

UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO

MAURA CRISTINA TUPINIQUIM CAVALCANTI LYRA

**SISTEMA DE INFORMAÇÃO SOBRE PESQUISA BASEADO NO
MODELO CRIS: CONCEITOS, TECNOLOGIAS, PADRÕES E
APLICAÇÕES**

RIO DE JANEIRO

2016

MAURA CRISTINA TUPINIQUIM CAVALCANTI LYRA

**SISTEMA DE INFORMAÇÃO SOBRE PESQUISA BASEADO NO
MODELO CRIS: CONCEITOS, TECNOLOGIAS, PADRÕES E
APLICAÇÕES**

Dissertação apresentada ao Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro como requisito parcial para obtenção do título de Mestre.

Orientador: Prof^o. Dr. Luís Fernando Sayão

RIO DE JANEIRO

2016

L992 Lyra, Maura Cristina Tupiniquim Cavalcanti.
Sistema de informação sobre pesquisa baseado no modelo CRIS:
conceitos, tecnologias, padrões e aplicações / Maura Cristina
Tupiniquim Cavalcanti Lyra, 2016.
118 f. ; 30 cm

Orientador: Luís Fernando Sayão.
Dissertação (Mestrado Profissional em Biblioteconomia) –
Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro,
2016.

1. United States.Cooperative State Research Service.Current
Research Information System. 2. CERIF (Common European Research
Information Format). 3. Sistemas de informação - Pesquisa. I. Sayão,
Luís Fernando. II. Universidade Federal do Estado do Rio de Janeiro.
Centro de Ciências Humanas e Sociais. Mestrado Profissional em
Biblioteconomia. III. Título.

CDD – 025.04

MAURA CRISTINA TUPINIQUIM CAVALCANTI LYRA

**SISTEMA DE INFORMAÇÃO SOBRE PESQUISA BASEADO NO
MODELO CRIS: CONCEITOS, TECNOLOGIAS, PADRÕES E
APLICAÇÕES**

Dissertação apresentada ao Programa de
Pós-Graduação em Biblioteconomia da
Universidade Federal do Estado do Rio de
Janeiro como requisito parcial para
obtenção do título de Mestre.

Aprovado em:

BANCA EXAMINADORA

Prof^o. Dr. Luís Fernando Sayão – Orientador
Comissão Nacional de Energia Nuclear / Universidade Federal do Estado do
Rio de Janeiro

Prof^a. Dra. Adriana Olinto Ballesté
Universidade Federal do Estado do Rio de Janeiro

Prof^a. Dra. Luana Farias Sales
Instituto de Engenharia Nuclear / Universidade Federal do Estado do Rio de
Janeiro

AGRADECIMENTOS

Agradeço, primeiramente, a Deus que me deu vida e saúde para chegar até onde estou hoje. Sem ELE nada seria possível! SENHOR, salvaste minha vida e me garantiste a vitória!

Agradeço ao meu marido, meu eterno namorado, que ao longo desses quatro anos de casados disponibilizou horas e vários dias de carinho, paciência e apoio para que meu sonho se tornasse possível. Agradeço a minha filha Maria Eduarda, por mais que você ainda não tenha nascido... tudo o que a mamãe fez até aqui foi por você, todo esforço, todo estudo, todo amor! Mamãe te ama demais, filha!!! Agradeço à minha mãe, genitora e amiga fiel, a qual disponibilizou toda sua vida em função da minha. Posso dizer que agora entendo todo o amor que você sente por mim!

Agradeço às minhas famílias, Tupiniquim e Lyra, que me proporcionaram amor em todos os momentos de dificuldade, em especial, agradeço a minha avó Jorgelina e a minha madrinha/tia Vera Lúcia.

E, finalmente, agradeço ao meu orientador, professor Dr. Luis Fernando Sayão, por todo suporte principalmente na reta final. Obrigada pela amizade e confiança dispensada. Agradeço ainda aos componentes da banca, professora Dra. Adriana Ballesté e Dra. Luana Sales por terem a disponibilidade de participarem desse momento importante da minha carreira acadêmica.

Dedico a vitória deste trabalho a Jesus Cristo, pois sem Ele nenhuma vitória é plena e nenhuma vitória é possível.

“Ainda que eu andasse pelo vale da
sombra da morte, não temeria mal algum,
porque tu estás comigo; a tua vara e o teu
cajado me consolam”.

(Bíblia Sagrada – Salmo 23:4)

RESUMO

O trabalho trata sobre o CRIS – Current Research Information System, um sistema de gestão de informações sobre pesquisa que arrola várias informações relevantes para o pesquisador o que auxilia durante todo o processo de pesquisa, como também os softwares de criação destes. Para chegar ao tema, busca-se embasamento teórico na e-Science e no Big data que justificam o grande volume de dados para que se fizesse necessário esse sistema. Aborda também outros sistemas de pesquisa, como os Repositórios Digitais e de Dados e Bibliotecas Digitais que antecederam o advento dos CRIS. Como método de trabalho foram analisadas normas e padrões consolidados, como o CERIF – Common European Research Information Format, que proporcionam a interoperabilidade destes sistemas.

Palavras-Chave: CRIS. Sistemas de informação sobre pesquisa. CERIF. Ciberinfraestrutura de pesquisa

ABSTRACT

The work deals about CRIS - Current Research Information System, a research information management system which lists various relevant information to the researcher which helps throughout the research process, as well as the creation of these softwares. To get to the theme, seeks to theoretical background in e-Science and Big data that justify the large amounts of data to make it necessary that system. Also discourses other search systems such as Digital Repositories, Data Repositories and Digital Libraries that preceded the advent of CRIS. As a working method, analyzed norms and consolidated standards such as CERIF - Common European Research Information Format, which provide interoperability of these systems.

