adesão e permanência dos principais atores, de gestão de conteúdo adequado. Aqui, de forma mais clara, estão envolvidas questões relacionadas ao envolvimento dos profissionais de informação e das bibliotecas, um grupo social extremamente relevante no que diz respeito à estruturação e gestão de conteúdo. A garantia do acesso é condição *sine qua non*, e questões de usabilidade e facilidade de interação são variáveis importantes. Para todos os casos, uma estruturação de conteúdos pobre e confusa e qualquer comprometimento no acesso aos conteúdos enfraquecem uma já tênue relação de confiança com o usuário, inerente ao uso de um "novo" serviço que, ainda na infância, precisa mostrar ao que veio.

Desenhado esse breve enquadramento de relações entre os grupos de interesse, o que se pode aprender com algumas experiências prévias?

## algumas lições do caminho já trilhado

Williams e Lawton (2005) apontam que, no que diz respeito ao desenvolvimento e implantação de RIs, há extensa literatura sobre seus aspectos técnicos e operacionais, tanto no que diz respeito ao **hardware** e **softwares** disponíveis, como nas questões sobre arquitetura do sistema, interoperabilidade, **metadados** e preservação digital. Nessa perspecti-

va, as experiências relatadas situam um RI essencialmente como um serviço de informação e, exceto pela forma diferenciada de captura de material, guarda grande semelhança com a literatura voltada ao desenvolvimento de sistemas de recuperação de informação.

No geral, a literatura retrata que são várias as demandas para desenvolvimento de RI, e que as mesmas continuam em evolução, e que não há nenhum roteiro/protocolo padrão indicando a melhor maneira de progredir. Sheereves e Cragin (2008) reconhecem que, vencida quase uma década após os primeiros RIs serem implantados no Reino Unido e nos EUA, a motivação que os guiou ainda não é clara. Antes, para, e, em cada país, uma demanda diferente, uma configuração de forças, recursos e interesses entre os grupos sociais relevantes específica, e um curso de ação particular.

Na perspectiva micro das instituições de pesquisa, ou, no espaço onde o RI deixa de ser conceito para ser operativo, de comum acordo o fato de ser um trabalho extremamente técnico, exigindo a concepção e implementação de infraestrutura tecnológica e de informação sólidas, mas também uma tarefa que exige um aporte gerencial de excelência, com planejamento contínuo, priorizando e coordenando as ações com as expectativas dos diversos grupos sociais relevantes.

No que diz respeito ao grau de sucesso dos repositórios institucionais, as evidências apontam ainda que é algo irregular. Em 2005, Lynch e Lippincott, (2005) promoveram um inquérito com universidades e faculdades de artes liberais nos Estados Unidos. Ali, 40% das universidades já contavam com RIs operacionais. DSpace foi o pacote de gerenciamento de conteúdo predominante listado pelos entrevistados. O tamanho descrito destes RIs variou entre centenas de milhares de objetos digitais (acima de 10 terabytes de espaço) para menos de uma dúzia (inferior a um gigabyte), embora houvesse confusão entre os entrevistados sobre o que seria um "objeto digital" – um banco de dados, ou um registro de informação.

A tipologia dos materiais armazenados era diversificada, incluindo e-prints, teses e dissertações eletrônicas, coleções especiais digitalizadas, materiais multimídia, materiais de cursos, dentre outros. De forma clara, essa variedade de tipologias está diretamente ligada à motiva-

ção que guiou a implantação do RI, o que, por outro lado, cria seus próprios problemas. Em sua pesquisa, Macdowell (2007) registra que somente cerca de 13% dos conteúdos dos RIs passaram por revisão pelos pares. Argumenta-se que essa diversidade, ou inconsistência de coleções, pouco contribui para quesitos de usabilidade e confiança do usuário, além de serem iniciativas que se distanciam do apelo central do Movimento de Livre Acesso (POYNDR, 2006).

A participação do corpo docente da instituição foi, para todos os entrevistados, apontada como voluntária, ou seja, sem caráter mandatório. No geral, é uma participação considerada muito baixa.

Sem surpresas, Lynch e Lippincott, (2005) apontam que alcançar uma taxa significativa de participação (o que quer que seja isso) é o principal indicador de sucesso dos RIs citado na literatura. Alguns gestores de RIs tentam recrutar a contribuição dos profissionais enfatizando a importância da preservação digital, e mesmo como aferição do uso dos resultados das pesquisas (DAY, 2004). Mas os fatores que motivam os acadêmicos são, de fato, mais complexos. Análises recentes (SWAN et al., 2005; FOSTER; GIBBONS, 2005; KENNAN; WILSON, 2006) confirmam as suspeitas de que o comprometimento com a visibilidade institucional e/ou a possibilidade de uma avaliação da produtividade científica na perspectiva institucional interna muito certamente não são os fatores motivadores para participar do empreendimento RI. Ao contrário, a motivação parece emergir fortemente do desejo de reconhecimento pessoal e do impacto entre os pares.

Uma pesquisa similar, também em 2005, foi realizada em universidades de dez países europeus - Bélgica, França, Reino Unido, Dinamarca, Noruega, Suécia, Finlândia, Alemanha, Itália e Holanda -, bem como no Canadá e Austrália (VAN WESTRIENEN; LYNCH, 2005). O número de repositórios institucionais em cada país variou de no mínimo 1,5% das universidades (caso da Finlândia, com uma única universidade), para no máximo 100%, na Alemanha, Noruega e Holanda. Aqui também a tipologia de conteúdos restringiu-se quase exclusivamente a publicações do corpo docente.

Como o estudo americano, o inquérito europeu também identificou a baixa participação do corpo docente no armazenamento de objetos

em seus repositórios. Van Westrienen e Lynch (2005) identificaram várias razões para essa não participação:

- Dificuldades para informá-los e convencê-los a participar;
- Confusão e incertezas a respeito de questões sobre propriedade intelectual;
- Crédito acadêmico e uso do material estocado (receio de plágio, por exemplo);
- A percepção de que os conteúdos de livre acesso sejam de baixa qualidade, e
- A falta de políticas obrigatórias para depósito.

Bailey Jr. (2006) também conduziu um inquérito em 2005, na perspectiva dos gestores de 123 bibliotecas da Association of Research Library (ARL), EUA. Aqui, as bibliotecas assumiram a responsabilidade de desenvolver e implantar RIs em resposta à demanda institucional. No geral, esses profissionais de informação foram, de longe, os atores mais ativos nesse empreendimento, com uma participação também ativa dos profissionais de TI. As principais motivações para o desenvolvimento de RI citadas foram, em ordem de importância: a promoção da visibilidade institucional, a preservação, o livre acesso e a administração dos conteúdos intelectuais da instituição. Um RI típico possui, em média, uma coleção de cerca de 3.800 objetos digitais, de tipologia diversificada, de *preprints a postprints*, relatórios técnicos, material multimídia, dentre outros. Grande parte dos RIs possuem políticas (normas/procedimentos) explícitas que orientam, entre outros pontos, o perfil de acesso dos diferentes usuários, o que significa que nem todos praticam o acesso livre em sua plenitude. O auto-arquivamento também era submetido a um perfil de autorização, e no geral, já eram depositados tendo passado pelo crivo das comunidades que definiam o perfil nas coleções. Enquanto a totalidade dos RIs almejam a preservação dos objetos digitais, poucos ainda a praticavam. Raras eram ainda as avaliações de uso e impactos dos RIs, o que explicita o pouco cuidado com a visão dos usuários. Novamente, o ponto nevrálgico apontado na pesquisa diz respeito às estratégias para recrutamento de material a ser depositado.

Palmer e Newton (2008), por fim, apresentam uma análise das abordagens de desenvolvimento adotadas por três bibliotecas universitárias norte-americanas, onde procurou comparar as escolhas, as estratégias e condições para atividades direcionadas ao desenvolvimento.

O estudo aponta que as diferenças mais acentuadas resultam de iniciativas como o equilíbrio, a aquisição de conteúdo e a prestação de serviços. A preocupação com propriedade intelectual é item predominante, e os objetivos e políticas do repositório são muitas vezes implícitas, e o valor do depósito para o corpo docente e para a universidade emergem de várias formas. O planejamento complexo, a gestão e o trabalho técnico dos desenvolvedores de repositório dependem cada vez mais de uma coordenação com os bibliotecários e a interação existente entre eles e os acadêmicos/pesquisadores.

Três abordagens foram identificadas em cada uma das três instituições analisadas. A instituição A distingue-se por uma forte ênfase em serviços, e tem uma abordagem pro-ativa ao trabalhar com pesquisadores para resolver seus problemas de gestão da informação e de curadoria de dados. A instituição B apresenta um equilíbrio de desenvolvimento de conteúdo orientado politicamente e inclui iniciativas de serviços seletivos que têm implicações para além do campus. Na instituição C, os desenvolvedores concentraram-se na construção de uma base forte de conteúdo, confrontando-se diretamente com desafios de propriedade intelectual.

Os três casos sugerem um leque de respostas produtivas para os muitos desafios que enfrentam os desenvolvedores de repositórios institucionais, à medida que os repositórios amadurecem, expandem-se e integram-se mais com as atividades das bibliotecas, e continuam a sua importante contribuição para o sempre mutante empreendimento da comunicação científica.

Dentro do tema mais amplo de equilíbrio de conteúdo e demandas de serviços, três dimensões de desenvolvimento de RIs mostraram-se importantes nas entrevistas: objetivos e política, propriedade intelectual e reconhecimento de valor.

Medir o sucesso de um repositório não é somente uma questão de números. Cabe antes perguntar quais eram as metas e se as mesmas

foram alcançadas. O perfil da instituição e as áreas disciplinares/comunidades de pesquisa que ela acolhe serão os guias para o escopo do RI, e para o quantitativo (e qualitativo) potencial que se pode alcançar. Se o comprometimento dos acadêmicos é fundamental, vários são os fatores que os levam a participar, e esses precisam ser melhor identificados e analisados. A sinergia com as atividades e estratégias já existentes, especialmente com os mecanismos de avaliação existentes, é fator fundamental para uma maior participação.

Ferreira e colaboradores (2008), no relato da política de desenvolvimento do RI da Universidade do Minho, Portugal, talvez reconhecidamente um case de sucesso internacional, comparam a experiência aos trabalhos de Hércules, ainda que não somente pelos aspectos técnicos ou garantia de financiamento institucional, mas principalmente pela forma com que os RIs interferem nas práticas tradicionais dos acadêmicos. E deixam aos leitores uma palavra de esperança: “[...] Nevertheless, as soon as an institutional repository is set up, all of the academy’s research output is expected to be placed in the repository in order to increase the academy’s visibility, usage and impact (among other things, such as constituting the long-term memory of the academy).” Talvez a grande diferencial da Universidade do Minho tenha sido a inteligência gerencial para desenvolver valor agregado no RI e prover formas criativas e legais para ter acesso a material com restrição de uso.

O que essas experiências deixam claro é que, a despeito do maior ou menor grau de sucesso dos RIs, os profissionais de informação, e as bibliotecas, são o grupo social relevante mais importante nessa configuração, e podem assim ocupar um papel de destaque. A proatividade desses profissionais pode, por exemplo, forjar alianças criativas com os gestores e acadêmicos/pesquisadores, levando a novas questões e novas soluções. Engajar os pesquisadores na primeira fase de concepção e desenvolvimento dos RIs é fundamental. Guédon (2009) observa que as demandas dos mesmos, enquanto leitores/usuários dos RIs, parecem estar sendo negligenciadas, ou, mal interpretadas.

Cabe, no caso brasileiro, perguntar se os profissionais de informação (obviamente aqui incluídos os bibliotecários) que estão ligados ao serviço nas universidades e instituições de ensino e pesquisa, estão

mobilizados e comprometidos com esse novo desafio. Quiçá a resposta seja positiva!

### **orientações para o futuro: uma resposta para cada pergunta**

Um leitor atento vai perceber que pouco se apresentou, ao longo do texto, uma perspectiva que venha das “macropolíticas” públicas, onde experiências como a norte-americana (capitaneada pela National Institutes of Health – NIH), a britânica (que nasceu no privado com a *Wellcome Trust* e se institucionalizou no Estado) e a australiana (com o programa governamental Asher) são experiências que, por certo, devem servir de aprendizado para outros países. Aqui, duas justificativas são dignas de nota.

Primeira, a ausência desse olhar das macropolíticas decorre da própria perspectiva metodológica adotada para fazer uma leitura dos RIs – no programa SCOT, como em todo conjunto de abordagens “construtivistas” da ciência e tecnologia, não há um *out there*, uma configuração política macro tomada como dada. Cada grupo social relevante, com seus recursos e seus interesses, traz embutido em si uma leitura das mesmas. E é no seio da concepção e do desenvolvimento das novas tecnologias, na resistência e na diferença, que as políticas, macro e micro, são forjadas. Políticas são consequências, não causas.

Segunda, há ainda um grande distanciamento entre as dimensões micro (o “colocar a mão na massa” e sua política cotidiana) e macro (orientações que emanam de níveis superiores, como de órgãos de financiamento) nas discussões sobre os RIs, novamente reflexo de suas diferentes concepções vigentes. Enquanto serviço de informação, as discussões estão muito mais orientadas para o âmbito interno das instituições; enquanto uma estratégia do Movimento do Livre Acesso, as discussões estão mais orientadas para o âmbito das políticas públicas. Essa flexibilidade interpretativa ainda não permite falar sobre os RIs em um *continuum* que vai da macro para micropolítica, e vice-versa. Fazer a defesa dos repositórios na perspectiva do Movimento

de Livre Acesso é uma estratégia *up-down*; desenvolver repositórios como serviços é uma estratégia *down-up*. Em algum momento, elas haverão de se encontrar.

Das citadas experiências de sucesso citadas acima, de três países diferentes, e descritas com cuidado por Guédon (2009), são pinçadas algumas conclusões, na perspectiva das macropolíticas: o envolvimento e a decisão firme do Estado, especialmente por meio de seu corpo legislativo, é fundamental. As agências de fomento parecem ser o fiel da balança, e são elas que melhor encarnam a legitimidade do caráter mandatório de depósito de artigos revisados por pares, tornando-se o melhor contraponto ao poder de fogo das editoras científicas comerciais.

O que essas experiências deixam antever ainda é que um setor de financiamento de pesquisa com perfil mais público que privado é uma vantagem, obviamente porque aí o Estado ganha mais independência em suas decisões. Mas o que é uma vantagem pode também ser uma desvantagem: se o Estado não está mobilizado ou tem pouca capacidade de escuta, os desafios são maiores.

Emerge também dessas experiências, de forma clara, o papel fundamental desempenhado pelos profissionais de informação, na figura de suas sociedades/organizações de classe. Talvez seja isso um testemunho do reconhecimento e do respeito que os mesmos despertam em gestores e pesquisadores, fruto de décadas de trabalho de excelência, especialmente no campo da pesquisa em saúde.

De todas as lições, a mais importante, e também a mais óbvia: a plena realização do potencial dos RIs, enquanto ampliação do acesso a artigos científicos revisados por pares, depende fundamentalmente de sua integração/harmonização com os mecanismos de avaliação científica em curso. Enquanto essas ligações não existirem, ou forem frágeis, qualquer estratégia de desenvolvimento será infrutífera, se não irrelevante.

Somente assim os RIs poderão ser a resposta para várias perguntas!