Key-Words: CRIS. Research Information System. CERIF. Research Cyberinfrastructure

LISTA DE ILUSTRAÇÕES

Figura 1: Fluxo da informação	38
Figura 2: Conceituando um CRIS	59
Figura 3: Modelo CERIF	75
Figura 4: Ligações entre as Entidades CERIF	77

LISTA DE TABELAS E QUADROS

Quadro 1: Análise dos artigos selecionados	19
Quadro 2: Evolução da Ciência através do tempo	24
Quadro 3: Eras da Informação	27
Quadro 4: E-Science x Pesquisa Tradicional	33
Quadro 5: Curadoria, Arquivamento e Preservação	48
Quadro 6: Tipos de Entidades CERIF	76
Quadro 7: Inovações nas atualizações do modelo CERIF	79

SUMÁRIO

1	INTRODUÇÃO	11
1.1	RELEVÂNCIA DO PROBLEMA	14
1.2	OBJETIVOS	16
2	METODOLOGIA	17
2.1	PRIMEIRA ETAPA – SELEÇÃO DO TEMA E FORMULAÇÃO DO PROBLEMA DE PESQUISA	17
2.2	SEGUNDA ETAPA – LEVANTAMENTO BIBLIOGRÁFICO SOBRE CRIS EM BASES DE DADOS DA ÁREA DE CI	18
2.3	TERCEIRA ETAPA – EXPLORAÇÃO DOS TIPOS DE CRIS E PONTUAÇÃO DOS CRIS MAIS CONSIDERÁVEIS NO MEIO DE PESQUISA ACADÊMICA	21
2.4	QUARTA ETAPA – TABULAÇÃO, ANÁLISE E DISCUSSÃO.....	21
3	AMBIENTE DE PESQUISA: E-SCIENCE, E-PESQUISA	22
3.1	A COMUNICAÇÃO CIENTÍFICA	22
3.2	MUDANÇAS NOS PARADIGMAS CIENTÍFICOS E AS RESPECTIVAS CONSEQUÊNCIAS	24
3.3	CIBERINFRAESTRUTURA	34
4	GESTÃO DE INFORMAÇÃO E DO CONHECIMENTO	37
5	BASES DE DADOS, REPOSITÓRIOS E BIBLIOTECAS DIGITAIS	42
5.1	BASES DE DADOS, REPOSITÓRIOS DIGITAIS, BIBLIOTECAS DIGITAIS	42
5.2	DADOS DE PESQUISA, REPOSITÓRIO DE DADOS DE PESQUISA, CURADORIA DIGITAL	45
6	CRIS/RIM (RESEARCH MANAGEMENT SYSTEM	54
6.1	HISTÓRICO	54
6.2	CONCEITOS E DEFINIÇÕES.....	56
6.3	TIPOS DE CRIS	64
6.4	CRIS X REPOSITÓRIOS INSTITUCIONAIS.....	66

6.5	NORMAS/PADRÕES	69
6.6	SOFTWARES DISPONÍVEIS: EXEMPLOS DAS PRINCIPAIS CARACTERÍSTICAS	83
6.6.1	PACOTES COMERCIAIS	84
6.6.2	SOFTWARE LIVRE	87
6.6.3	DESENVOLVIMENTO IN-HOUSE	87
6.7	PRINCIPAIS EXPERIÊNCIAS	88
7	O PAPEL DAS BIBLIOTECAS DE PESQUISA NA GESTÃO DE INFORMAÇÃO PARA A PESQUISA	94
8	GUIA DE SISTEMATIZAÇÃO DA LITERATURA DO MODELO CRIS: CONCEITOS, TECNOLOGIAS, PADRÕES E APLICAÇÕES	96
9	CONCLUSÃO	106
9.1	TRABALHOS FUTUROS	109
	REFERÊNCIAS	111

1 INTRODUÇÃO

O primeiro contato da autora com a biblioteconomia foi há 9 anos no início do curso de graduação em Biblioteconomia na UNIRIO. Desde então, a convivência com os demais alunos e com os professores possibilitou à autora a reflexão de que a informação é atualmente um recurso valioso não só para a comunidade científica, mas para a sociedade como um todo. Tal fato teve como decorrência o anseio por dar continuidade às atividades de pesquisa que já tinha iniciado na graduação, o que deu origem ao desejo de iniciar o mestrado de biblioteconomia.

Mas houve outro ponto primordial para a motivação do desenvolvimento desse trabalho que foi o fato de estar atuando como bibliotecária na PUC-Rio, no Sistema Maxwell. O Sistema Maxwell é o repositório institucional da Pontifícia Universidade Católica do Rio de Janeiro. Nele é possível encontrar parte da produção científica da universidade. Foi criado em 1995 e registrado no INPI – Instituto Nacional de Propriedade Industrial em 1998. Desde então, vem sendo alimentado pelas produções intelectuais dos alunos e professores da PUC-Rio. Nele é possível encontrar: trabalhos de final de curso, monografias de especialização, ETDs (teses e dissertações em texto completo em formato digital), séries, periódicos, atas e exercícios interativos. Através do OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting - mecanismo de interoperabilidade para repositórios) os metadados são abertos aos coletores nacionais e internacionais.

Durante todos este tempo de trabalho na universidade, os administradores da biblioteca onde a autora trabalha têm participado de congressos internacionais e têm oferecido inovações a respeito da biblioteconomia internacional. Contudo, um dos congressos em que os administradores da biblioteca estiveram presentes foi singular, pois tiveram a oportunidade de conhecer uma infraestrutura denominada CRIS (Current Research Information Systems) cujo conceito trouxeram para a equipe. Os administradores da biblioteca explicaram que, além de dados relativos à própria pesquisa, esse sistema suporta outras informações como: contratos, publicações, planos de estudos, projetos em andamento e concluídos, informações sobre pesquisadores, resultados de pesquisas, patentes, dentre outras informações que serão tratadas no decorrer do trabalho.

Ao dar início aos estudos para a dissertação notou-se que o avanço da sociedade como um todo e a passagem da era industrial para a pós-industrial

estruturaram a sociedade da informação ou do conhecimento. Contudo, o avanço trouxe em seu contexto complexidades que requerem novos estudos e metodologias que comportem questões de pesquisa de maneira mais sólida e eficiente.

Na segunda metade do século XX, a necessidade do aumento das atividades de planejamento, de organização da informação e dos seus respectivos serviços, em conjunto das áreas de Biblioteconomia, a Documentação, a Recuperação da Informação e a Ciência da Informação fez com que suas estratégias relacionadas aos problemas informacionais, a organização, ao tratamento e ao uso da informação. Na esfera da gestão, a questão das práticas organizacionais tem gerado discussões metodológicas que suscitaram às organizações o reexame os métodos de gestão que utilizavam. Esses novos métodos de gestão da informação têm desafiado concomitantemente organizações e pesquisadores que analisam teorias, conceitos, metodologias e técnicas de gestão da informação e do conhecimento.

A contínua evolução da tecnologia, principalmente das ferramentas de comunicação que são mediadas por computador e pela Internet, está tornando possível o intercâmbio de informações até mesmo com pequenas organizações.

O ambiente de pesquisa tem se tornado mais complexo, arrolando vários atores que se relacionam: pesquisadores, projetos, agências de fomentos dentre outros. Esses atores precisam de um sistema que comporte infraestruturas mais complexas do que os repositórios institucionais, bibliotecas digitais dentre outros sistemas, pois são muito diferenciados. Mesmo que os Repositórios Institucionais (RIs), os Repositórios de Dados (e-repositórios) e alguns instrumentos disponíveis, como Portal de Periódicos CAPES e Google Scholar, permitam ao pesquisador acessar mais eficientemente as publicações, não suportam a heterogeneidade das informações que são primordiais para gerir por completo todos os ambientes de pesquisas existentes. Além disso, as agências de fomento visam introduzir sistemas mais sofisticados que sejam capazes de suportar todo o ambiente de pesquisa, sem deixar de prestar a assistência necessária aos pesquisadores. Um sistema que pode atender essa necessidade é conhecido como CRIS (Current Research Information System). O CRIS é orientado para a gestão de um grande volume de dados e metadados sobre os atores que compõem o cenário da atividade de pesquisa.

O desenvolvimento de sistemas de informação de pesquisa, como o CRIS, oferece uma ferramenta mais poderosa que é incomparável com dispositivos de pesquisa da internet, pois é potencializado quando baseado em padrões bem

consolidados, através de uma combinação que agrega informações de sistemas diferenciados com uma interface homogênea que possibilita ao pesquisador ter acesso às informações de modo análogo. O que é útil nas duas principais funções da gestão do conhecimento: construir e adicionar à estrutura da comunidade de práticas de um rico acervo de conhecimento, no intuito de que este possa ser usado e reutilizado. Neste raciocínio, o CRIS é capaz de registrar um grande volume de metadados que representam todas as entidades e suas respectivas relações dentro da série de atividades de uma determinada organização de pesquisa. Com isso, o CRIS permite aos pesquisadores tomarem decisões relacionadas aos procedimentos de gerenciamento de pesquisa, tais como produtividade, análises, resultados de pesquisas, patentes e outros. Fato que resulta na maior visibilidade para os trabalhos institucionais.

Por um lado, os pesquisadores almejam contar com um sistema que localize todas as informações relevantes mediante a integração e comparação de informações entre as fontes de dados, em uma interface homogênea. Desejam, ainda, um alto grau de precisão com o mínimo de ruído possível para que um maior número de usuários possam recuperar informações relevantes atrelados a sistemas que contribuam com a organização e a gestão desse processo. Isso justifica o motivo pelo qual os sistemas de gestão de informações sobre pesquisas se propõem a sistematizar a informação de modos cada vez mais eficazes para que os cientistas possam recuperar a informação rapidamente.

Por outro lado, a internet é capaz de facilitar a união dos documentos que estão descentralizados. Assim, cada pesquisador tem acesso a um número maior de informações e a produção de conhecimentos novos que são acessados também por outros pesquisadores que por sua vez podem produzir novos conhecimentos. Este processo é cíclico e induz ao aumento no fluxo da produção de novas informações. O acesso facilitado à informação envolve desde a disponibilidade de informações com uma definição comum e formato, até a forma com que os pesquisadores podem recuperar essa informação. A indexação da informação é essencial neste contexto, a classificação deve ser consistente para todas as fontes de informação da pesquisa; caso contrário, as pessoas recuperam informações que não condizem com o assunto explorado. Contudo, apesar dos pesquisadores possuírem várias vertentes de pesquisa, o que quer dizer que não dependem da biblioteca física, isto não faz com que seja necessário um profissional de

biblioteconomia com novas aptidões, pois devido a fenômenos como a “explosão de dados”, dentre outros que serão tratados no decorrer do trabalho, os pesquisadores precisam de um profissional qualificado para estruturar a informação de modo eficiente, pois só assim será possível não só a economia de tempo como também de recursos financeiros.

No panorama da comunicação científica, os dados de pesquisa são materiais significativos e necessitam de um sistema que os comporte com toda sua complexidade e diversidade. Para tal, serão apresentados também conceitos que abrangem a questão desses objetos digitais informacionais denominados CRIS/ RIMs (Research Information Management). Como mencionado por Sales (2014, p. 19), “O tratamento fornecido a esses dados de pesquisas, no ambiente de um repositório com a finalidade de gestão, preservação e futuro reuso é o que se pretende através das técnicas de curadoria digital”. A falta de preservação destes dados pode significar a perda de anos de pesquisa científica.

Devido à urgência de novos métodos para a produção científica foi evidenciada a inevitabilidade em elaborar infraestruturas de informação que atendam aos novos métodos científicos. Portanto, o propósito dessa dissertação é, mediante a pesquisa científica, tratar e discutir conceitos como “Bases de dados”, “Repositórios Digitais”, “Bibliotecas Científicas”, “Big data”, “Big Science”, “Curadoria digital” e “Sistemas de Gestão de Dados sobre Pesquisa” e desfragmentar os conceitos do CRIS e comprovar a escassez de literatura no Brasil sobre esse tema que responderá as seguintes questões: Quais são as principais características de um CRIS? Qual é a relevância de um CRIS para a sociedade? Existe a necessidade da criação de um CRIS no Brasil? Por quê?

1.1 RELEVÂNCIA DO PROBLEMA

A pesquisa científica influencia a evolução tecnológica, econômica e social. E está vinculada aos sistemas de informação de gestão de pesquisas que dão suporte a todo o processo de produção de pesquisa. Percebe-se que esses sistemas são capazes de auxiliar no desenvolvimento da sociedade, pois reúnem informações valiosas para os pesquisadores. Desde modo, pode-se mensurar a importância desse estudo para a comunidade acadêmica e para sociedade como um todo.

Não obstante, é cada vez mais relevante que os pesquisadores das unidades de informação sejam capazes de encontrar informações relevantes; caso contrário, a produtividade será comprometida pelo tempo empregado na mineração desses dados. Por isso, a gestão dessas informações se torna de grande importância e um grande desafio para os profissionais de informação, para os gestores e para os pesquisadores. Esses profissionais precisam do auxílio de sistemas que armazenem, gerenciem, disseminem e recuperem as referidas informações de forma eficiente e eficaz e que auxiliem os pesquisadores, administradores e usuários em geral. Os profissionais da informação ainda são necessários para fornecer informações na tomada de decisão e dar suporte ao processo de criação de conhecimento.