KENNAN, M. A.; WILSON, C. Institutional repositories: review and an information systems perspective. *Library management*, v. 27, n. 4/5, p. 236-248, 2006. Disponível em: <<http://unsworks.unsw.edu.au/vital/access/manager/Repository/unsworks:24>>. Acesso em: 15 dez. 2009.

KING, C. et al. **Scholarly communication: academic values and sustainable models**. UC Berkeley: Center for Studies in Higher Education, 2006. Disponível em: <<http://www.escholarship.org/uc/item/4j89c3f7>>. Acesso em: 16 dez. 2009

KLING, R. Social informatics: a new perspective on social research about information and communication technologies. *Prometheus*, v. 18, n. 3, p. 245-266, 2000.

\_\_\_\_\_. Learning about information technologies and social change: the contribution of social informatics. *The Information Society*, n. 16, p. 217-232, 2000.

\_\_\_\_\_. What is social informatics and why does it matter? *The Information Society*, v. 23, n. 4, p. 205-220, July, 2007.

KURAMOTO, H. Informação científica: proposta de um novo modelo para o Brasil. *Ci. Inf.*, Brasília, v. 35, n. 2, Aug. 2006. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19652006000200010&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652006000200010&lng=en&nrm=iso)>. Acesso em: 5 dez. 2009. doi: 10.1590/S0100-19652006000200010.

LYNCH, C. A. Institutional repositories: essential infrastructure for scholarship in the digital age. *ARL Bimonthly Report*, 2003. Disponível em: <<http://www.arl.org/newsitr/226/ir.html>>. Acesso em: 15 dez. 2009.

LYNCH, C. A.; LIPPINCOTT, J. K. Institutional repository deployment in the United States as of early 2005. *D-Lib Magazine*, v. 11, n. 9, Sept. 2005. Disponível em: <<http://www.dlib.org/dlib/september05/lynch09lynch.html>>. Acesso em: 20 dez. 2009.

MACDOWELL, C. S. Evaluating institutional repository development in American academe since early 2005: repository by the numbers, part 2. *D-Lib Magazine*, v. 13, n. 9/10, Sept/Oct. 2007. Disponível em: <<http://www.dlib.org/dlib/september07/mcdowell09mcdowell.html>>. Acesso em: 18 dez. 2009.

OPEN SOCIETY INSTITUTE. **Open access publishing and scholarly societies: a guide**. New York: Open Society Institute, 2005. Disponível em: <[www.soros.org/openaccess/scholarly\\_guide.shtml](http://www.soros.org/openaccess/scholarly_guide.shtml)>. Acesso em: 16 dez. 2009.

PALMER, L. C. T.; NEWTON, M. P. Strategies for institutional repository development: a case study of three evolving initiatives. *Library Trends*, v. 57, n. 2, p. 142-67, 2008 (special issue). Disponível em: [http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1131&context=lib\\_research](http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1131&context=lib_research). Acesso em: 10 jan. 2010.

PINCH, T.; BIJKER, W. The social construction of facts and artifacts or how the sociology of science and the sociology of technology might benefit each

other. In: BIJKER, W.; HUGHES, T.; PINCH, T. 1984. **The social construction of technological systems: new directions in the sociology and history of technology**. Cambridge, MA: MIT Press, 1987.

POYNDRER, R. **Clear blue water**. (Eprint insiders) 7 Mar. 2006 Disponível em: <http://www.eprints.org/community/blog/index.php?archives/43-Clear-blue-water.html> >. Acesso em: 20 dez. 2009.

SHEEREVERS, S.; CRAGIN, M. H. Institutional repositories: current state and future. **Library Trends**, v.57, n.2, p. 89-97, Fall, 2008.

STIGLITZ, J. Knowledge as a global public good. In: KAUL, I.; GRUNBERG, I.; STERN, M. A. (Ed.). **Global public goods: international co-operation in the 21st century**. New York: Oxford University, 1999. p. 308-326.

SWAN, A. et al. Developing a model for e-prints and open access journal content in UK further and higher education. **Learned Publishing**, v. 18, n. 1, p. 25-40, 2005.

SWAN, A.; CARR, L. Institutions, their repositories and the Web. **Serials Review**, v. 34, n. 1, 2008. Disponível em: <http://eprints.ecs.soton.ac.uk/14965/>. Acesso em: 27 nov. 2009.

VAN WESTRIENEN, G.; LYNCH, C. A. Academic institutional repositories: deployment status in 13 nations as of mid 2005. **D-Lib Magazine**, v. 11, 2005. Disponível em: <doi:10.1045/september2005-westrienen>. Acesso em: 17 dez. 2009.

WARE, M. **Pathfinder research on web-based repositories**. London: Publisher and Library/Learning Solutions, 2004. Disponível em: <<http://www.palsgroup.org.uk/palsweb/>

[palsweb.nsf/79b0d164e01a6cb880256ae0004a0e34/8c43ce800a9c67cd80256e370051e88a/\\$FILE/PALS%20report%20on%20Institutional%20Repositories.pdf](http://www.palsweb.nsf/79b0d164e01a6cb880256ae0004a0e34/8c43ce800a9c67cd80256e370051e88a/$FILE/PALS%20report%20on%20Institutional%20Repositories.pdf). Acesso em: 15 dez. 2009.

WILLIAMS, R. The Social shaping of information and communications technologies. In: KUBICECK, H.; DUTTON, W. H.; WILLIAMS, R. **The social shaping of information superhighways**. Frankfurt: Campus Verlag; New York: St Martin Press, 1997.

WILLIAMS, R.; EDGE, D. The social shaping of technology. **Research Policy**, 25, p. 865-899, 1996.

WILSON, M.; HOWCROFT, D. Re-conceptualising failure: social shaping meets IS research. **European Journal of Information Systems**, v. 11, n. 4, p. 236-250, Dec. 2002.

## ANEXO B – Marcação e coleta do texto 2

# Invisible institutional repositories: addressing the low indexing ratios of IRs in Google

Kenning Arlitsch and Patrick S. O'Brien

J. Willard Marriott Library, University of Utah, Salt Lake City, Utah, USA

K.

**Componentes em fundo vermelho** – coleção de elementos componentes  
**Itens ambientais em verde** – coleção de itens que não fazem parte do sistema mas atuam ou sofrem ação por algum componente  
**Ligações em azul claro** – coleção de ligações entre componentes e entre esses e itens do ambiente

### Abstract

**Purpose** – Google Scholar has difficulty indexing the contents of institutional repositories, and the authors hypothesize the reason is that most repositories use Dublin Core, which cannot express bibliographic citation information adequately for **academic papers**. Google Scholar makes specific recommendations for repositories, including the use of **publishing industry metadata** schemas over Dublin Core. This paper aims to test a theory that transforming metadata schemas in institutional repositories will lead to increased indexing by Google Scholar.

**Design/methodology/approach** – The authors conducted two surveys of institutional and disciplinary repositories across the USA, using different methodologies. They also conducted three pilot projects that transformed the metadata of a subset of papers from USpace, the University of Utah's institutional repository, and examined the results of Google Scholar's explicit harvests.

**Findings** – Repositories that use GS recommended metadata schemas and express them in HTML meta tags experienced significantly higher indexing ratios. The ease with which **search engine crawlers** can navigate a repository also seems to affect indexing ratio. The second and third metadata transformation pilot projects at Utah were successful, ultimately achieving an indexing ratio of greater than 90 percent.

**Research limitations/implications** – The second survey is limited to 40 titles from each of seven repositories, for a total of 280 titles. A larger survey that covers more repositories may be useful.

**Practical implications** – Institutional repositories are achieving significant mass, and the rate of author **citations from those repositories** may affect **university rankings**. Lack of visibility in Google Scholar, however, will limit the ability of IRs to play a more significant role in those citation rates.

**Social implications** – Transforming metadata can be a difficult and tedious process. The Institute of Museum and Library Services has recently awarded a National Leadership Grant to the University of Utah to continue SEO research with its partner, **OCLC** Inc., and to develop a toolkit that will include automated transformation mechanisms.

**Originality/value** – Little or no research has been published about improving the indexing ratio of institutional repositories in Google Scholar. The authors believe that they are the first to address the possibility of transforming IR metadata to improve indexing ratios in Google Scholar.

**Keywords** Search engines, Digital libraries, Google Scholar, Institutional repositories, Search engine optimization, Metadata

Paper type Research paper

### Introduction

Search engine optimization (SEO) research conducted at the University of Utah has revealed that many institutional repositories (IRs) have a low indexing ratio[1] in Google Scholar (GS). IRs were developed to manage and ensure long-term **access to research**.

© Emerald Group Publishing Limited  
 0737-8833  
 DOI: 10.1108/07378831211213129

The authors would like to thank Dr. Awesome for her expertise, edits, and unflinching support.

publications, and GS was created as a search engine for those publications, whether they reside in IRs, at publisher repositories, or other research-oriented sites. This paper addresses the reasons for the low indexing ratio of many IRs, which the authors believe stem mainly from the metadata requirements of GS and which stand in contrast from the practices of many IRs. The authors conducted two surveys of IRs across the country and implemented three pilot projects designed to increase the indexing ratio of the IR at Utah. Transforming the metadata schema in those pilot projects led to a significant improvement in GS indexing ratio of the sample set. Additional reasons for the low indexing ratio of IRs can be tied to the ease with which GS's crawlers can navigate a given repository.

While much has been written about search engine optimization for general websites, very little has been published about SEO specifically for digital repositories and even less for institutional or disciplinary repositories. The subject of this paper developed from more general digital repository SEO research the authors have conducted at the University of Utah's J. Willard Marriott Library for the past 18 months, and whose continued research has recently been funded by a National Leadership Grant from the Institute of Museum and Library Services (IMLS). OCLC is a formal partner on this grant. The authors have dramatically improved the indexing ratio of Utah's digital library (including IR) in Google's main index, and some of that work will be discussed as background.

Digital repositories relevant to this article are defined as databases that store digitized or born-digital objects, making them freely accessible to the public. These repositories typically run on web server technologies, use descriptive metadata, and because their missions are generally directed at open access they all benefit from successful harvesting and indexing by internet search engines. IRs are that subset of digital repositories that capture and manage the intellectual output of academic institutions or disciplines.

#### Research question

The USpace IR at the University of Utah currently experiences an indexing ratio of less than 0.1 percent in GS, even though the indexing ratio for the same repository improved from approximately 18 to 98 percent in Google's main index after numerous SEO problems were addressed. The authors hypothesized that the average indexing ratio in GS of IRs across the USA is low, and that altering repository metadata to follow one of the publishing industry schemas recommended by Google Scholar would lead to a substantial improvement in the indexing ratio of USpace content.

#### Research method

The authors conducted two surveys to identify the indexing ratio of other IRs, and the methodologies used for both are explained more fully in the survey methodologies section. In brief, the authors selected repositories from the OpenDOAR directory (University of Nottingham, 2011), gathered total item numbers from each repository, and then systematically sampled GS for evidence that those items had been included in its index. Two surveys (survey 1 and survey 2) were conducted, each using a substantially different searching methodology.

To test improvements to the USpace IR the authors gathered indexing ratio statistics and created a feedback loop to measure results of changes they made to repository items, demonstrated in three pilot studies. They submitted sitemaps

LHT  
30,1

62

containing URLs for all items in USpace to GS, and then used Google Webmaster Tools (Google, 2011) to observe the activity of crawlers and the resulting indexing ratios. The authors adjusted the metadata for a subset of repository articles, and following a re-harvest they gathered additional statistics (pilot 1). After this approach failed, the authors conducted discussions with OCLC and GS to confirm a new approach, which led to the development of metadata templates for various academic paper types whose effectiveness was tested during explicit harvests by GS (pilot 2 and 3).

#### Google and Google Scholar

Google and Google Scholar are separate indexes, and GS has a different focus from its much larger parent. Dr Amurag Acharya, GS's founding engineer, has stated that the goal is to offer the "most comprehensive list of research papers available on the Web," and that GS limits its results to "peer reviewed papers, theses, books, abstracts, and technical reports" (Assisi, 2005). More recently, GS has added patents and legal cases to the items it indexes.

GS has its own crawlers (also known as spiders or robots) that visit repositories and publisher sites, among others, to harvest content appropriate for its index. A peculiarity of GS's presentation of academic papers is that it generally provides a link directly to the PDF document. This is expedient for users as it gets them directly to the content, but it also strips any context that may have been provided by the repository's HTML display. In other words, metadata institutional logos, and other information normally displayed to users are lost unless they are inserted into the PDF itself. The practice can also affect the reporting of visitation statistics through website analytics software that utilize page tagging. For instance, Google Analytics requires a tracking code inserted in the HTML of each page of a given website to gather statistics. Each time that page is displayed in a web browser it is counted as a visit by Google Analytics, but separating the PDF file from the HTML display means that the visit will not be counted because the required tracking code is not executed when the PDF is called directly. This problem can be overcome by having the webserver execute a PHP script containing the tracking code before serving the requested PDF, but it is unlikely that many repository managers are doing this, or are even aware that their visitation and download statistics may be underreported as a result of GS's item display practice.

#### Literature review

Internet search engines dominate general information-seeking behavior of users and Google is by far the most popular search engine, consistently grabbing 65 percent share of the "explicit core search market" (Comscore, 2011). Bing powers Microsoft and Yahoo! search sites, capturing another 30 percent of market share. The dominance of search engines is also apparent in the academic sector. A 2005 survey by OCLC demonstrated that 89 percent of college students began their research with internet search engines, and that only 2 percent began at library websites (DeRosa and OCLC, 2005). A repeat of that survey five years later demonstrated that the situation for libraries had only worsened, as 0 percent of respondents reported visiting library websites at the outset of their research (DeRosa et al., 2010). That same report saw a slight drop in traditional search engine use, but also noted for the first time the use of social media search engines for initial research. Another 2005 survey in the UK found that "students prefer to locate information or resources via a search engine above all options, and Google is the search

---

engine of choice" (Griffiths and Brophy, 2005). The information-seeking behaviors of young academic researchers in Sweden displayed an "almost complete dominance of Google as a starting point for searching scientific information" (Haglund and Olsson, 2008).

Faculty search behavior is similar. A study of active faculty researchers at four major universities reported that "researchers find Google and Google Scholar to be amazingly effective" for their information retrieval needs and accept the results as "good enough in many cases" (Kroll and Forsman, 2010). Rieger reports a high degree of use and satisfaction with internet search engines. She notes that "both faculty and students prefer search engines over other resources to support their academic work" and that "there is a broader awareness of specialized Google tools such as Google Scholar and Google Book among faculty members and graduate students" (Rieger, 2009). In a comparison of GS to Web of Science, Mikki states that "the amount of qualified scholarly content has increased considerably in Google Scholar since it was launched in 2004," and that it has developed into a serious research and citation study tool that should be included in information literacy programs (Mikki, 2009).

A review of the literature pertaining to SEO in libraries reveals that much of the published research deals with general websites (e.g. Cahill and Chalut, 2009; Rushton et al., 2008). The minimal research dealing with digital repositories sometimes concludes by suggesting that content be replicated outside the database in a static format in order to make it friendlier to search engines, a method that seems arcane and burdensome, but may have been the best option at the time. "Unless links are located on a static web page, crawlers won't find them, and many such links are not followed" (DeRidder, 2008). Page rank in search engines is another factor that plays into repository visibility. Malaga has shown that 62 percent of users click only on results that appear in the first search engine results page (Malaga, 2008). The high use of internet search engines as primary search mechanisms suggests that digital repositories created by libraries are likely to be nearly invisible to users if their contents are not indexed in these search engines. Search engine and metadata optimization for institutional repositories are also addressed only minimally in the published literature, and the value and use of GS is sometimes questioned. McKay offers that "authors are quite right in perceiving [IRs] as 'islands of information,' [...] a condition that can be addressed by search-engine harvesting [...]". She goes on to say "Google Scholar is not usually the first information source" consulted by academics, though that may have been truer when the article was published in 2007, when GS was relatively new and contained much less content (McKay, 2007). Increased use of GS is demonstrated in a more recent University of Mississippi study in which use rose from 4 to 27 percent of major library databases over a four-year period (Herrera, 2010). A 2006 article on optimizing metadata for search engines acknowledges "the problem may not lie with the search engines but with the data providers," and introduces the concept of "data shoogling" to offer more Google-friendly metadata in digital collections (Dawson and Hamilton, 2006). It does not, however, specifically address institutional repositories or GS. A survey of 540 librarians at 108 ARL libraries notes complaints of "inadequate use of metadata by Google Scholar" (Drewry, 2007), which may support this paper's hypothesis that GS does not find metadata supplied by libraries to be appropriately structured or unique.

Beel, Gipp, and Wilde offer related and significant strategies for optimizing academic papers themselves for better inclusion in search engines, and in GS in

---

particular. Their advice includes optimizing graphics for indexing purposes, writing relevant document titles, and selecting appropriate keywords (Beel et al., 2010). While optimization of the academic papers themselves merits continued exploration and testing, it is beyond the scope of this article.

#### Background on SEO research for digital repositories at the University of Utah

Digital repositories of every type face a common challenge: having their content found by interested users in a crowded sea of information on the internet. Getting found means the repository items must be included in the indexes of major search engines, because that is where the vast majority of users start looking. Unfortunately, many digital repositories show poorly in the results from major search engines. In 2010 the authors conducted a survey of 650 known objects across the thirteen repositories of the Mountain West Digital Library (MWDL), and revealed a disturbing pattern: only 38 percent of **known objects** searched by title were found in Google's index. Worse, this Google search engine results page (SERP) consisted mostly of links back to a search results screen in the local repository, rather than linking directly to the objects. Only 15 percent of the hits on the SERP provided users with direct links to the objects. The known-item title searching method employed by the survey probably produced the best results possible at the time; searching by keyword or subject term would likely have presented even fewer items from the repositories of the libraries and archives in the MWDL.

Search engines can be thought of as "users with substantial constraints: they can't read text in images, can't interpret JavaScript or applets, and can't "view" many other kinds of multimedia content" (Hagens, 2005). In order to be indexed by internet search engines, repository databases and the servers on which they reside must be receptive to crawlers sent out by the search engines. The crawlers follow links to each digital object in the repository; a process greatly facilitated by the submission of sitemaps that function as formal invitations and guides, revealing the preferred URL for each of the repository's objects. The crawlers "harvest" metadata and other information about the objects, sending that information back to the search engine where it is analyzed by algorithms that take many factors into account in deciding whether to add the metadata to the search engine's index. Crawlers that encounter difficulties in the harvesting phase will throw off errors that can be analyzed and addressed using free Webmaster Tools services offered by both Google and Bing. Errors that are not addressed in a timely manner may discourage crawlers from returning, leading to continuing low search engine indexing ratios, or worse, being dropped from the index altogether. Technical problems encountered by crawlers may include, but are not limited to the following:

- conflicts between sitemaps and robots.txt files;
  - slow server response time;
  - dead links or failure to provide appropriate redirects;
  - labyrinths created by repository software, including poorly implemented framesets and JavaScripts, as well as multiple URLs for the same object;
  - poor application of metadata, including re-use of the same metadata terms for multiple objects; and
  - metadata schemas deemed unacceptable by the specific search engine.
-

Additional challenges with search engine optimization for digital repositories may be framed as administrative. These include:

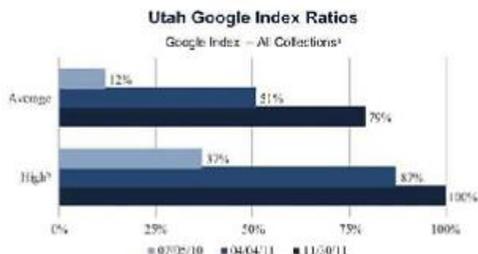
- aligning the goals of the digital library with institutional goals;
- informing, training, motivating, and coordinating staff from various departments;
- establishing an environment of continuous monitoring and addressing crawler errors as they arise; and
- institutionalizing tools to analyze metrics and using them to inform and convince stakeholders of the impact of the digital library.

#### Results at Utah

Analyzing the SEO problems and applying a variety of solutions has resulted in dramatic improvements to the indexing ratio of Utah's digital repositories in Google. The digital repositories (including the IR) managed by the J. Willard Marriott Library at the University of Utah currently run on CONTENTdm v5.4 (OCLC, 2011). While version 5.4 includes some features that are considered unfriendly to search engines, such as JavaScripts, framesets (for compound objects), and multiple URLs for digital objects, those ultimately proved not to be barriers to successful SEO (see Figure 1). Version 6, which was released in late 2010, eliminates those barriers altogether, and Utah is planning a migration.

Figure 1 shows increases in average indexing ratios for all digital collections over a period of 15 months. It also shows improvements in the highest indexing ratio achieved for collections with more than 500 URLs.

Increased indexing ratios have thus far led to a 200 percent increase in referrals from Google, and an 80 percent increase in visits to all digital collections. Indexing ratios of USpace, the University of Utah's IR, have also increased from approximately 18 percent to 98 percent, but only in Google, not in Google Scholar (see Figure 2).

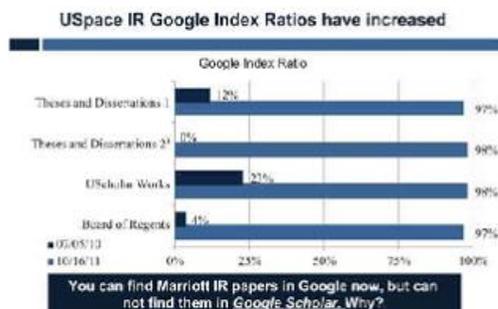


Notes: <sup>1</sup>Google Index Ratio = URLs submitted/URLs Indexed by Google for about 150 collections containing ~170,000 URLs; <sup>2</sup>Highest Index Ratio achieved for collections with over 500 URLs submitted to Google

Figure 1.  
Google index ratio  
improvement for general  
digital collections at Utah

LHT  
30,1

66



Notes: <sup>1</sup>The Theses and Dissertations 2 collection was created after July, 2010. Thus, it had a 0% Google Index Ratio on 07/05/10.

Figure 2.  
Increase in USpace  
indexing ratios in Google

#### Open access and institutional repositories

The **open access movement** was launched to improve access to publicly funded research, and to help libraries deal with rampant inflation in journal subscription prices. According to Peter Suber the open access movement is dependent on **internet technologies** and the **consent of the author** or **copyright holder** (Suber, 2004). Institutional repositories were one product of this movement; they capture the **intellectual output** of the **faculty staff and students of universities** or academic disciplines, and assure perpetual and free access to that output (barring embargo periods or other publisher restrictions). IRs often include electronic **theses** and **dissertations** (ETD), and most are managed by **academic libraries** and some by **scholarly societies**. Over the past decade IRs have variously enjoyed advances and suffered setbacks, but through the consistent work of many individuals at numerous institutions they are achieving enough mass to become viable sources of research publications. They also hold the promise of contributing significantly to author citation rates. Recent research in the UK suggests that institutional repositories may play a crucial role in measuring research output, and in turn may affect university rankings (Key Perspectives and Brown, 2009). The Times Higher Education publishes an annual ranking of the top world universities, and research citations contribute 32.5 percent toward each university's score (The Times Higher Education, 2010).

Libraries have not developed a mechanism to aggregate and search IRs, and thus GS has become the best de facto search engine available for IR content. But just as institutional repositories are gaining enough mass to make them useful and credible sources of **research output**, the difficulties associated with SEO threaten to undermine their potential. Faculty and other authors who contribute publications to IRs may lose interest if their publications can't be located (and cited) in academically-oriented search engines like GS.

#### Surveys of IR indexing ratios in Google Scholar

In October and December 2011 the authors conducted two surveys of institutional and disciplinary repositories to arrive at a preliminary determination of how well GS was indexing them. The IRs were identified through the Directory of Open Access Repositories, also known as OpenDOAR (University of Nottingham, 2011).

Only institutional or disciplinary repositories housed in the USA were selected for these surveys. They were chosen for their academic content, and to represent an approximate real-world distribution of several repository software types: DSpace, Digital Commons, EPrints, [IRIS](#), [OpenStax Commons](#), [Institutional Repository](#), and [arXiv](#) (see Table I). While there are a number of other software types in use, many of them are not found in the U.S. Some repositories found in OpenDOAR were ruled out because it was immediately obvious that they included other types of non-IR digital collections, such as photographs. According to OpenDOAR the arXiv repository software is used only by arXiv, but it was included in Survey 1 because of its size and importance to the scientific community (see Table I for a complete listing of the repositories selected for survey 1).

#### Survey methodologies

Search engine indexing is a dynamic environment. Crawlers return to repositories periodically to pick up new additions, sometimes discarding items if they run into errors, and the repositories themselves are (hopefully) continually growing. Therefore these surveys should be understood to be a snapshot from a specific moment in time.

OpenDOAR records list the number of items in most repositories, but those figures are usually outdated. The authors determined the current number of repository items from figures available on the sites themselves, and in one case by contacting the repository manager. DSpace repositories make it easy to browse by title to reveal all the items in the repository. In the case of digital commons, a dynamic script posts the current total items in the repository. EPrints repositories had a page that listed the number of items by type, the sum of which represented all the items in the repository. Other sites offered similar methods of determining the total number of items contained in the repository.

#### Survey 1 Methodology

In the first survey searches were conducted to determine the number of items indexed by GS from a given repository by using the "site" operator, i.e. search queries in GS were structured in the following manner: "site:repositoryURL." This operator must be used with caution, because in GS it only searches the primary versions of academic papers. In other words, a paper that has been formally published in a journal will be considered the primary version. Additional versions of that paper, including those that appear in IRs may be indexed by GS, but are considered other versions and will only be revealed by clicking the "versions" link (see Figure 3). Because the other versions do not appear on the initial search results page, it is incorrect to assume that the number of results of a search using the site operator shows all the items that GS has indexed from that repository.

The data from this survey confirm a low average primary publication indexing ratio of only 30 percent (see Table I and Figure 4). Being mindful of casting aspersions, the authors are fully aware that their own IR (USpace) currently shows a near zero percent indexing ratio in GS.

Table 1  
 of the  
 primary  
 publications  
 in  
 the  
 field  
 of  
 the  
 study  
 of  
 the  
 history  
 of  
 the  
 world  
 in  
 the  
 19th  
 century

68

LHT  
301

Repository name	Repository software	Repository URL	Repository items	Items in Google Scholar	Indexing ratio (%)
Boston College eScholarship@BC	DigiTool	dollections.bc.edu	1,625	1	0
UW – ResearchWorks Archive	DSpace	digital.lib.washington.edu/dspace	11,285	304	3
Univ of Rochester Research Commons	IR_p	uresearch.rochester.edu	16,184	983	6
caltechsummers	epprints	summers.library.caltech.edu	22,000	2,200	10
D-Scholarship@Pitt	Epprints	d-scholarship.pitt.edu	5,888	696	12
Columbia Univ-Academic Commons	Fedora Backlight	academiccommons.columbia.edu	4,631	586	13
TU Scholarworks	DSpace	scholarworks.tu.edu/dspace	7,782	1,050	13
Texas A&M Repository	DSpace	repository.tamu.edu	46,324	7,530	16
UW Madison-Minds@UW	DSpace	minds.wisc.edu	15,078	2,530	17
eCommons@Cornell	DSpace	ecommons.library.cornell.edu	18,544	3,410	18
Harvard Univ – DASH	DSpace	dash.harvard.edu	6,193	1,710	28
Univ of Oregon-Scholar Bank	DSpace	scholarbank.uoregon.edu/umbank	9,740	2,840	29
Michigan – Deep Blue	DSpace	deepblue.lib.umich.edu	66,038	22,200	34
BYU Scholars Archive	CONTENTdm	scholarsarchive.lib.byu.edu	7,421	2,520	34
RUPUI Scholar	DSpace	scholarworks.rupui.edu	2,109	800	38
Cornell-Digital Commons@ILR	Digital Commons	digitalcommons.ilr.cornell.edu	14,669	5,880	40
Cornell-arXiv	arXiv	arxiv.org	706,936	380,000	47
Aquatic Commons	Epprints	aquaticcommons.org	5,722	3,230	56
Virginia Tech – CS Tech Reports	Epprints	epprints.cs.vt.edu	983	586	60
Digital Commons@UNLincoln	Digital Commons	digitalcommons.unl.edu	50,657	30,200	60
Baylor U – BearDoc	DSpace	beardocs.baylor.edu	928	829	89



LHT  
30,1

70

The authors created a data set for seven repositories from survey 1 by using crawler software to harvest titles from each repository. This method mimicked the process used by internet search engine crawlers, and collected 500 to 1,400 article titles from each repository and saved them into Excel spreadsheets. In some cases, [open access titles](#) in the IRs were easy to identify and entire collections could be crawled. In other cases, it was difficult to isolate the publicly available scholarly papers for an automated crawler because the repositories do not follow the GS recommendations. These difficulties in crawling the IR resulted in less than optimal sampling of titles within the IR collection; in fact the sample may have been biased to favor a higher indexing ratio because the authors made efforts to harvest only academic papers. Using a sampling methodology developed for verifying database backups (LaRock, 2010) those titles were then randomized, and forty titles from each set were searched by copying article titles from the spreadsheets and pasting them into the GS search box. The authors used Zotero to create metadata records and snapshots for each search result, whether the article was found or not. "Versions" links were followed whenever found and the resulting screen was also captured as a snapshot attached to the same metadata record in Zotero.

Of the seven repositories that were sampled, three showed very high indexing ratios (88-98 percent), while the other four showed ratios below 50 percent (see Table II). A discussion about the likely reasons for these differences follows in the section titled "Survey Conclusions."

#### Survey conclusions

The first survey had limitations in terms of calculating a complete index ratio for each IR. However, since use of the site operator in GS reveals only the primary versions of the articles, the average indexing ratio of 30 percent indicates that most IRs do not contain very many primary articles. This raises some interesting questions about the purpose of IRs. Specifically, how much value is really derived from having [open access titles](#) in the IR, given the amount of labor required to put them there, particularly if the primary publisher is open access as well? On the other hand, IRs that largely contain grey literature that is not published elsewhere will likely see a higher indexing ratio with GS precisely because those are the primary articles.