Em um ambiente de contínuas descobertas, os cientistas possuem a ambição de acessar as informações sobre pesquisa através de uma interface familiar e uniforme que seja capaz de integrar e comparar a informação entre as fontes de dados. Essa interface deve ser capaz de realizar as seguintes tarefas: encaminhar o conteúdo e seu formato pela interface; ao menos no resumo, apresentar a informação de maneira uniforme; as fontes de informação de pesquisa devem possuir classificação consistente; tornar obrigatória a indexação do resumo; fazer com que o vocabulário controlado tenha o mesmo significado em todas as línguas e trabalhar ao máximo no intuito de poupar esforço do pesquisador, sendo capaz de realizar buscas inteligentes em um ambiente multi-linguístico, com a assistência de um vocabulário controlado (ontologia / taxonomia) (ZIMMERMAN, 2002). Por isso, a necessidade de um CRIS que supra os interesses dos pesquisadores com suas características.

Com base nas características apresentadas por Zimmerman, os CRIS sempre se adaptam às novas tecnologias e à ciência como um todo e são capazes de relacionar o ciclo de vida da pesquisa científica por completo. Este será o objeto de estudo desta dissertação.

1.2 OBJETIVOS

✓ Objetivo Geral:

- ✓ Investigar os conceitos, tecnologias e padrões subjacentes aos sistemas de informação para a pesquisa preconizados pelo modelo CRIS.

✓ Objetivos Específicos:

- ✓ Estudar a trajetória evolutiva, identificando as principais iniciativas no país e no mundo de sistemas de informação baseados no modelo CRIS;
- ✓ Caracterizar o papel do bibliotecário frente às novas tecnologias de pesquisa científica e frente ao desenvolvimento de um CRIS.

Esta dissertação será dividida do seguinte modo: após esta Introdução que elenca também o tópico de Relevância do Problema e Objetivos; Seção 2 demonstrará os métodos que foram utilizados durante a confecção do trabalho; Seção 3 tratará do ambiente de pesquisa e das mudanças na comunicação científica; Seção 4 abordará a “Gestão da Informação e do Conhecimento”; Seção 5 analisará as “Bases de Dados, Repositórios, Bibliotecas Digitais”, dados de pesquisa e curadoria digital; Seção 6 tratará do CRIS em seus mais diversos aspectos: histórico, definições, padrões, softwares de desenvolvimento, pacotes comerciais e principais experiências do CRIS no mundo; Seção 7 mencionará “O Papel das Bibliotecas de Pesquisa na Gestão de Informação para a Pesquisa”; Seção 8 elaborará um mapeamento com os principais conceitos, tecnologias, padrões e aplicações e, por fim, a Seção 9 tratará da “Conclusão” desta dissertação e dos trabalhos futuros que serão desenvolvidos pela autora.

2 METODOLOGIA

Para estudar as características que envolvem os sistemas de gestão de informação para a pesquisa, o ambiente em que funcionam, os padrões e tecnologias que dão sustentação a esse tipo de sistema de informação e as suas aplicações mais significativas, delimitou-se o objeto de estudo nos seguintes pontos:

- Definições e características dos Sistemas CRIS;
- Os principais tipos de sistemas CRIS;
- Principais aplicações;
- As principais funções e usuários dos sistemas CRIS;
- Os softwares utilizados na implementação de sistemas CRIS;
- As normas e padrões que dão sustentação aos sistemas CRIS.

De acordo com o compreendido através da obra de Marconi e Lakatos (2003), a presente pesquisa atende a abordagem indireta, pois reúne informações já existentes sobre os temas. É de cunho bibliográfico, uma vez que levanta dados de fontes secundárias e é uma pesquisa exploratória, pois tem o intuito de se familiarizar com o tema ainda pouco conhecido.

No decorrer da investigação foram estabelecidas algumas etapas metodológicas com o objetivo de responder a questão de pesquisa e, simultaneamente, atender aos objetivos que foram anteriormente pautados.

2.1 PRIMEIRA ETAPA – SELEÇÃO DO TEMA E FORMULAÇÃO DO PROBLEMA DE PESQUISA

O primeiro momento da pesquisa envolveu a definição do objeto empírico a ser investigado: “sistemas de informação sobre pesquisa”, mais especificamente os baseados no modelo CRIS. Definido o tema, o esforço de pesquisa seguiu na direção de compreender a natureza dos sistemas CRIS e as suas relações com os outros sistemas de informação voltados para a pesquisa, como os repositórios institucionais e as bibliotecas científicas. A partir desse ponto, se coloca como

problema de pesquisa a caracterização dos sistemas CRIS em termos históricos, conceituais, tecnológicos e de padronização.

Isso posto, segue-se o estabelecimento do objetivo geral dos objetivos específicos, tratados na seção anterior.

2.2 SEGUNDA ETAPA – LEVANTAMENTO BIBLIOGRÁFICO SOBRE CRIS EM BASES DE DADOS DA ÁREA DE CI

O CRIS é um assunto ainda muito raramente discutido nas bibliografias nacionais. Por isso, recuperou-se pouca bibliografia na literatura brasileira. Durante todo o processo de estudo, a pesquisadora se empenhou em desenvolver uma dissertação que caracterizasse o modelo de sistema de informação para pesquisa - CRIS, bem como seus principais conceitos, tecnologias, padrões, aplicações e sua importância para a comunidade científica.

As obras consultadas foram constituídas pelas esferas da Ciência da Informação, da Biblioteconomia e da Ciência da Computação, nas quais foram destacadas publicações científicas apropriadas da temática relacionada à comunicação científica (*big data*, *big science*, curadoria digital, *e-Science*), sistemas de gestão de informações para pesquisa e seus padrões (repositórios institucionais, CRIS e CERIF) e esses termos foram pesquisados nos idiomas português e inglês. Os conteúdos abrangeram o tema e conduziram os componentes do referencial teórico preparado para a análise.

A realização das buscas nas bases de dados prolongou-se entre setembro de 2013 a outubro de 2015, e houve preferência por publicações mais atuais. A triagem dos materiais para pesquisa foi processada nas bases de dados de trabalhos científicos, nacionais e internacionais, tais como: Portal de Periódico da Capes, DSpace do EuroCRIS, BRAPCI, e-LIS e LISA. Uma das técnicas desenvolvidas para busca se dedicava à exploração da bibliografia de alguns artigos quando constatada a relevância do texto para a pesquisa. Outra técnica utilizada foi a indicação da literatura fornecida por especialista. Por intermédio desses métodos, foi possível encontrar artigos e outros materiais significativos para a pesquisa.