Data from the second survey are much more interesting. Because the authors used crawler software to harvest article titles, they encountered many of the same problems that Internet search engine crawlers face when trying to harvest institutional repositories. The crawling and indexing guidelines shown in Table II were drawn from stated requirements and recommendations from GS's Webmaster Inclusion Guidelines website (Google Scholar, 2010). In general, IRs that followed these guidelines had a much higher indexing ratio (88-98 percent) than sites that did not (38-48 percent). For the purposes of this paper, the most validating differences were found in the expression of publisher metadata schemas (Bepress, Highwire Press, PRISM, or Eprints) in the meta tags within the header tags of the HTML display pages (see Figure 5). Those repositories that did not make their metadata available in one of the recommended publisher schemas within the HTML meta tags generally fared much more poorly than those that did. Further, the repositories that offered absolute URLs to the PDF files for their documents also had far higher indexing ratios than those that did not. Finally, improving crawler efficiency by providing chronological listings of papers, recently

	Cornell	Oregon	Cal Tech	Texas A&M Faculty	UW Aquatic Tech Reports	Columbia	Rochester
Indexing ratio (%)	98	88	88	48	46	45	38
Software	Digital Commons	DSpace	ePrints	DSpace	DSpace	Fedora/Backlight	IR p
Titles available/captured	Unknown/1,421	4,067/1,463	24,146/1,306	763/757	563/539	3,819/1,432	1,543/926
Crawling guidelines							
Browse by date	No	Yes	Yes	Yes	Yes	No	No
Recently added	No	No	Yes	No	No	No	No
10 clicks from home page	Yes	Yes	Yes	No, only first 200	No only first 200	Yes	No
Robots.txt	Yes	Yes, not in not	Yes	Yes, disallows browse by date	Yes, disallows browse by date	Yes, configured	not Yes
Sitemap index	Yes	No	No	Yes, not compliant with standards	No	No	No
Indexing guidelines							
Meta Tag Schema in HTML headers	BePress	DC	ePrints and DC	None	DC and DCTERMS	None	None
Title	Yes	Yes	Yes	No	Yes	No	No
Author	Yes	Yes	Yes	No	Yes	No	No
Pub Date	Yes	Yes	Yes	No	DCTERMS	No	No
Publisher	Yes	Yes	Yes	No	No	No	No
Journal	No	No	Yes	No	No	No	No
Volume	No	No	Yes	No	No	No	No
Issue	No	No	Yes	No	No	No	No
First page	No	No	Yes	No	No	No	No
Last page	No	No	Yes	No	No	No	No
Absolute URL to PDF	Yes	Yes	Yes	No	No	No	No
Instruction	n/a	n/a	n/a	n/a	n/a	No	n/a
Dissertation name	n/a	n/a	n/a	n/a	n/a	No	n/a

Table II  
Shows 2 indexing ratios  
for seven institutional  
repositories

Invisible  
institutional  
repositories  
71

LHT  
30,1

added papers, and a limited number of clicks to publicly available scholarly papers also seemed to positively affect indexing ratio.

GS makes specific recommendations for IR software on its Inclusion Guidelines for Webmasters site (see reference below), but the surveys in this paper demonstrate that software makes little or no difference, the problem cuts across institutions, repository focus, and repository software. Instead, indexing ratio success has much more to do with how carefully a repository follows the guidelines described, above:

72

If you're a university repository, we recommend that you use the latest version of Eprints ([eprints.org](http://eprints.org)), Digital Commons ([digitalcommons.bepress.com](http://digitalcommons.bepress.com)), or DSpace ([dspace.org](http://dspace.org)) software to host your papers. If you use a less common hosting product or service, or an older version of these, please read the rest of this document and make sure that your website meets our technical guidelines (Google Scholar, 2010).

#### Why Google Scholar has difficulty with institutional repositories

Librarians are great believers in standards, and while building digital repositories they have dutifully followed them for scanning, metadata creation, harvesting, and web Services, among others. Search engines, however, are not required to honor standards. For example, in August 2008 Google announced that it was "Retiring support for OAI-PMH[2] in Sitemaps" (Mneller, 2008), causing consternation across the library community. Two years later, GS made the following announcement on its Webmaster inclusion guidelines site: "Use Dublin Core tags (e.g. DC.title) as a last resort – they work poorly for journal papers [...]" (Google Scholar, 2010).

Although Dublin Core is recognized to be a standard of the lowest common denominator, libraries have used it widely for most digital repositories, including IRs. The Dublin Core schema works "poorly for journal papers" because it does not include adequate fields for citation data and because it is interpreted inconsistently. Citation information such as journal name, volume and issue number, and page numbers span of the article is usually entered into a single field, such as DC.Relation or DC.Source in simple Dublin Core, and there is no specified format or consistency. This makes it difficult for a search engine like GS to accurately parse and index the data into their individual bibliographic components. The Dublin Core Metadata Initiative website (DCMI, 2005) does include guidelines for encoding bibliographic citation information using a qualification of the DC.Identifier field (called "bibliographicCitation") but this is still only a single field. It is also unlikely that many repositories have updated to reflect the relatively recent development of DC Qualifiers. Dublin Core also does not facilitate various academic paper types: there is no specific field to distinguish a pre-print from a journal article, a book chapter from a book, a working paper from a conference proceeding, or a dissertation. In short, libraries are not focusing enough on making metadata machine-readable.

Instead of Dublin Core GS recommends using one of the following schemas: Highwire Press, Eprints, Bepress, and PRISM. These schemas are more adept at

Figure 5.  
Example of HTML meta  
tags using of Bepress  
schema

```

1 <meta name="bepress_citation_author" content="Weller, Douglas A." />
2 <meta name="bepress_citation_publisher" content="Cornell University" />
3 <meta name="bepress_citation_author" content="Friedberg, Harold C." />
4 <meta name="bepress_citation_publisher" content="Cornell University" />
5 <meta name="bepress_citation_title" content="Do Expenditures Other Than Institutional Expenditures Affect
6 <meta name="bepress_citation_year" content="2005" />
7 <meta name="bepress_citation_off_url" content="http://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?

```

structuring citation data appropriately. Highwire Press, a division of Stanford University, developed its schema for journal articles and GS extended the tags to cover additional academic paper types, such as working papers, dissertations, manuscripts, conference papers, books and book chapters. The authors used the extended Highwire Press tags in their pilot projects to test the hypothesis that transforming metadata would lead to an increase in indexing ratio in GS for an IR.

#### Pilot 1

Due to the USpace's non-existent showing in GS, the authors began to strategize methods to modify USpace metadata to fit the recommendations. GS explains how Highwire Press tags could map to Dublin Core fields (Google Scholar, 2010). Thus the first step was to begin aligning existing Dublin Core fields with those mappings (see Table III).

The indexing ratio for USpace at the University of Utah prior to the pilot (July 5, 2010) was poor, at best, and can be summarized as follows:

- (1) Index ratio for the three primary USpace IR collections containing 6,482 papers:
  - ranged between 4 percent and 23 percent within Google;
  - average overall Google Index Ratio was 18.33 percent (1,188/6,482); and
  - index ratio within GS was less than 0.1 percent.

The following steps were taken to address the poor indexing ratio:

- (1) Sitemaps representing three IR collections were submitted through Google Webmaster Tools:
  - A total of 6,482 URLs were submitted:
    - Each collection contained between 500 and 4,200 academic papers.

Highwire press tags	Dublin Core tags
citation_author	DC.creator
citation_date	DC.issued
citation_title	DC.title
citation_publisher	DC.publisher
citation_journal_title	DC.relation.ispartof
citation_volume	DC.citation.volume
citation_issue	DC.citation.issue
citation_firstpage	DC.citation.spage
citation_lastpage	DC.citation.epage
citation_issue	n/a
citation_isbn	n/a
citation_keywords	DC.subject
citation_dissertation_institution	DC.publisher
citation_technical_report_institution	DC.publisher
citation_technical_report_number	n/a
citation_language	DC.language
citation_conference_title	DC.publisher
citation_pdf_url	DC.identifier

Table III.  
Map used in first GS pilot

LHT  
30,1

(2) Errors generated during Google crawls were analyzed using Webmaster Tools and improvements were made:

- improved server performance;
- implemented unique title and description tags containing the paper's name and abstract, respectively; and
- implemented "rel ¼ canonical" tags, indicating the preferred URL of each digital object (there were often multiple URLs pointing to each paper).

74

To address the metadata requirements per the Google Scholar inclusion guidelines the authors did the following:

(1) Mapped Dublin Core to Google-supported Highwire Press tags:

- Extended Dublin Core fields according to GS recommendations:
  - journal volume (DC.volume);
  - journal issue (DC.issue);
  - starting page number (DC.citation.spape); and
  - ending page number (DC.citation.epage).

A total of 20 papers were selected for a pilot:

- Verified metadata was accurate and mapped correctly to the HTML "meta name ¼ " fields on display templates as understood from GS inclusion guidelines (see Table III and Figure 6).
- Ensured each of the 20 papers had a full-text PDF that met GS inclusion guideline requirements.
- Embedded the metadata schema directly into five of the PDF files of the papers.
- Provided a "landing page" per GS inclusion guidelines, containing links to the 20 IR pilot papers that was within a few clicks of the home page. This landing page contained links to both a paper's HTML page and its full-text PDF.

Below is an example of converting the bibliographic data being stored in a single field. The DC.Relation field data, were parsed and presented in the HTML header for Google Scholar.

*McKeever, M. & Wolfinger, N.H. (2006). Thanks for Nothing: Changes in Income and Labor Force Participation for Never-Married Mothers since 1982. Institute of Public & International Affairs (IPIA), 4, 1-43.*

```

1 <meta name="DC.title" content="Thanks for nothing: changes in income and labor force
2 <meta name="DC.creator" content="Wolfinger, Nicholas H."
3 <meta name="DC.issue" content="316/317">
4 <meta name="DC.citation.spape" content="1" />
5 <meta name="DC.citation.epage" content="43" />
6 <meta name="DC.publisher" content="University of Utah">
7 <meta name="DC.contributors" content="Institute of Public and International Affairs (&
8 <meta name="DC.identifier" content="https://search.proquest.com/staff/114.000/

```

Figure 6.  
Converting bibliographic  
data

The experiment delivered a significant increase in the Google index ratio for the IR collections (see Figure 2), and as of October 16, 2011 the Google Index ratio for the IR collections was 97.82 percent (10,306/10,536). However there was no effect on the IR's GS index ratio. In fact, not one of twenty USpace papers that had been isolated and optimized was included in the GS index[3].

#### Pilot 2

During the summer of 2011 the authors consulted with OCLC and Google Scholar with the aim of developing and testing a second pilot project. Nineteen papers from USpace were selected for the second pilot:

- (1) Six of seven GS paper types were represented and the full text PDF document was included for each paper. The book paper type was out of scope for this pilot (see Appendix for examples of each paper type):
  - dissertation and thesis;
  - conference article;
  - working paper;
  - manuscript and pre-print;
  - journal article; and
  - book chapter.
- (2) CONTENTdm v6.0 display templates were augmented:
  - embedded Highwire Press meta tags in the HTML page header of display templates using an automated script (see Figure 7);
  - created a browse by year page that provided links to papers in chronological order of publishing date; and
  - created a recently added page that listed papers added to the IR within the last 30 days.

The second pilot was a moderate success, with 62 percent of papers indexed on the first harvest. However, due to unexpected campus network and power outages that took down the test server for an extended period, the pilot was cut short and the results were dropped from GS's index.

#### Pilot 3

For the third and final pilot project, the authors uploaded 56 papers with full-text PDF files, and transformed the Dublin Core metadata to Highwire Press tags as described

```

1 <meta name="citation_title" content="Thanks for nothing: changes in income and life
2 <meta name="citation_author" content="Kohlinger, Nicholas E." />
3 <meta name="citation_author" content="Rufener, Ralfing" />
4 <meta name="citation_date" content="2009-01-01" />
5 <meta name="citation_firstpage" content="1" />
6 <meta name="citation_language" content="en" />
7 <meta name="citation_keywords" content="bioethics; Single mothers; income; single;
8 <meta name="citation_language" content="en" />
9 <meta name="citation_language" content="en" />
10 <meta name="citation_language" content="en" />
11 <meta name="citation_language" content="en" />
12 <meta name="citation_language" content="en" />

```

Invisible  
institutional  
repositories

75

Figure 7.  
Highwire Press tags  
embedded in HTML  
headers

---

LHT  
30,1

earlier. The same six paper types were represented as before. This time more than 90 percent appeared in the GS index after four weeks. Continuing conversations with GS and OCLC will help address lingering issues, but the authors consider this success to be a significant breakthrough.

---

76

#### Transforming metadata

The thought of manually transforming metadata for an IR might induce nausea in repository managers. Fortunately, the IMLS NLG grant recently awarded to the University of Utah intends, as one of its deliverables, to help address this problem. OCLC is a partner in the grant and will develop formal crosswalks between Dublin Core and one or more of the publishing industry schemas recommended by GS. Automated transformation and linked data mechanisms will also be developed to minimize the work required to express citation data more effectively for indexing. The products of that grant will be published in a toolkit by 2014 or sooner.

#### Conclusion

Transforming metadata to GS-preferred metadata schemas is very likely to raise indexing ratio of IRs. The second and third pilot projects described in this paper were successful, demonstrating that transforming from Dublin Core metadata tags to more precise bibliographic Highwire Press tags increased the sample data set GS indexing ratio from 0 percent to 62 percent in the second pilot, and then to more than 90 percent in the third. The authors are cautiously optimistic that continuing discussions with GS and OCLC will eliminate most remaining indexing problems. Transforming metadata to EPrints, PRISM, and Bepress schemas is also likely to have a positive effect, though this assertion will require additional testing.

The low indexing ratio of IRs in GS cuts across institutions and repository software. Despite GS's endorsement of three software packages, the surveys conducted for this paper demonstrates that software is not a deciding factor for indexing ratio in GS. Each of the three recommended software packages showed good indexing ratios for some repositories and poor ratios for others. Rather, the major deciding factors seem to lie in:

- whether the IR has provided crawlers an efficient method to access its scholarly papers, and
- whether acceptable metadata schemas are provided that offer precise bibliographic information within the HTML page header tags.

While transforming metadata seems to be an effective route to getting indexed, individual IRs may have additional SEO-related problems that must be addressed as well. Slow or misconfigured servers, failure to submit viable sitemaps, crawler errors that remain unresolved, failure to provide appropriate server response codes, lack of communication across the organization, and a host of other potential problems must be considered for effective SEO that will raise repositories' visibility in all search engine indexes. Advanced methods for optimizing PDF files may also help to assure inclusion in the GS index. More research and testing is needed, but it is fair to say that a crawler-friendly repository will fare much better in GS than one that poses difficulties to crawlers. Upgrading to current repository software packages may help in this endeavor as product development teams become aware of and address SEO issues.

---

The growing use of GS by researchers underscores the need to address the problem of low IR indexing ratio. As the economic recession has tightened university budgets, more emphasis is being placed on assessment and measurement of outputs. IRs have the potential to raise author citation rates, and in turn to affect university rankings, but this potential may be seriously hampered if IR content is redundant or invisible to researchers who use GS.

## Notes

1. Indexing ratio is defined here as the number of unique URLs from a given repository found in a search engine's index divided by the total number of URLs in the repository.
2. (Open Archives Initiative Protocol for Metadata Harvesting, a common standard for sharing metadata in the library community).
3. USpace added a second theses and dissertations collection after the first GS pilot was started in July, 2010.