A tabela abaixo fornece: uma compilação dos artigos selecionados entre setembro de 2013 e outubro de 2015; analisa a quantidade dos artigos encontrados,

em quais bases foram encontrados e; quais desses foram efetivamente utilizados. Menciona também a quantidade de artigos encontrada na bibliografia de artigos relevantes e a quantidade de artigos indicada por especialistas.

Quadro 1: Análise dos artigos selecionados

Base de Dados	Termos utilizados	Artigos encontrados	Artigos utilizados
Portal de Periódico da Capes	Ciberinfraestrutura	3	1
	e-Science	5	2
	Current Research Information Systems ou CRIS	7	3
BRAPCI	Bases de Dados	9	3
	Bibliotecas Digitais	1	1
	e-Science	6	4
	Big Data	3	3
	Comunicação Científica	5	4
	Curadoria Digital	10	3
	Current Research Information Systems ou CRIS	0	0
	CERIF	0	0
	Gestão da Informação e Gestão do conhecimento	18	4
	Repositório Digital	5	2
	Repositório de Dados	3	1
	Repositório	5	2

	Institucional		
DSpace do EuroCRIS	CERIF	14	1
	Current Research Information Systems ou CRIS	30	11
e-LIS	CERIF	0	0
	Current Research Information Systems ou CRIS	0	0
LISA	CERIF	0	0
	Current Research Information Systems ou CRIS	0	0
Bibliografia encontrada em artigos relevantes	-----x-----	12	12
Bibliografia indicada por especialistas	-----x-----	14	14
Total de artigos encontrados: 150		Total de artigos utilizados: 71	

Fonte: A autora

Através deste quadro, pode-se observar que termos encontrados nas bases nacionais, em sua maioria, não foram procurados nas bases internacionais, pois existem em grande quantidade no Brasil. Mas, vocábulos imprescindíveis para a análise como CRIS e CERIF não foram encontrados nas bases nacionais e encontrados somente em algumas bases internacionais como as do Portal de Periódico da CAPES, como por exemplo o Elsevier, e o DSpace do EuroCRIS.

Um ponto relevante a ser ressaltado é que nem todos os materiais foram utilizados, pois após a análise da autora e o auxílio de especialista verificou-se que nem todos os materiais recuperados eram relevantes para a dissertação, porque

mencionavam superficialmente o assunto ou já havia materiais selecionados anteriormente que eram de maior importância para o trabalho. Como por exemplo, os artigos que analisavam o CRIS em sua maioria mencionavam características do CERIF, por este motivo os artigos procurados com o termo CERIF quase não foram utilizados.

Assim, decidiu-se que os arquivos selecionados seriam categorizados em duas classes: 1 - da comunicação científica e 2 - sistemas de gestão de informações para pesquisa e seus padrões.

2.3 TERCEIRA ETAPA – EXPLORAÇÃO DOS TIPOS DE CRIS E PONTUAÇÃO DOS CRIS MAIS CONSIDERÁVEIS NO MEIO DE PESQUISA ACADÊMICA

Foi possível identificar os exemplos mais significativos de sistemas CRIS em âmbito nacional e internacional. A análise desses sistemas combinados com o levantamento bibliográfico que abrangeu todo o contexto que envolve os sistemas CRIS possibilitou a sistematização que será apresentada na última etapa.

O critério para a escolha dos sistemas avaliados foi a indicação destes sites nas fontes de informação averiguadas e a indicação de especialistas. E este processo pôde ser legitimado pela combinação de uma análise crítica da literatura pertinente, porque alguns desses sistemas foram apontados em numerosas fontes. Logo, a influência e a importância desses portais de gestão de informação de pesquisa puderam ser observadas por sua constante menção na literatura científica.

2.4 QUARTA ETAPA – TABULAÇÃO, ANÁLISE E DISCUSSÃO

Essa etapa consistiu em tabular, categorizar e analisar as informações que caracterizam os sistemas CRIS e seus contextos, e, a partir daí, juntamente com o material bibliográfico que possibilitou a elaboração das seções que definem o referencial teórico, foi possível a análise e discussão e a conclusão da presente pesquisa.

3 AMBIENTE DE PESQUISA: E-SCIENCE, E-PESQUISA

Esta seção trata do contexto no qual evoluiu o cenário do CRIS, além de elucidar algumas definições que apoiam e favorecem o surgimento da presente pesquisa.

Portanto, inicia-se com o esclarecimento do conceito de comunicação científica, o que remete aos paradigmas científicos, sua cronologia e os fenômenos que desencadearam tais progressos, como: *Big data*, *Big Science*, *e-Science* e Ciberinfraestrutura de pesquisa. Esses termos apresentam um cenário favorável para a evolução e importância dos sistemas CRIS.

3.1 A COMUNICAÇÃO CIENTÍFICA

A comunicação científica designa o método específico de produção, consumo e transferência da informação no saber científico. Abrange desde a identificação da questão a ser investigada até a etapa em que a informação produzida é absorvida por outro pesquisador. A comunicação científica é empregada nos meios de relacionamentos entre membros dos grupos científicos, em que permanece cooperando com a criação, transferência, consumo do conhecimento e progresso da ciência. Isto é, há uma relação intrínseca entre a criação do conhecimento científico e a comunicação científica (CUNHA, 2008; LEITE; COSTA, 2007).

Com o advento da internet, as produções em rede inovam a pesquisa e a exploração das informações potencializam o crescimento e a simplicidade do acesso aos materiais. Em virtude disso, a comunicação científica que é particularizada pela busca de maior velocidade na permuta e divulgação das informações tem sido beneficiada. Como consequência, possibilita que o cientista faça uso do sistema de comunicação durante todo o processo de pesquisa, simplificando o processo de consumo da informação (LEITE; COSTA, 2007; SALES, 2014).

O ambiente de pesquisa científica é revolucionado continuamente por novas tecnologias que despontam no meio científico, ou seja, o progresso na ciência é assistido pelos avanços dos recursos tecnológicos. Pinheiro (2003) e Leite e Costa (2007) corroboram esta observação ao relatarem que a análise da comunicação científica engloba a informação científica e tecnológica, porque esta é parte

substancial da infraestrutura da ciência e tecnologia. Por este motivo, tratar sobre comunicação científica indica realçar a importância dos padrões de comunicação entre pares e, em igual proporção, incluir tanto as informações que auxiliam em suas pesquisas, quanto o conhecimento que produzem e disseminam por canais de comunicação adequados.

Torna-se evidente, portanto, que um pesquisador utiliza os canais de comunicação em todo o ciclo do conhecimento – desde a sua criação até a sua divulgação –, o que nos permite afirmar que existe um complexo sistema de comunicação científica que permeia as comunidades científicas e - instituições acadêmicas. [...] É importante destacar que o uso da Internet e de tecnologias emergentes no contexto da comunicação científica tem proporcionado e ampliado, ao longo do tempo, uma série de novas possibilidades e oportunidades de inovação nesse campo. [...] [A] introdução em grande escala das tecnologias de comunicação e informação gerou demanda para o uso da Web para a disseminação dos resultados de pesquisas. Isto fez surgir modelos alternativos para comunicação científica, tais como repositórios institucionais e temáticos, periódicos de acesso aberto, [...] É importante dar destaque ao seu caráter revolucionário, visto que, além de questionar, tem causado transformações paradigmáticas [...], tais como a criação de canais alternativos de veiculação de literatura científica validada ou não (repositórios institucionais ou temáticos), o aumento da visibilidade da produção científica e o auto arquivamento por parte dos próprios pesquisadores. Essas transformações incidem diretamente no modo como os indivíduos em uma comunidade científica criam, compartilham e utilizam o conhecimento. (LEITE; COSTA, 2007, p.93-94)

Targino (2000) argumenta que a comunicação científica é imprescindível à atividade científica, pois agrega o trabalho de cada integrante da comunidade científica, permitindo a permuta e absorção de conceitos de outros teóricos. É a comunicação científica que permite à produção científica e aos cientistas criarem evidências e confiabilidade no meio científico. A autora ainda declara que “o processo de comunicação científica consiste na interação psicológica entre os interesses individuais e grupais, mediante influência recíproca e permanente” (TARGINO, 2000, p.12). Sobretudo, afirma que a ciência evolui a partir do processo de ratificação ou contraposição de princípios ou conceitos, após ponderação realizada pela comunidade científica.

Segundo Gómez (2003), o crescimento da ciência é importante “porque exerce enorme influência nas atividades de comunicação entre os cientistas: o volume de

pesquisas e o de literatura científica”. No intuito de seguir o mesmo raciocínio, a próxima seção trata das transformações dos paradigmas científicos e das consequências nas tecnologias de apoio à pesquisa.

3.2 MUDANÇAS NOS PARADIGMAS CIENTÍFICOS E AS RESPECTIVAS CONSEQUÊNCIAS

O desenvolvimento da ciência foi esquematizado por Costa e Cunha (2014), Gomes (2013), Medeiros e Caregnato (2012), Hey, Tansley e Tolle (2009) da maneira elucidada no quadro abaixo:

Quadro 2: Evolução da Ciência através do tempo

Linha do tempo	Ciência	Características
Milhares de anos atrás	Experimental ou Empírica	- Descrição dos fenômenos naturais.
Últimos 100 anos	Teórica	-Utiliza modelos ou generalizações.
Últimas décadas	Computacional	- Simula fenômenos complexos.
Atualmente	E-Science / Exploração de Dados	- Unificação de teoria, experimento e simulação; - Dados capturados por instrumentos; - Dados gerados por simulações; - Processador por software; - Informação armazenada no computador; - Pesquisador analisa base de dados; - Gerenciamento de dados e estatística.

Fonte: Baseado em Costa e Cunha (2014); Gomes (2013); Medeiros e Caregnato (2012); Hey, Tansley e Tolle (2009).

Após a análise do quadro acima, testifica-se que a ciência desponta acompanhada de seu primeiro paradigma. Este é particularizado pelo seu cunho empírico ou experimental, ou seja, o método de descobertas é executado mediante experimentos e é capaz de descrever fenômenos naturais. Já o segundo paradigma emerge após alguns séculos, no esforço de esclarecer acontecimentos através da teoria. Isso possibilita melhor compreensão dos acontecimentos constatados de maneira empírica, bem como das previsões sobre o comportamento de novos fenômenos, por exemplo: Leis de Kepler, as Leis de Newton ou a Lei de Boyle-Mariotte (CORDEIRO; et. al, 2013).

Esses dois primeiros paradigmas citados – experimental e teórico – são na verdade os dois motores da ciência, que impulsionam o fazer científico. Apesar das tentativas de definição os colocar em lados distintos, atualmente - principalmente devido à interdisciplinaridade que marca a construção do conhecimento científico - esses dois paradigmas são mais complementares do que opostos, sendo difícil, em muitos casos, classificar uma pesquisa unicamente como teórica ou experimental (SALES, 2014, p.40)

Entretanto, quando os cientistas percebem que criaram padrões tão complexos que só poderiam ser resolvidos mediante simulações e os métodos criados no segundo paradigma já não eram suficientes para comprovar suas teorias, desponta o terceiro paradigma. No mesmo cenário, surge a computação científica para solucionar tais simulações; como por exemplo, em 1943, durante a Segunda Guerra Mundial, um dos primeiros computadores de uso geral foi criado para testes balísticos e reutilizado para desempenhar simulações em vários ramos do conhecimento. Fato que viabilizou que muitos estudos fossem alicerçados em simulações numéricas e, não somente, na teoria (CORDEIRO; et. al, 2013).

Após a Segunda Guerra Mundial, o constante desenvolvimento das tecnologias incentivou diversas áreas do conhecimento a desempenharem experimentos em maiores proporções, demandando a operação dos *softwares* de simulações que se revelaria a principal característica do terceiro paradigma posteriormente. Isso resultou no aumento exponencial da quantidade de dados, - foi reputado por “dilúvio de dados” e deu origem à *big science* (CORDEIRO; et. al, 2013; COSTA; CUNHA, 2014).

Na década de 60, o termo *big science* foi mencionado pela primeira vez em um ensaio na revista *Science* onde foi retratada a diversidade de parcerias, instituições e pesquisadores membros, como também o aumento no volume da iniciativa científica. A revista expôs um novo estilo de pesquisa científica, o qual foi elementar para a definição de grande parte das características das pesquisas nas áreas de física, de astronomia e de ciências biológicas (GOMÉZ, 2003).