## References

- Assisi, F.C. (2005). "Amurag Acharya helped Google's scholarly leap". *INDOLink – Science & Technology*, available at: [www.indolink.com/SciTech/tf010305-075445.php](http://www.indolink.com/SciTech/tf010305-075445.php) (accessed 13 October 2011).
- Beel, J., Gipp, B. and Wilde, E. (2010). "Academic search engine optimization", *Journal of Scholarly Publishing*, Vol. 41 No. 2, pp. 176-90.
- Cahill, K. and Chalut, R. (2009). "Optimal results: what libraries need to know about Google and search engine optimization", *The Reference Librarian*, Vol. 50 No. 3, pp. 234-47.
- Comscore (2011). "comScore releases September 2011 US search engine rankings", available at: [www.comscore.com/Press\\_Events/Press\\_Releases/2011/10/comScore\\_Releases\\_September\\_2011\\_US\\_Search\\_Engine\\_Rankings](http://www.comscore.com/Press_Events/Press_Releases/2011/10/comScore_Releases_September_2011_US_Search_Engine_Rankings) (accessed 22 October 2011).
- Dawson, A. and Hamilton, V. (2006). "Optimising metadata to make high-value content more accessible to Google users", *Journal of Documentation*, Vol. 62, pp. 307-27.
- DCMI (2005). "Guidelines for encoding bibliographic citation information in Dublin Core metadata", *Dublin Core Metadata Initiative*, available at: <http://dublincore.org/documents/dc-citation-guidelines/> (accessed 26 October 2011).
- DeRidder, J.L. (2008). "Googling a digital library", *The Code4Lib Journal*, No. 2, available at: <http://journal.code4lib.org/articles/43> (accessed 5 October 2011).
- DeRosa, C. and OCLC (2005). *Perceptions of Libraries and Information Resources: A Report to the OCLC Membership*, OCLC Online Computer Library Center, Dublin, OH.
- DeRosa, C. et al. (2010). *Perceptions of Libraries, 2010: Context and Community*, OCLC Online Computer Library Center, Inc., Dublin, OH, available at: [www.oclc.org/reports/2010perceptions.htm](http://www.oclc.org/reports/2010perceptions.htm) (accessed 4 October 2011).
- Dewry, J.M. (2007). *Google Scholar, Windows Live Academic Search and Beyond: A study of new tools and changing habits in ARL libraries*, University of North Carolina at Chapel Hill, Chapel Hill, NC, available at: <http://etd.ils.unc.edu/dspace/handle/1901/429> (accessed 21 October 2011).
- Google (2011). "Google Webmaster Central", available at: [www.google.com/webmasters/](http://www.google.com/webmasters/) (accessed 29 October 2011).
- Google Scholar (2010). "Inclusion Guidelines for Webmasters", available at: <http://scholar.google.com/intl/en/scholar/inclusion.html> (accessed 4 October 2011).

- Griffiths, J.R. and Brophy, P. (2005), "Student searching behavior and the web: use of academic resources and Google", *Library Trends*, Spring, pp. 539-54.
- Hagens, A. (2005), "High accessibility is effective search engine optimization", *A List Apart*, available at: [www.alistapart.com/articles/accessibilityseo](http://www.alistapart.com/articles/accessibilityseo) (accessed 4 October 2011).
- Haglund, L. and Olsson, P. (2008), "The impact on university libraries of changes in information behavior among academic researchers: a multiple case study", *The Journal of Academic Librarianship*, Vol. 34 No. 1, pp. 52-9.
- Herrera, G. (2010), "Google Scholar users and user behavior: an exploratory study", *College and Research Libraries*, available at: <http://crf.acrl.org/content/early/2010/07/23/crl-125f1.abstract> (accessed 4 October 2011).
- Key Perspectives and Brown, S. (2009), "A comparative review of research assessment regimes in five countries and the role of libraries in the research assessment process: a pilot study commissioned by OCLC Research", OCLC Research, Dublin, OH.
- Kroll, S. and Forman, R. (2010), *A Slice of Research Life Information Support for Research in the United States*, OCLC Research, Dublin, OH.
- LaRock, T. (2010), "Statistical sampling for verifying database backups", *simple-talk*, available at: [www.simple-talk.com/sql/database-administration/statistical-sampling-for-verifying-database-backups/](http://www.simple-talk.com/sql/database-administration/statistical-sampling-for-verifying-database-backups/) (accessed 10 December 2011).
- McKay, D. (2007), "Institutional repositories and their 'other' users: usability beyond authors", *Ariadne*, No. 52, available at: [www.ariadne.ac.uk/issue52/mckay/](http://www.ariadne.ac.uk/issue52/mckay/) (accessed 15 October 2011).
- Malaga, R.A. (2008), "Worst practices in search engine optimization", *Communications of the ACM*, Vol. 51 No. 12, p. 147.
- Mikki, S. (2009), "Google Scholar compared to Web of Science: a literature review", *Nordic Journal of Information Literacy in Higher Education*, Vol. 1 No. 1, pp. 41-51.
- Mueller, J. (2008), "Retiring support for OAI-PMH in Sitemaps", *Official Google Webmaster Central Blog*, available at: <http://googlewebmastercentral.blogspot.com/2008/04/retiring-support-for-oai-pmh-in.html> (accessed 19 October 2011).
- OCLC (2011), "CONTENTdm Digital Collection Management Software", *CONTENTdm* (OCLC - Digital Collection Services), available at: [www.oclc.org/contentdm/default.htm](http://www.oclc.org/contentdm/default.htm) (accessed 27 October 2011).
- Rieger, O.Y. (2009), "Search engine use behavior of students and faculty: user perceptions and implications for future research", *First Monday*, Vol. 14 No. 12, available at: <http://firstmonday.org/ftbin/cgiwrap/bin/ojs/index.php/fm/article/view/2716/2385> (accessed 21 October 2011).
- Rushon, E.E., Kelehan, M.D. and Strong, M.A. (2003), "Searching for a new way to reach patrons: a search engine optimization pilot project at Binghamton University Libraries", *Journal of Web Librarianship*, Vol. 2 No. 4, pp. 525-47.
- Suber, P. (2004), "Very brief introduction to open access", available at: [www.earlham.edu/~peters/foi/brief.htm](http://www.earlham.edu/~peters/foi/brief.htm) (accessed 15 October 2011).
- (The) Times Higher Education (2010), "The Times Higher Education World University Rankings 2010-2011", available at: [www.timeshighereducation.co.uk/world-university-rankings/](http://www.timeshighereducation.co.uk/world-university-rankings/) (accessed 4 October 2011).
- University of Nottingham (2011), "OpenDOAR - Home Page - Directory of Open Access Repositories", available at: <http://opendoar.org/> (accessed 12 October 2011).

Appendix			Invisible institutional repositories
Meta tag	Pre-print	Journal article	
1. citation_author	Maloney, Krisellen; Antelman, Kristin; Arltisch, Kenning; Butler, John	Maloney, Krisellen; Antelman, Kristin; Arltisch, Kenning; Butler, John	79
2. citation_date	2009	2010	
3. citation_title	Future leaders' views on organizational culture	Future leaders' views on organizational culture	
4. citation_publisher	N/A	Association of College and Research Libraries	
5. citation_journal_title	N/A	College and research libraries	
6. citation_volume		71	
7. citation_issue		4	
8. citation_firstpage	1	323	
9. citation_lastpage	56	347	
10. citation_doi			
11. citation_issue			
12. citation_issue			
13. citation_keywords	Organizational culture	Organizational culture	
16. citation_technical_report_institution	Uspace Institutional Repository, University of Utah	N/A	
17. citation_technical_report_number		N/A	
18. citation_language	en	en	
21. citation_pdf_url	<a href="http://cdm06s.lib.utah.edu/utrls/getfile/collection/uspac/id/10/filename/3.pdf">http://cdm06s.lib.utah.edu/utrls/getfile/collection/uspac/id/10/filename/3.pdf</a>	<a href="http://cdm06s.lib.utah.edu/utrls/getfile/collection/uspac/id/16/filename/17.pdf">http://cdm06s.lib.utah.edu/utrls/getfile/collection/uspac/id/16/filename/17.pdf</a>	
22. citation_abstract_html_url	<a href="http://cdm06s.lib.utah.edu/cdm/singleitem/collection/uspac/id/10/rec/1">http://cdm06s.lib.utah.edu/cdm/singleitem/collection/uspac/id/10/rec/1</a>	<a href="http://cdm06s.lib.utah.edu/cdm/s1">http://cdm06s.lib.utah.edu/cdm/s1</a>	

Table AI.  
Highwire Press metadata mappings for seven paper types

LHT  
30,1

80

Meta tag	PhD	Masters
1. citation_author	Rague, Brian William	Wu, Shanghai
2. citation_date	2010/08	2010/07
3. citation_title	A CS1 pedagogical approach to parallel thinking	Electronic structure and transport property of two-dimensional graphene
8. citation_firstpage	1	1
9. citation_lastpage	234	84
13. citation_keywords	Computer; CS1; Education; Parallel; Programming	Disorder; Electronic structure; Graphene; Transportproperty; Electronic structure
14. citation_dissertation_institution	University of Utah, College of Engineering	University of Utah, College of Science
15. citation_dissertation_name	PhD	MS
18. citation_language	en	en
21. citation_pdf_url	<a href="http://cdmogs.lib.utah.edu/units/getfile/collection/uspac/id/5/filename/19.pdf">http://cdmogs.lib.utah.edu/units/getfile/collection/uspac/id/5/filename/19.pdf</a>	<a href="http://cdmogs.lib.utah.edu/units/getfile/collection/uspac/id/0/filename/4.pdf">http://cdmogs.lib.utah.edu/units/getfile/collection/uspac/id/0/filename/4.pdf</a>
22. citation_abstract_html_url	<a href="http://cdmogs.lib.utah.edu/cdm/singleitem/collection/uspac/id/5/rec/1">http://cdmogs.lib.utah.edu/cdm/singleitem/collection/uspac/id/5/rec/1</a>	<a href="http://cdmogs.lib.utah.edu/cdm/singleitem/collection/uspac/id/0/rec/1">http://cdmogs.lib.utah.edu/cdm/singleitem/collection/uspac/id/0/rec/1</a>

Table AII.

Highwire Press metadata mappings for seven paper types

Notes: Not relevant; 4 – citation\_publisher; 5 – citation\_journal\_title; 6 – citation\_volume; 7 – citation\_issue; 10. citation\_doi; 11 – citation\_issn; 12 – citation\_isbn; 16 – citation\_technical\_report\_institution; 17 – citation\_technical\_report\_number; 19 – citation\_conference\_title; 20 – citation\_inbook\_title

Meta tag	Book chapter	Book
1. citation_author	Riloff, Ellen M.	Ram, Ashwin
2. citation_date	1999	1999
3. citation_title	Information extraction as a stepping stone toward story understanding	Understanding language: understanding computational models of reading
4. citation_publisher	MIT Press	MIT Press
8. citation_firstpage	435	1
9. citation_lastpage	460	519
12. citation_isbn	0-262-18192-4	0-262-18192-4
13. citation_keywords	Information extraction; Story understanding	Information extraction; Story understanding
18. citation_language	en	en
20. citation_inbook_title	Understanding language: understanding computational models of reading	N/A
21. citation_pdf_url	<a href="http://cdmogs.lib.utah.edu/units/getfile/collection/uspac/id/9/filename/3.pdf">http://cdmogs.lib.utah.edu/units/getfile/collection/uspac/id/9/filename/3.pdf</a>	
22. citation_abstract_html_url	<a href="http://cdmogs.lib.utah.edu/cdm/singleitem/collection/uspac/id/9/rec/1">http://cdmogs.lib.utah.edu/cdm/singleitem/collection/uspac/id/9/rec/1</a>	

Table AIII.

Highwire Press metadata mappings for seven paper types

Notes: Not relevant; 5 – citation\_journal\_title; 6 – citation\_volume; 7 – citation\_issue; 10. citation\_doi; 11 – citation\_issn; 14 – citation\_dissertation\_institution; 15 – citation\_dissertation\_name; 16 – citation\_technical\_report\_institution; 17 – citation\_technical\_report\_number; 19 – citation\_conference\_title

Meta tag	Working paper	Invisible institutional repositories
1. citation_author	Wolfiger, Nicholas H.; McKeever, Matthew	
2. citation_date	2006-07-26	
3. citation_title	Thanks for nothing: changes in income and labor force participation for never-married mothers since 1982	
6. citation_volume		81
7. citation_issue		
8. citation_firstpage	1	
9. citation_lastpage	43	
10. citation_doi		
13. citation_keywords	Motherhood; Single Mothers; Income; Population surveys	
16. citation_technical_report_institution	Institute of Public and International Affairs (IPIA), University of Utah	
17. citation_technical_report_number	2006-07-04	
18. citation_language	en	
19. citation_conference_title	101st American Sociological Association (ASA) Annual meeting; 2006 Aug 11-14; Montreal, Canada	
21. citation_pdf_url	<a href="http://cdm6gs.lib.utah.edu/utis/getfile/collection/uspac/id/7/filename/21.pdf">http://cdm6gs.lib.utah.edu/utis/getfile/collection/uspac/id/7/filename/21.pdf</a>	
22. citation_abstract_html_url	<a href="http://cdm6gs.lib.utah.edu/cdm/singleitem/collection/uspac/id/7/rec/1">http://cdm6gs.lib.utah.edu/cdm/singleitem/collection/uspac/id/7/rec/1</a>	

Notes: Not relevant: 4 – citation\_publisher; 5 – citation\_journal\_title; 11 – citation\_issue; 12 – citation\_issue; 14 – citation\_dissertation\_institution; 15 – citation\_dissertation\_name; 20 – citation\_inbook\_title

Table A1V.  
Highwire Press metadata mappings for seven paper types

#### About the authors

Kenning Arlitsch is Associate Dean for IT Services at the J. Willard Marriott Library, University of Utah. He recently completed a 12-month sabbatical, during which he conducted research with OCLC on search engine optimization and network level library technologies. Mr Arlitsch began building the Marriott's digital library program in 1999, and founded the multi-state Mountain West Digital Library, the Utah Digital Newspapers program, and co-founded the Western Soundscape Archive. His department is responsible for digitization, interface design and development, ILS, repository management, and server infrastructure for the library and its extended digital programs. He holds a BA in English from Alfred University in New York, and a Master's degree in Library and Information Science from the University of Wisconsin-Milwaukee. He is also a graduate of the Frye Leadership Institute (2005) and of the Research Libraries Leadership Fellows program (2009), sponsored by the Association of Research Libraries. Kenning Arlitsch is the corresponding author and can be contacted at: [kenning.arlitsch@utah.edu](mailto:kenning.arlitsch@utah.edu)

Patrick S. O'Brien is an expert in customer focused, data driven sales and marketing operations. He specializes in the use of new media channels and internet marketing to increase product visibility, acquire new customers, and improve customer satisfaction. He first began incorporating search engine optimization (SEO) into demand generation marketing programs in 1997. He is a former Accenture Strategy Consultant with over 15 years' experience working with business executives on converting marketing strategy into actionable results within the pharmaceutical, biotechnology, healthcare, financial services and telecommunications industries. Mr O'Brien holds a BA in Economics from UCLA and an MBA in Marketing and Finance from The University of Chicago, Booth School of Business.

To purchase reprints of this article please e-mail: [reprints@emeraldinsight.com](mailto:reprints@emeraldinsight.com)  
Or visit our web site for further details: [www.emeraldinsight.com/reprints](http://www.emeraldinsight.com/reprints)



## 1. FRAMING THE ISSUE OF REPOSITORIES

Repositories are very good examples of what historians and sociologists of technology call socio-technical devices. With such a term, they intend to keep the links between social networks and technical knowledge at the center of their analyses. A technology is neither purely technological nor the direct emanation of a social structure: it belongs to networks that are partly human, partly technical. As a society, human beings form a collective cyborg. Technology studies often label this conclusion as "social construction of technology" (SCOT).

Repositories ought to be examined in the light of the concepts and ideas encountered in the context of SCOT debates. Because this approach has been largely ignored by Open Access discussions, it may be useful to revisit it briefly, if only to bring to the fore some interesting theoretical tools better to understand the repository as a socio-technical object. It should be added that if repositories appear simple, this is largely an illusion.

One of the earliest examples of SCOT-inspired studies was a fascinating examination of the emergence of the modern bicycle (Pinch and Bijker, 1984). In several ways, the present situation of repositories, either institutional or subject-based, is reminiscent of the situation of the bicycle before it stabilized in the now familiar form that has remained the same for over a century: everyone wanted some sort of individual vehicle propelled by human muscle but neither the shape, nor the primary function of these vehicles was clear. Moreover, many partial solutions were in competition. For example, was it better to protect the rider by putting some sort of elastic layer between the wheel and the road, or was it better to build some kind of suspension system in the very frame of the machine?

Historians and sociologists of technology have demonstrated that the bicycle was designed and shaped by interactions between various stakeholders. More precisely, the kind of analysis first put forth by Bijker and Pinch, rested on a number of fundamental ideas that included:

- The notion of "interpretive flexibility": This means that technological design is a process open to various possibilities according to the social context where it emerges;

- The idea of "relevant social groups" (RSG). Nowadays, the buzzword would be "stakeholders", except that the earlier form of expression was clearer as to its essentially collective nature;
- The idea of closure or stabilization: the process of technological design does not go on forever. At some point, it must stop according to some mechanism that will involve some social dimension such as consensus, a vote, etc.<sup>2</sup>

These basic insights have been either criticized or complemented in a variety of ways. For example, more importance has been attributed to the wider social context. Others have striven to clarify the nature of the interactions between the RSG's, in particular by showing greater sensitivity to power relationships and their consequences. Attempts have also been made to move beyond what is perceived to be the agency approach of Bijker and introduce a more structural approach which, of course, dovetails nicely with concern for the "wider context". However, the three basic ideas above remain sufficiently important to be singled out and used to approach the issue of repositories.

Nowadays, the system of **repositories and digital preservation** also involves a wide variety of stakeholders and the debates about Open Access have provided the means by which they have identified themselves. However, knowing which RSG's are involved in the shaping of repositories does not mean that we have a clear perception of what is really at stake from the perspective of each one of these groups. This situation is well reflected in the vocabulary used: everyone seems to know what a repository/depository/archive is and is meant to do; meanwhile the vocabulary has vacillated between several terms and it is only gradually that the word "repository" is becoming dominant in English. In other languages, the issue, paradoxically, emerges more clearly and more quickly, probably because the need to translate an English-language terminology forces the translator to make conscious choices. If we recast this reasoning in SCOT terms, it can be said that, in languages other than English, the translators form a RSG of their own, albeit a minor one, that may or may not overlap with other RSG's.

The idea of power has been introduced more recently within SCOT analyses and it adds important insights into

the social shaping of technologies. RSG's harbor hierarchies of their own, and even oligarchies. Power struggles certainly account for their history and evolution. Also, RSG's display various degrees of social concentration or dispersion. For example, consumers will tend to be quite dispersed and, because of this, will find it relatively more difficult to mount coordinated actions than the tight executive of a well-run company. The more dispersed a RSG is, the more likely its power will be diffuse and more difficult to marshal.

RSG's also compete with each other and, to do so, they rely on access to various resources. Economic resources are always crucial; political resources can be important, especially when policy issues emerge. Less obvious perhaps are what some analysts call "cultural resources," also referred to as "discursive legacies". More simply put, these resources deal with the extent to which a particular technological artifact may resonate with a given cultural theme and how this theme is perceived. An example of this situation is provided by Sport Utility Vehicles: nowadays, with economic worries and a great deal of talk about global warming, SUV's are often described as going against the grain of environmental concerns. Not so long ago, particularly in advertisements, they were portrayed as symbols of power and adventure. Publicity constantly promotes cultural meanings of technical objects to stimulate sales and in so doing reveals the power of cultural resources.

Cultural resources are found at all scales of society; in particular, they can work at macro-social levels as in the SUV example above, but they can also be linked to the "culture" of a particular institution: for example, setting up a computer network will lead to concerns, in particular security concerns, that will vary greatly from one institution (e.g. a Church) to another (e.g. a police precinct or a bank).

Repositories, like any other socio-technical entity, find themselves under the stewardship of various RSG's. Understanding what these groups are, how they relate, what their implicit or explicit agendas are, and what resources they have at their disposal, will clarify the situation greatly. If, occasionally, repositories appear as fuzzy and incoherent objects, this is because they are examined from a variety of viewpoints that are not clearly brought to the fore, and it is also because they are not yet stabilized. Although various stakeholders or RSG's may have a very precise and clear

idea of what they would like repositories to be, the end result may remain in the balance for a period of time that can extend over several years. In short, repositories are technical artifacts and applying some theories developed to analyze such objects should be helpful. In particular, for those committed to Open Access, such an approach would provide a more systematic turn to strategic planning.

## 2. HOW ARE REPOSITORIES PRESENTLY PORTRAYED IN THE LITERATURE?

Two very recent texts provide an excellent entry point to this question. The first one, by Charles Bailey, is designed to introduce "the reader" to "key aspects of institutional repositories" (Bailey Jr., 2008). Unsurprisingly, Bailey begins by a quest for a definition, but leaves us with three distinct possibilities: for Clifford Lynch, it is a "set of services that a university offers", for Mark Ware, it is "a web-based database... of scholarly materials which is institutionally defined", in contrast to a subject-based repository. Finally, for Raym Crow, it is made up of "tools, resources capturing and preserving the intellectual output of a single or multi-university". All three definitions place the preservation function front and center. Lynch and Ware also underscore organization and access issues related to repositories but Lynch places a greater accent on stewardship as a generic goal while Ware adds concerns for interoperability. Crow's document presents a different perspective by emphasizing the "reforming [of] the system of scholarly communication..." and by mentioning that an institutional repository (IR) can be useful for the ways in which a university presents itself to the world.

In themselves, these three definitions are good examples of what SCOT defenders would call "interpretive flexibility". They also show how strongly situated forms of discourse can be. Cliff Lynch, for example, is head of the Coalition for Networked Information (CNI) and, as such, has played an important role in trying to chart a path for the professional evolution of librarianship in an increasingly digital environment. His emphasis on services, organization, and access stand at the heart of the library profession, independently of print or digitization. Mark Ware works at the interface between publishers (in this case the Association of Learned and Professional Society Publishers (ALPSP)

and the Publishers Association) and the Joint Information Systems Committee (JISC) anchored within the British system of higher education. His reference to the neutral term "data-base", and his insistence on interoperability certainly fit well with his need to chart a ground common to two worlds, and identify some division of labor acceptable to both sides.

Raym Crow framed his definition in a position paper designed for an OA advocacy group: the Scholarly Publishing and Academic Resources Coalition (SPARC). His talk of reform and his attempt to gain the support of university managers show strong strategic motivation. Interestingly, both CNI and SPARC are offshoots of the Association of Research Libraries (ARL), but their concerns are quite different: the former studies the "transformative promise of networked information technology", while the latter deals with the politics of scientific and scholarly electronic publishing. It was very engaged in the recent passage of the law mandating the archiving of all the researchers' publications supported by the National Institutes of Health in the USA<sup>4</sup>.

Charles Bailey's article provides more examples of "interpretive flexibility": for example, the reasons supporting the development of a local IR range from increasing the visibility and impact of the institution's scholarship, especially if OA is attached to the IR, to providing unified access to the institution's scholarship. The former point largely reflects the concerns of the researchers; the second point corresponds more to the concerns of university managers.

In another article, Alma Swan and Leslie Carr discuss "Institutions, their Repositories and the Web". They too underscore the role repositories can play to "maximize the visibility of the institution's research outputs". While they equate the issue of visibility with providing Open Access, they also stress the importance of a "mandatory policy on the use of the repository for collecting outputs". In defending IR's, Swan and Carr have chosen the point of view of the institution's managers: thanks to their survey of European institutions, the authors can state that "the primary reason for establishing a digital repository is to increase the visibility of the institution's research outputs by making it Open Access". They also add that "... a repository is a tool that enables senior management in research institutions to collate and assess research, to market their

institution, to facilitate new forms of scholarship and to enable the tools that will produce new knowledge". A concrete example is then briefly analyzed in which it is shown that three rankings of the University of Southampton differ drastically. While the *World University Rankings of the Times Higher Education Supplement* and the *Shanghai Jiao Tong University Academic Ranking of World Universities* place Southampton between the 100<sup>th</sup> and the 200<sup>th</sup> places, the *G-Factor International University Ranking* places the same university in the 25<sup>th</sup> position. The difference is simply that the last ranking system relies on university-to-university web links. If Southampton shows up well there, it is because it has a very rich and useful OA IR.

If we look now at IR's from the perspective of publishers, we come across very different perspectives, including negative or even panicked reactions<sup>5</sup>. At best, they emphasize the role that IR's can play for preservation and for data, perhaps for long-term preservations<sup>6</sup>, but the basic reaction remains very cautious and even hostile. In short, in the literature, IR's are always described from the perspective of some RSG.

### 3. UNDERSTANDING THE "INTERPRETIVE FLEXIBILITY" ATTACHED TO INSTITUTIONAL REPOSITORIES

#### a) Identifying the RSG's

The definition and design of IR's is contested, as the SCOT tells us it should be. All socio-technical entities go through an uncertain phase when their designs have yet to stabilize. During that phase, RSG's and their respective visions become visible and they become all the more visible in moments of divergence and even opposition. In such moment, success or failure will depend on the amount of resources available to RSG's, including their ability to create alliances with other RSG's.

The RSG's affecting the shaping of IR's are actually quite diverse:

1. Researchers and advanced students;
2. Librarians, library consortia and associations;
3. University and research center administrators.

4. Large **commercial publishers**, some society publishers and publishing professional associations.
5. Most society publishers and **university presses**.
6. **Public and private research granting agencies**.
7. **Education funding agencies** (ministries or governmental departments).
8. **Foundations**.
9. Law making bodies.
10. Public interest advocacy.

To these RSG's could be added **infrastructural** elements that can heavily affect the shaping of repositories. **Copyright** laws certainly weigh in on the deposit and access **rules of a repository**, as do the ready availability (or not) of licensing schemes such as **Creative Commons**. Some possible RSG's have been left aside, for example computer services and programmers.

This list would of very limited usefulness, were it not for the possibility of creating some hierarchy within it so as to identify the most important groups. Let us remember that the ability of a group to influence the design of a socio-technical object rests on a number of resources, but that the status of a group within an institution can severely constrain the same group. For example, librarians will be considered as service providers within a university, and will generally find themselves at some disadvantage in comparison **with faculty**.

Overall, repositories are often under the responsibility of libraries, but they can also be under the responsibility of computer services. Sometimes, repositories regroup a number of institutions and can even be national in scope. In Europe, with the DRIVER<sup>9</sup> project, supra-national coordination of repositories begins to appear. This suggests that repositories exist within variable scales involving different kinds of RSG's. In short, repositories cannot be taken as a homogeneous set and their social shaping varies greatly according to the size, status and type of their RSG's.

#### b) **SummitRSG's**

Publishers generally are the owners of the materials that ought to go into the repositories. As can be expected, they contemplate the rise of repositories with some anxieties; at the same time, they have been careful, for the most part, to avoid direct confrontation with two important

interlocutors: researchers and **librarians**. As a result, they have carried out a dual strategy at two different levels: on the one hand, they have aimed at confusing the issues around repositories; on the other, they resist any quest for mandatory deposits that tends to be national in scope or involve a high-level national institution. To refer back to the early discussion on the SCOT approach to technology, publishers know that the dispersion of potential opponents can turn to their collective advantage. By making some issues confusing, they manage to generate some degree of bewilderment and thus stand in the way of concerted action. For example, when publishers allow **self-archiving**, but each of them places a different set of constraints on the process, they will discourage many researchers from going down a route whose meaning is not entirely clear to them<sup>10</sup>. Maintaining control over the "reference" version of any given **article** (the one that can be cited) minimizes concessions while offering the public the facade of a sensible attitude. Publishers generally know well the rhetorical injunction to ensure *capta benevolentiae*. In short, publishers have some very powerful tools to divide and conquer:

1. With regard to researchers, publishers have the privilege of intervening in the selection of **editors-in-chief** if they own the journal. It must be remembered that promotion to **gatekeeper** status is very prestigious for any researcher and comes with a good deal of implicit power.
2. With regard to librarians, the digital context has led to licensing, rather than selling, documents. This transformation has also led to a concentration of forces where nowadays groups of libraries working within a consortium negotiate "Big deals". As a result, publishers often work with consortia leaders and generally manage to assign them to a procurement role with relatively few possibilities for free exchange of information among consortia<sup>11</sup>. This may explain why some consortia have not supported repositories or Open Access journals very vigorously<sup>12</sup>. In fact, with consortia and their peculiar relationship with publishers, the library voice appears to have grown more hesitant, less assured, less unified. For some consortial leaders, maintaining good relations with vendors is more important than solidarity with other consortia. Of course, behaving this way can have its rewards too.



N° 737

IT'S A REPOSITORY, IT'S A REPOSITORY, IT'S A REPOSITORY... OPEN ACCESS, DIGITAL COLLECTIONS AND VALUE

such as invitations to conference events sponsored by publishers<sup>12</sup>.

In short, the richest and most powerful among the publishers have maneuvered very efficiently to divide researchers and librarians and coopt a fraction of each group. They have also managed to cloud the issues sufficiently to make researchers indecisive and librarians less militant. In particular, library associations encounter difficulties in keeping the consortial groups and the advocacy groups on the same page. The International Coalition of Library Consortia (ICOLC) went through difficult discussions before adopting recommendations in 1998 in favor of sharing the results of individual negotiations. Publishers, of course, voiced artful opposition<sup>14</sup>. When the ICOLC principles were revised in 2004, not all consortia signed.

The publishers oppose Open Access but sometimes settle for slowing down its progress sufficiently to discourage its supporters. However, OA moves forward, as the recent law passed in the United States testifies: researchers financed by the National Institutes of Health must deposit their papers in PubMed Central within twelve months<sup>15</sup>. The composition of the alliance needed to obtain such a result will tell a great deal about the relative strengths of the RSG's that have been shaping IR's at the highest level.

The recent political success of Open Access in the United States relies on an institution that is part of a very specific group of RSG's: the **funding agencies**. Some are private, such as the Wellcome Trust in the United Kingdom, while a majority of them in a majority of countries are public. In the United States, medical research is supported both by private (e. g. Howard Hughes Medical Institute (HHMI)) and public (National Institutes of Health) organizations. In many countries, such funding agencies have taken the lead in favor of Open Access. The reasons are clearly laid out by the Wellcome Trust:

The Wellcome Trust has a fundamental interest in ensuring that the availability and accessibility of this material is not adversely affected by the copyright, marketing and distribution strategies used by publishers (whether commercial, not-for-profit or academic)<sup>16</sup>.