De fato, conforme manifestado por Gómez (2003), esse conceito possui atributos que se aproximam da pesquisa estratégica, como: mudança de escala e a tentativa de suprimir as fronteiras entre as categorias de ciência pura e da ciência aplicada. É capaz de ser particularizada por seus mecanismos facilitadores que são financiados por agências do governo ou por agências internacionais, nas quais pesquisas são realizadas por equipes de cientistas e técnicos. Ao mesmo tempo, é considerado um modelo atual de organização do conhecimento “em grandes programas de pesquisa ‘orientados por missão’, os quais, para serem viabilizados, requerem a participação de mais de um grupo e instituição de pesquisa, assim como a coordenação efetiva entre o governo federal, a universidade e o setor privado, interagindo de modo cooperativo” (GOMÉZ, 2003, p.64).

A *big science* “ênfatisa os produtos da atividade científica e desenvolve uma representação objetivadora da produção de conhecimentos científicos, plausível de leituras e intervenções externas e quantitativas”. A produção científica na *big science* pretende guiar os cientistas para um novo método de pesquisa, no qual os resultados dos estudos sejam veiculados em artigos produzidos por dezenas, ou até mesmo por centenas de coautores. Essa pesquisa “cooperativa” visa enriquecer a estrutura da produção científica que, em primeira instância, baseava-se em trabalhos individuais ou entre poucos colaboradores. Tão importante quanto os relatórios publicados são os dados gerados pela pesquisa que podem ser reutilizados pelos pesquisadores por diversas vezes (GOMÉZ, 2003).

No panorama global, a ciência também pode ser fragmentada em fases denominadas de “Eras da Informação”. A primeira era da informação é demarcada pela introdução dos sistemas de informação dentro das instituições e é conhecido como a “Era Digital”. A partir da segunda era da informação, identificada como “Era da Rede”, o foco foi redirecionado para relações com o ambiente externo. Tal era é caracterizada pelo advento da internet e o progresso na globalização. Na metade da década de 90, o arranjo entre a internet e o WWW (World Wide Web) já configurava

uma transmutação na quantidade e na diversidade de dados disponíveis. Este era o princípio da terceira era da informação que tem por diferencial a análise de dados na sua forma original (GOMÉZ, 2003).

Para melhor compreensão da cronologia dos fatos das “Eras da Informação”, segue o quadro que demonstra a evolução da utilização dos dados e informação desde sua origem até 2013:

Quadro 3: Eras da Informação

Ano ou Período	Fato
8000 a.C.	Comerciantes sumérios usavam tábuas e fichas de barro para denotar os bens comercializados.
1086	Livro Domesday, um dos mais venerados tesouros britânicos, foi uma contagem abrangente -e sem precedentes -dos ingleses, suas terras e propriedades.
1439	Primeira revolução da informação: Impressora de Gutenberg.
1453 e 1503	De acordo com a historiadora Elizabeth Eisenstein, oito milhões de livros foram impressos. Volume de informações dobra na Europa, em 50 anos.
1662	O comerciante britânico chamado John Graunt queria saber a população de Londres na época da peste negra. Em vez de contar as pessoas, ele inventou uma abordagem -que hoje chamamos de "estatística"-que lhe permitiu estimar a população.
1880	O Census Bureau dos Estados Unidos contratou o inventor Herman Hollerith, que desenvolveu um sistema de cartões perfurados e máquinas de tabulação para o censo de 1890. Ele conseguiu diminuir o tempo de tabulação do censo de oito anos para um.
1920	Descobertas da mecânica quântica abalaram para sempre o sonho das medições abrangentes e perfeitas.
1934	Jerzy Neyman, estatístico polonês, demonstrou que a abordagem da amostragem levava a erros e que para isto deveria ser considerada a aleatoriedade na escolha da amostra. A amostragem tornou-se a solução para o problema da profusão de dados.
1950-1960	Implementação de um sistema de folha de pagamento para General Electric Corporation, por Joe Glickauf e Arthur Andersen em um computador eletrônico digital. Início das ideias de "Revolução da informação" e "Era Digital" ou " Era da informação ".
1980-1990	Nesta época, Leonard Kleinrock, Lawrence Roberts, Robert Kahn, e Vint Cerf inventaram a internet. 40% de toda a potência computacional do mundo existia na forma de calculadoras de bolso. " Era da rede ".
2000	Dados: 75% dados analógicos (papel, filme, vinil, fitas magnéticas, livros, fotografia); 25% dados digitais.
2007	Dados: 7% dados analógicos; 93% dados digitais.
2010	Ciências como a astronomia e a genômica, vivenciaram uma explosão informacional, e cunharam o termo <i>big data</i> , que representa uma transição em termos de armazenamento e

	análise. " Era big data ".
2013	Dados: 2% dados analógicos; 98% dados digitais - 1200 <i>exabytes</i> . Volume de informações dobra a cada 3 anos.

Fonte: NESELLO; FACHINELLI, 2014, p. 23-24.

Com a observação do quadro, é possível perceber que no século XXI a conexão contínua da nossa sociedade produz e armazena uma imensa quantidade de dados. Assim, a atividade de pesquisa é intensificada pela quantidade de informações disponíveis predominantemente em suportes digitais (COSTA; CUNHA, 2014),

[...] Isto acontece essencialmente porque a tecnologia digital se torna cada vez mais um elemento onipresente nos processos da construção do conhecimento científico, seja por aumentar a capacidade dos instrumentos científicos, seja por reconstruir realidades por meio de simulação, ou ainda inaugurando formas inéditas de colaboração e compartilhamento de dados e informações. A capacidade dos computadores, que não para de crescer, aliada às possibilidades abertas pela Internet abrem novas aplicações para as fontes básicas de pesquisa – os dados de pesquisa – dando um novo impulso ao trabalho científico (SAYÃO; SALES, 2013, p.1).

Nos anos 2000, a explosão informacional iniciou o fenômeno *big data*. Este é um termo vasto que pode ser definido como dados em grande escala e compreende também o agregado de soluções tecnológicas para tratar desses dados. Alguns teóricos retratam-no como essencial na tomada de decisões, ou ainda, uma nova fonte de valor econômico e informação (COSTA; CUNHA, 2014; NESELLO; FACHINELLI, 2014).