Funding agencies such as the Wellcome Trust want to improve the system of scientific communication because

a better system will lead to the better and faster scientific progress. In the particular case of the Wellcome Trust, we are talking about health issues so that the benefits are very easy to comprehend as well as concrete. The mixture of fundamental research needed for progress and the familiarity with painful health situations around all of us makes the position of the private charity easy to grasp.

Similar arguments are relevant for public agencies funding medical research, but they are couched differently. While NIH and its sister institutions elsewhere all share the simple, yet lofty, goals of the Wellcome Trust, they must also do so in such a way as to convince lawmakers. For this reason, value for money has to be shown, as well as concern for the fairness of the decisions taken and the public (voter) good. Symmetrically, publishers have pushed back, particularly in the United States, with arguments having to do with the fundamental importance of market mechanisms, and warnings about misdirected governmental interventions. Large publishers and some powerful scientific societies have lobbied members of Congress to try blocking any legislation favoring the mandating of article deposits in suitable repositories, in particular in PubMed Central<sup>17</sup>.

The difficulty of the battle around the NIH deposit mandate as well as its intensity can be evaluated by the length of time to pass the legislation<sup>18</sup>. The long and tedious path through committees, the false starts, lost votes, vetoed laws and, sometimes, the less than felicitous solutions proposed all bear witness to the incredible work that was accomplished. But who did it? What were the RSG's involved?

The answer to this question is relatively simple but it does include a surprising element: an advocacy group within a library professional organization, the Scholarly Publishing and Academic Resources Coalition (SPARC), most ably led by Heather Joseph, was constantly supported by a motley crew of researchers that had been involved in all of these battles ever since 2000-1. The people behind PLoS, some Nobel Prize winners such as Harold Varmus and a number of other first-rank scientists were involved. The majority of researchers, however, heard the sound and fury of this epic battle only vaguely, if at all. Administrators, for the most part, remained indifferent as well. But another, somewhat unexpected, RSG came to the fore and began to play an

688

ARBOR CLXXX 737 mayo-junio [2009] 561-686 ISSN: 0210-1963

doi: 10.3916/abor.2009.185n.737

extremely important role. Made up of ordinary people with ordinary concerns, the Alliance for Tax Payer Access began to weigh in<sup>19</sup>. This organization is self-described as follows: "The Alliance for Taxpayer Access is a coalition of patient groups, physicians, researchers, educational institutions, publishers and health promotion organizations that support barrier-free access to taxpayer-funded research".

RSG's, let us remember, also draw upon cultural resources to influence the shaping of socio-technical objects. In this case the Alliance occupied an enviable position in the sensitive issue of access to health. In the United States, with the largely privatized system of health insurance and the very high costs of both medicine and drugs, this issue has been very much at the forefront of political debates for nearly two decades and it generates a great deal of anxiety among voters. Although the Alliance was not speaking directly about institutional repositories, it was clearly addressing issues of access to medical knowledge for everyone. The way to obtain this result was to get a law passed that would mandate the deposit of publicly-supported research results into a suitable repository or a set of repositories. For their part, library advocacy groups had access to some financial and political resources, but, with the Alliance, they had a great deal more of both and it is with this particular combination that the lobbying efforts of SPARC got the added traction it needed.

The United States omnibus law that contained the language needed to mandate the deposit of articles in a repository within twelve months of their publication was ultimately obtained through the convergence of librarians, a well-placed and highly-motivated minority of researchers, an advocacy branch of the Association of Research Libraries and an institution allocating billions of dollars each year for medical research (NIH).

A number of lessons can be drawn from this example:

1. The battle waged by large commercial publishers to neutralize the institutional repositories was lost because of the coming together of a number of RSG's, some of whom were not obvious.
2. While publishers succeeded in keeping many librarians and most researchers out of the push in favor of Open Access, they could not prevent some powerful

minorities in both of these groups to continue working actively for Open Access and even join forces.

3. The emergence of funding agencies as strong supporters of Open Access was a crucial moment, and a very dangerous one for the publishers. They had to stop very powerful institutions, with deep pockets, that have a constant, positive, relationship with researchers. The mixture of private and public institutions in that group of funding agencies or charities also prevented publishers from designing a single, simple strategy. In particular, the presence of public institutions meant that the battle had to shift level and instead of being waged at the lower level of universities or below, it had to be fought at the highest levels of government.
4. The Alliance for Taxpayers Access suddenly provided librarians, militant researchers and the funding agency with a very powerful cultural resource. From that point on, the Open Access informal coalition had all the resources it needed: political, cultural and even financial. By contrast, publishers had only their financial resources. They could use those resources to try buying some political resources, but these are not always of the best quality. Finally, publishers had almost no viable cultural resource.

This episode also shows that, ultimately, the battle ground shifted in scale and ended up reaching the highest level of government. This shift should not come as a surprise as it had already been observed in Britain with the debates surrounding the Select Committee on Science and Technology in 2004. In this case, the attempt to move the government itself failed despite a favorable report from the Select Committee<sup>20</sup>. However, when the UK Government directed the question of Open Access to the Research Councils, it gave the Councils the opportunity to choose what they preferred. As research funding agencies, they obviously were sensitive to the arguments that had already moved the Wellcome Trust and nudged NIH in the United States. Not too surprisingly by June 2006, Research Councils UK had drafted and made public a document supporting access to publicly supported research. The press release of 2006 contained an interesting sentence worth quoting in full: "Ideas and knowledge derived from publicly-funded research must be made available and accessible for pub-



N° 737

IT'S A REPOSITORY, IT'S A REPOSITORY, IT'S A REPOSITORY... OPEN ACCESS, DIGITAL COLLECTIONS AND VALUE

688

lic use, interrogation and scrutiny, as widely, rapidly and effectively as practicable<sup>29</sup>. Although the British scene witnessed a to and fro trajectory from Select Committee to Government back to Councils, it is clear that the parameters that have been involved in the United States debates were also at work in Britain: the granting agencies in these two countries (as in France with INSERM, and in Germany with the Max-Planck Institute) stand for the widest access to the research they finance, especially when that financing is public. Researchers and librarians played their role too in Britain. In the UK, however, there was no Alliance for Taxpayers Access and this may have been the missing element that would have allowed the government to accept the Select Committee's recommendations.

In summary, the large commercial publishers, confronted with the growing push of funding agencies in favor of Open Access, carried the battle to the highest legislative and executive levels only to find themselves either thrown back into the hands of the Councils, as in Britain, or, ultimately defeated in Congress in the United States. The lessons learned were crucial and must be carefully mediated for similar battles in other countries. However, this analysis deals with only one layer of the whole repository scene. Other, more modestly scaled, debates have also taken place and are being waged at this very moment. To those we will now turn.

#### c) Inside Research Institutions

For most authors, depositing their research results in a repository does not fulfill their publishing needs: both validation and branding are lacking. Peer review and the prestige of the journal harboring a particular article are perhaps even more important to the researcher than knowing that the journal is going to be available in a number of libraries around the world. Various bureaucratic forms from universities, or granting agencies, will ask researchers about articles appearing in peer-reviewed journals, not about the runs (and costs) of these journals. In publishing, researchers often respond more to institutional pressures ("publish or perish"), than to a deep-seated desire to produce and disseminate new knowledge. For this reason, the ritual passage through the journal phase is essential to researchers because many evaluative procedures are linked to their presence: tenure and promotion committees pay

great attention to the journal variable, as do juries helping to allocate grants. Repositories come only later, if at all.

Except where mandates have been instituted at some institutional level (a department, a faculty or a whole university), the lack of obvious links between the act of depositing one's papers in the local institutional repository and the evaluative procedures of the institution makes the gesture appear largely irrelevant. It is all the more irrelevant that depositories are generally not on the radar screen of researchers when they look for information, except if their libraries are poorly stocked. As for the Open Access advantage<sup>21</sup>, if one compares it with something like the "impact factor" of journals, it simply is not treated as valid or useful argument in most administrative evaluation procedures. In short, many researchers do not pay attention to institutional repositories, do not know what they can really do for them, and do not feel they are losing anything in terms of their careers when they benignly neglect them. One of the fundamental paradoxes of the institutional repositories is that, until now, researchers have not placed them at the center of their preoccupations, be they career promotion or knowledge acquisition (with the exception, once again, of researchers working in developing countries: they know the value of Open Access because they are not subsidized readers like their richer colleagues).

Inside the universities and the research laboratories, a movement largely led by librarians (again) has driven the growth of institutional repositories in hundreds of institutions<sup>29</sup>. Here, the impetus is on collecting the production of a local institution and displaying it to the world. The alliances that have allowed this situation to develop are very specific to each institution but almost always involve the library and the administration. The library, guided by recommendations coming from professional library associations worldwide such as the International Federation of Library Associations and Institutions (IFLA)<sup>30</sup> have developed guidelines, best practices, institutional strategies and a number of other documents to grow repositories. In some universities, much energy has been and is being spent on filling these repositories, often with unconvincing results.

Let us look at some statistics. Repository holdings vary from a few dozen records to several thousand. In the [OpenAire](#) directory of repositories, a search engine al-

doi: 10.9309/Arbor.2009.185A.737

lows us to search by category of documents and offers fourteen possibilities. Dissertation sites (528) are the most frequent and unpublished works (488) follow. Neither are very convincing categories from the viewpoint of researchers looking for authoritative information. Then come publications without any definition. They are found in 345 repositories but we do not know what they are. Preprints occur in 82 cases, and postprints in 256 cases. Even assuming that the two groups do not overlap and deal only with refereed articles, this would mean that less than a third of repositories hold peer-reviewed literature. Other categories are not very helpful: they include software, learning objects, references, multimedia and even patents. In short, while the Open Access movement often reiterates the need to collect the refereed journal literature of the world, the repositories seem to include about anything that can be gathered inside a university. Obviously, this state of affairs will not be terribly attractive to the researcher looking for reliable and authoritative literature. The noise level is much too high. Neither does this situation incite researchers to deposit their articles in such a repository.

What are the SRG's that have shaped repositories in this very peculiar manner in a great many institutions? The answer is quite obvious: librarians. They probably decided to build repositories as part of their general positive feelings for Open Access<sup>28</sup>. However, in order to proceed, they checked with their administrator and probably were asked for a justification, especially if they requested help to build the repository. It is not difficult to imagine the form of the argument taken: our repository will expose the intellectual output of our institutions to the whole world and it will be much easier to monitor what people are producing within the university. It will also, with time, build the intellectual memory of this institution. With such a starting point, repositories came to collect all the local intellectual output, from courses to articles, from student theses to research published in the most **prestigious journals**. At the same time, because the librarian did not have the power to bring about a deposit mandate, and because the administrators were (and always are) worried about creating tensions within the institution, most especially by requiring that faculty do something, the repository limps along with deposit rates that rarely exceed 20 %<sup>29</sup>. And publishers are delighted because institutional repositories, and with it Open Access, have been marginalized.

It is also worthy of note that researchers are largely absent from this effort with few exceptions: occasionally, a librarian may organize a colloquium to advertise the repository to faculty. Typically, and we write here from personal experience, only a few members of the faculty come. Some e-mail campaigns may go on for a while, including visits to various departments, again with mitigated success.

The real needs of the researchers appear to be misconstrued or totally neglected. As readers, they need to find good literature fast, and without too much noise. This is the strength of traditional bibliographies and specialized journals in their particular fields. So far, the tools linked with repositories, while giving access to a great deal of literature, do so in a rather unreliable way<sup>27</sup>.

Symmetrically, we have seen that researchers largely ignore or neglect repositories. For tenure and promotion, or for grant proposals, the repositories do not offer anything yet judged valid by administrators and juries. At the level of the whole institution, the **metrics** used to assess the research of research institutions have created a situation where branding by journals and impact factors dominate the scene: this is how "excellence" is "measured" nowadays<sup>30</sup>. Such methods leave little room for Open Access even though IR's hold the promise of improved impact. However, most researchers are but dimly aware of this so-called "OA advantage" and administrators appear even more oblivious to it.

This leads to a fairly absurd paradox: institutional repositories have been designed for researchers to help them both in their work and their career. However, as the rest of the institution does not integrate IR's in its working practices, researchers do not see their importance and neglect them. On the other hand, librarians design their repositories to capture the research output of the university, but often end up designing a device that gathers all forms of intellectual output, not just research. Moreover, researchers are not consulted very efficiently or very well.

Institutions appear to organize their institutional repositories and set up mandates only when they become part of national research assessment exercise. Such a situation can be observed in Australia with the Asher program<sup>31</sup> that lies at the center of the proposed Research Quality framework procedures. In short, it makes the evaluation



N° 737

IT'S A REPOSITORY, IT'S A REPOSITORY, IT'S AN AKAHOHE... OPEN ACCESS, DIGITAL COLLECTIONS AND VALUE

600

of the research in a given university depend on what is actually collected within the local institution<sup>26</sup>. The Asher program provides help to universities so that they can set up their repository.

This survey of the local forces shaping repositories reveals some striking elements. If we think about the way in which researchers are supposed to be at the center of repositories and their *raison d'être*, the result appears contradictory: researchers are absent both in terms of preoccupations and ability to influence the design of these new institutional devices. Administrators appear more reactive than pro-active: they tend to go along with the librarian's desire to build a depository, but remain skeptical. Meanwhile, the real issues, such as integrating the new device into important evaluative practices are not touched. Students, in particular doctoral students, can also be added to the list of important, yet neglected, partners. As a result, they too practice neglect, not because they are mean spirited, but because the issues do not appear clearly to them.

When progress is noted, it generally signals that more central organizations intervene, as the Australian example shows. The hierarchy of institutions and the action of power centers reappear in this context. It is clear that the financing arm of higher education can be determinant. In some ways, ministries of higher education and granting agencies, although not identical, may find themselves adopting congruent attitudes. Globally, the intervention of ministries, of funding agencies and similarly positioned national institutions points to the fact that policy changes seem to require movement from the very top authorities of a country or administrative region. At the same time, local actors within institutions must also act, and that is where the minorities of researchers advocating Open Access, the militant segments in the library community and the odd administrator can begin to agitate to make the institutional "middleware" (i.e. the cadres of higher education and research) move in useful directions.

Having now completed a quick tour of the main RSG's involved in defining and designing IR's, we can now turn to the last part of this paper which will deal with the possible ways to bring about a positive closure to this long debate. This is the point where analysis and strategy begin to converge in earnest.

#### 4. THE PATH TO CLOSURE: DESIGNING THE RIGHT INSTITUTIONAL REPOSITORIES

Anyone surveying the repository landscape for the first time would be immediately struck by its diversity. And anyone doubting the validity of the "interpretive flexibility" concept would have to explain why such a wide variety of designs exist. Many types of repositories stand beside institutional repositories: national repositories (such as "Hyper-articles en ligne" or HAL in France)<sup>27</sup>, subject-based repositories (such as PubMed Central at NIH or ArXiv at Cornell University)<sup>28</sup> or repository consortia (such as White Rose for the Universities of York, Leeds and Sheffield in the UK or, on a full national scale, DAREnet in Holland)<sup>29</sup>. Contents also vary enormously from one repository to another, as noted earlier, including non-research materials, and materials that are not in Open Access. Some digital collections have been designed in part to create value, as in the case of the "Cream of Science"<sup>30</sup> project, a national repository located within DAREnet in Holland. The scope of these variations may signal the fact that the stabilization of the repository format is still some distance away.

More cohesive is the movement of most repositories as they use a common [metadata](#) standard, [OA-PMH](#), and this standard is being completed by the [Object Reuse and Exchange \(ORE\)](#) specifications<sup>31</sup>. These are important tools that offer the promise of well ordered repositories with the possibility of searching through specified subsets of documents such as [preprints](#) or [theses](#). Protocols such as [OA-PMH](#) are indifferent to proprietary or limited-access digital collections, and to OA materials.

How can repositories and their contents move toward some stabilized, coherent framework? To bring back the vocabulary of SCOT, let us call this stabilization process "closure". Common sense immediately tells us that closure does not mean erasing diversity. Far from being a problem, diversity brings some very positive elements in support of creativity and innovation. Closure should lead to developing a framework in which various types of repositories can function in a complementary and mutually supportive manner.

The vocabulary just used did not emerge by chance: it is inspired by history, in particular the history of computer networks. With computers, diversity could neither be ignored

by the relevant RSG's, nor controlled. A "good" solution in emerging computer networks had to incorporate diversity and even make use of it. The [Internet](#), as we know, was built upon a network of networks where diversity could contribute to the robustness of the whole.

The Internet analogy demonstrates the need for diversity in the design of repositories. It also underscores the potential importance of a networked solution which means thinking about more than a single repository and thinking about their possible relationships. Networked solutions are increasingly being favored because they protect individual autonomy and diversity while ensuring some degree of coordination at a more global level. The result generally benefits everyone (Benkler, 2006). In short, repositories should be conceived as nodes within networks of networks.

The DRIVER project in Europe gives some hints that this networking process has begun. DRIVER seeks to create a "confederation of repositories" across Europe. A recent overview of European repositories brings out a number of interesting results (Van der Graaf, 2007). It identifies some "important stimuli" for the development of IR's and helps understand how repositories are evolving. Among the stimuli, increased visibility and [citations of the publications](#) as well as a simple depositing process were favored by nearly 50 % of the respondents; increasing interest and awareness among faculty and administrators were mentioned by about 33 % of the repositories. Conversely, the lack of a depositing mandate was seen as an inhibitor by over 50 % of the respondents. Nearly a third deplored the failure of funding agencies to mandate deposit.

The stimuli and inhibitors just mentioned have one point in common: they pay attention to the concerns of researchers and administrators. The problem is to engage these two important RSG's. Both groups must help shape the repositories, in direct and frequent dialog with librarians and/or computer services, but this is the detail that has been missing in most cases. Libraries try spreading the good word to faculty and administrators but, as they remain largely outside the development process, the librarians' message often remains unheard and the strategy largely ineffective. The institutional hierarchies and the traditional division of labor through delegation appear in this case to work against true collaborative engagement.

Faculty should take the lead in the design of the repository characteristics that serves them most directly, i.e. with the functions that support research and increase their impact in their own peer communities. Administrators should help develop repositories in such a way that they enrich and improve procedures they use or need to administer. As a result, they would understand better the value and importance of mandating deposits. They would also have a clearer view of the relationship between career management (tenure and promotion for example), evaluation metrics and the search for excellence.

Researchers and faculty members, if invited, will quickly point out their needs. It can be expected that these will include clearly separating the peer-reviewed documents from the other digital collections. They will point out that they work in specific subjects and, therefore, need subject-based sets of documents. They will add that sets must be of a sufficient size to attract any researcher. Consequently, they will most likely suggest that their university should team up with other institutions to reach the size necessary to attract the attention of scholars. The beauty of digitized documents is that they can be indefinitely repackaged to suit the most obscure needs; Open Access documents lend themselves even more readily to this reconfiguration of content.

Researchers also want to make themselves more visible and increase their impact. They will, therefore, ask for tools that make harvesting easier and tools that create some forms of branding. The harvesting issue is essentially being solved with OAI-PMH and with ORE, as well as with various [search engines](#), including [Google Scholar](#) and [Caisler](#). The branding issue is much newer to the repository world, but it is not entirely unknown. Repositories can lean upon the experience of the "Faculty of one thousand" of Biomed Central<sup>18</sup>. They can also pay attention to the Dutch project, "Cream of Science" mentioned earlier. More generally, subject-based repositories, especially if they are the result of networked institutional repositories, can powerfully contribute to complementary forms of branding. There is no question that researchers will deposit their articles far more enthusiastically if they understand that this simple gesture adds to their impact, visibility and, ultimately, authority.

The branding capacity of repositories should also attract the administrators' attention. New forms of branding can



N° 737

IT'S A REPOSITORY, IT'S A REPOSITORY, IT'S A REPOSITORY... OPEN ACCESS, DIGITAL COLLECTIONS AND VALUE

602

go a long way toward correcting some of the most glaring deficiencies of scientific evaluation nowadays, including peer-review. They include quality rating by users<sup>39</sup> or new metrics (for example [downloads](#) and [webometrics](#)). The point here is that Open Access repositories can become very important means to enrich and improve existing forms of scientific and scholarly evaluations.

Creating alliances between researchers, administrators and librarians will inevitably lead to new questions and new solutions. For example, the repackaging of peer-reviewed papers within subject-based collections will lead to redefining relationships between subject-based and institutional repositories. Not only can subject-based repositories harvest institutional repositories, but they can provide some added services that would serve researchers. Symmetrically, institutional repositories can endeavor to build prestigious subsets of articles within various subjects or disciplines.

This argument requires a working alliance between administrators, researchers and librarians beside or even beyond the hierarchies and divisions of labor that have stood in the way of these collaborations so far. In many ways, the present stage of development of repositories brings us back to the Renaissance: then, printers were trying to imagine how to make humanists and metallurgists work together despite deep social divisions and strong hierarchical biases. In this, they were helped by various commands and demands from the power structure that provided both rewards and some degree of enforcement. Likewise, our modern funding agencies are in a position to open opportunities, but also to express specific demands, and so are educational agencies that conduct various evaluation exercises, as documented by the Australian example mentioned earlier.

One final remark may help better to understand the potential of repositories when they begin to provide evaluation and branding services. If we change perspective for a moment and look at Open Access journals, some interesting consequences follow. Perhaps because so much has been made of sustainable models for OA journals and perhaps because the major form of sustainability found so far has been the so-called "author-pays" model (in actuality, the model should be called "author-proxy-pays" model), little notice has been made of the majority of OA journals

that do not require any fees from any authors because they rely on other forms of support. The SoEIO collection of journals<sup>40</sup> in Latin America, Spain and Portugal is supported by governments. Governments do so because they support research and they simply include the cost of publishing in the research budgets. For this reason, the only requirement to publishing in a SoEIO journal (there are 528 titles in the SoEIO collection as of January 2008) is acceptance by peer review. Once the peer-review process is completed, the collections of articles from these journals (over 160,000 articles) form a repository, but a repository endowed with the ability to manage peer review and enriched by a variety of branding tools that are associated with it. In other words, looking at repositories from this angle reveals that OA journals can begin to look suspiciously like subject-based repositories with new functions and services added. Repositories have the potential to present themselves by subject and to add some peer review and branding capacity. And that begins to look suspiciously like a journal. In short, repositories and OA journals are not two entities that shall never meet; on the contrary, they appear perfectly capable of converging, overlapping and even of merging.

Because of this convergence phenomenon, it becomes possible to bring to the design of repositories notions borrowed from journals. For example, journals are important elements in building intellectual communities. Repositories can then be envisioned as part of the architecture of various communities. Librarians who have worked on building communities within universities should feel at ease in this role. But, to repeat, this approach cannot succeed without a strong sense of networking beyond the walls of the local institution. Knowledge is woven across the whole world. Consequently, designing knowledge tools on a parochial basis will fail. Bringing closure to repositories requires harmonizing these devices with the requirements of truly globalized knowledge. Repositories will then become part of a "knowledge society".

In the end, it does not matter if repositories are repositories, collections or archives; in fact they should be all of these. When bicycles stabilized, they fulfilled a whole range of functions. They were tools to go to work, to race, to explore the countryside with the family. They were also toys for children and they adapted one way to men and another to women. Repositories can be just





N° 737

IT'S A REPOSITORY, IT'S A REPOSITORY, IT'S A REPOSITORY... OPEN ACCESS, DIGITAL COLLECTIONS AND VALUE

694

- tory shared by three UK universities (Leeds, Sheffield and York) but this "consortium is not a licensing consortium, unlike OhioLink".
- 13 For examples related to the Canadian consortium, CRKN, see the Elsevier publication *Library Connect*, vol. 4, n.° 3 (August 2006), p. 12 (available on-line at [http://www.elsevier.com/framework\\_librarians/LibraryConnect/LCN403/LCN403.pdf](http://www.elsevier.com/framework_librarians/LibraryConnect/LCN403/LCN403.pdf) and vol. 5, n.° 2 (April 2007), p. 1 (available on-line at <http://libraryconnect.elsevier.com/conn/0502/conn050213.html>).
- 14 As early as March 1998, ICCLC issued a "Statement of Current Perspective and Preferred Practices for the Selection and Purchase of Electronic Information" in which it was stated that "Non-disclosure language, if necessary, should not preclude library consortia from sharing pricing and other significant terms and conditions with other consortia". Elsevier responded in the Fall of 1998 (<http://www1.elsevier.com/homepage/about/esifisue1101.pdf>) with careful language asking not to be pushed into identical arrangements with all consortia.
- 15 See the text of H.R.2784 at [http://thomas.loc.gov/cgi-bin/query/D?r1109:--:temp--110E02YR#](http://thomas.loc.gov/cgi-bin/query/D?r1109:--:temp--110E02YR#:); [the double colon is part of the URL].
- 16 [http://www.wellcome.ac.uk/doc\\_WTD002786.html](http://www.wellcome.ac.uk/doc_WTD002786.html).
- 17 <http://www.pubmedcentral.nih.gov/>. For an example of lobbying language, see [http://www.publishers.org/main/PressCenterPRIISM\\_launch.htm](http://www.publishers.org/main/PressCenterPRIISM_launch.htm).
- 18 See note 5.
- 19 <http://www.taxpayeraccess.org/>.
- 20 <http://www.publications.parliament.uk/pa/cm/200304/cmselect/cmstech/1200/120002.htm>.
- 21 [The full report can be found at http://www.rozk.ac.uk/news/20060826openaccess.htm](http://www.rozk.ac.uk/news/20060826openaccess.htm).
- 22 On the OA advantage, there exists a large collection of texts. For a bibliography of these, as well as comments, see <http://opcit.eprints.org/oaicitation-biblio.html>.
- 23 The "Roar" site (or Registry of Open Access Repositories) lists 980 repositories while "OpenDoar" (directory of open access repositories) holds 1,035 repositories (January 23<sup>rd</sup>, 2008). See <http://www.opendoar.org> and <http://roar.eprints.org>.
- 24 <http://www.ifa.org>.
- 25 In OpenDoar, we find that 838 are institutional repositories, 128 are disciplinary, 53 aggregate their documents from subsidiary repositories and 18 are governmental repositories.
- 26 One of the first institutions to achieve a deposit mandate is the University of Minho in Portugal. See Ferreira, Baptista, Rodrigues and Saraiva (2008).
- 27 Oaister (<http://www.oaister.org/>), a search engine harvesting from 934 repositories gives access to over 14 million documents. Oaister harvests repositories that hold materials in Open Access but in so doing, it regularly points to toll-gated materials that are also kept in a repository.
- 28 On a related subject, William L. Cheung shows how evaluation methods and the peculiar needs of post-doc students intersect each other and affect the building of what he calls a "portfolio of objectives". See Cheung (2008).
- 29 See <http://www.dest.gov.au/NR/rdonlyres/3785571A-6853-4E30-90A1-0057245609C8/18673/FactsheetASHER30May07.pdf>.
- 30 With the recent change of government in Australia, this project may undergo significant changes.
- 31 <http://hal.archives-ouvertes.fr/>
- 32 <http://www.pubmedcentral.nih.gov/> and <http://arxiv.org/>.
- 33 <http://www.darenet.nl/en/page/language.viewsearch.page>.
- 34 <http://www.creamofscience.org/en/page/language.view/eur.page>.
- 35 <http://www.openarchives.org/> OAI-PMH stands for Open archive initiative - Protocol for Metadata Harvesting.
- 36 See <http://www.openarchives.org/ore/documents/ORE-Announcement.html>. A succinct objective of ORE might be as follows: "Object Reuse and Exchange (ORE) will develop specifications that allow distributed repositories to exchange information about their constituent digital objects".
- 37 <http://www.oaister.org/>.
- 38 <http://www.facultyof1000.com/>.
- 39 PloS One has been a most interesting pioneer in this kind of development. See <http://www.plosone.org/home.action> for further information.
- 40 See <http://www.scielo.org/ghp/index.php?lang=en>.

## BIBLIOGRAFIA

- Bailey Jr., Charles W. (2008): "Institutional Repositories, 'Tout de suite'", available on-line at <http://www.digital-scholarship.org/sintoutsuite.pdf>.
- Benkler, Yochai (2006): *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, New Haven, Yale University Press, passim.
- Cheung, William L. (2008): "The Economics of Post-Doc Publishing", *Ethics in Science and Environmental Politics*, vol. 8. Preprint 2008, available on-line at <http://www.in-res.com/articles/essp2008/8/e008pp4.pdf>.
- Drake, Miriam A. (2007): "Scholarly Communication in Turmoil", vol. 24, n.° 2,

- Feb., available on-line at <http://www.infotoday.com/it/feb07/Drake.shtml>.
- Ferreira, M.; Baptista, A. A.; Rodrigues, E. y Saraiva, R. (2008): "Carrots and Sticks. Some Ideas on How to create a Successful Institutional Repository", *D-Lib Magazine*, vol. 14, n.º 1-2, available online at <http://www.dlib.org/dlib/january08/ferreira01ferreira.html>.
- Klein, Hans J. y Kleinman, Daniel Lee (2002): "The Social Construction of Technology: Structural Considerations", *Science, Technology & Human Values*, vol. 27, n.º 1, Winter, pp. 28-52.
- Kohl, David and Sanville, Tom (2008): "More Bang for the Buck: Increasing the Effectiveness of Library Expenditures Through Cooperation", *Library trends*, vol. 54, n.º 3, pp. 394-410.
- Peters, Thomas (2002): "Digital repositories: individual, discipline-based, institutional, consortial, or national?", *Journal of Academic Librarianship*, vol. 26, n.º 6, Nov.-Dec., pp. 414-417.
- Pinch, Trevor J. y Bijker, Wiebe E. (1984): "The Social Construction of Facts and Artifacts: or, How the Sociology of Science and the Sociology of Technology Might Benefit Each Other", *Social Studies of Science*, vol. 14, pp. 399-441, available on-line at: <http://arno.unimaas.nl/show.cgi?fid=1270>.
- Van der Graaf, Maurits (2007): "DRIVER: Seven Items on a European Agenda for Digital Repositories", *Anadne*, vol. 52, July, available on-line at <http://www.anadne.ac.uk/issue52/vandergraaf/>.