Segundo Xexeo (2013), o *big data* retrata os obstáculos e as respectivas resoluções tecnológicas para dados que apresentam características de tratamento complexo. Numerosos teóricos chegaram à conclusão que as principais características desse fenômeno são volume, velocidade e variedade, conhecidas como os três “Vs”. No propósito de agregar maior conhecimento, Taurion (2012) indica outras características, como veracidade e valor.

As características, dispostas por Xexeo (2013) e Taurion (2012), são mais satisfatoriamente elucidadas abaixo:

- ✓ Volume: É uma das maiores complexidades do *big data*. Os sistemas atuais não estão preparados para comportar, e tratar, a escala de dados que já existe e os dados que irão existir daqui a algum tempo.
- ✓ Velocidade: Os dados são enviados para os sistemas com o índice de *bytes* por tempo tão alto que se torna inviável armazenar todos os dados produzidos.
- ✓ Variedade: O sistema é apropriado para processar dados em um determinado tipo de formato ou estrutura. Porém, esses novos dados aparecem de modos diversificados e alguns tipos possuem modelo livre ou estrutura própria.
- ✓ Veracidade: Conhecer a autenticidade e a procedência dos dados.
- ✓ Valor: *Feedback* às organizações que aplicaram recursos para implantação dos projetos de *big data*.

A priori, a expressão *Big Data* foi criada no mercado com o intuito de caracterizar a produção de um grande volume de dados, mas obteve um grande impacto no ambiente de pesquisa e, conseqüentemente, suscitou a apreensão no domínio da pesquisa científica em virtude da diversidade, da quantidade e da velocidade com que esses dados são disseminados (SAYÃO; SALES, 2014; PIMENTA, 2013).

O *big data* propicia o acesso de todas as formas de informação digital em tempo real, o que favorece a competitividade, a produtividade e a inovação na esfera científica. De acordo com Borba, Santos e Kawamoto Jr (2013), o *big data* é motivado pela habilidade de ampliação linear que viabiliza a análise de grandes conjuntos de dados. No entanto, deixam claro que o *big data* necessita de uma estrutura automatizada nas atividades de captação, ordenação, classificação e organização dessas variadas fontes de dados inerentes à pesquisa.

A metodologia do fazer científico também está mudando e a comunidade científica está no início de uma era singular em que a ciência é comandada por dados. Esse grande volume de dados é determinante para a origem do quarto paradigma – a *e-Science*. Esse novo paradigma tem por objetivo reunir teoria, experimento e simulação e, concomitantemente, explorar um grande volume de informação. A *e-Science* é capaz de analisar esses dados por meio do emprego de métodos e instrumentos direcionados para gestão e reutilização desses dados (SALES, 2014; COSTA; CUNHA 2013).

A *e-Science*, ou ciência orientada por dados como também é conhecida, permite um diálogo entre os cientistas da computação, da informação e especialistas de diferentes áreas com o intuito de desenvolver novos conceitos e teorias, partindo dos dados disponibilizados. Pode ser considerada o suporte que pesquisadores utilizam no progresso de diversos campos do conhecimento e no aperfeiçoamento de técnicas para o tratamento do vasto volume de dados primitivos criados. O método de constituir a ciência foi transformado e, deste modo, evidenciou-se conveniência de compartilhar esses dados produzidos e do trabalho em equipe. Com o objetivo de preservar esses dados, a ciência deve ser constituída de três etapas principais: captura, curadoria e análise (FERREIRA; SILVA, 2013; SAYÃO; SALES, 2012)

A *e-Science* emerge como um atalho para o fazer científico e tecnologias necessárias para um suporte colaborativo, uma ciência em rede em que a produção científica é reajustada por meio da intensidade do uso da rede e pelo uso dos dados disseminados. Esse campo multidisciplinar de pesquisa apresenta novas infraestruturas para armazenar, acessar e analisar grupos de dados em um contingente maior do que a gestão convencional. As tecnologias de informação em rede de apoio a atividades de pesquisa científica, bem como a distribuição de dados e publicação do produto final, possibilitam que os pesquisadores tenham acesso aos dados primitivos de modo remoto, capacitando-os a acessar, mover, manipular e extrair os dados (SALES, 2014; HEY; HEY, 2006; CORDEIRO, et al., 2013; SALES, 2014; PRADO; BARANAUSKAS, 2013; MEDEIROS; CAREGNATO, 2012)

Apesar dos dados gerados pela *e-Science* acarretarem grandes impactos sobre a ciência e, por isso, necessitarem de uma análise cuidadosa dos papéis das instituições dedicadas ao desenvolvimento da ciência e ao apoio aos pesquisadores, o movimento da *e-Science* é sustentado pela disponibilidade de uma grande quantidade de dados, de experimentos científicos e de pesquisas, que fornecem informações riquíssimas (COSTA; CUNHA 2013; Hey e Hey, 2006).

[...]e-Science altera fundamentalmente a maneira com que os cientistas realizam seu trabalho, as ferramentas que usam, os tipos de problemas que abordam e a natureza da documentação e da publicação que resulta da sua pesquisa. E-Science requer novas estratégias de suporte à pesquisa e significativo desenvolvimento de infra-estrutura (JOINT TASK FORCE ON LIBRARY SUPPORT FOR

E-SCIENCE, 2007, p. 6 apud MEDEIROS; CAREGNATO, 2012, p.315).

Essa nova metodologia de pesquisa nos quais dados experimentais são reunidos através de ferramentas ou constituídos por simulação, são processados por sistemas de *software* complexos, e só assim a informação decorrente é depositada em servidores. Lida com uma modificação profunda no desenvolvimento do pensamento científico, pois os dados só serão examinados no fim do processamento. Isto é, o processo de “formulação de hipótese \implies experimentação \implies análise de resultados” está sendo sucedido por “formulação de hipótese \implies busca da resposta no banco de dados”. Esse novo fazer científico, respaldado no processamento e exame de uma extraordinária quantidade de dados, preconiza tecnologias e métodos extremamente inovadores (CORDEIRO, et al., 2013).

Nesta fase da pesquisa, fez-se imprescindível a distinção de alguns conceitos. Para isso, far-se-á citação de autores conceituados na área com o propósito de sanar possíveis equívocos na terminologia:

É importante mencionar que na literatura de língua inglesa o conceito de *eScience* aparece também como *e-Research*, *cyberinfrastructure*, *cyberscience*, *enhanced science*, *data-driven-science*, entre outros nomes (CESAR JUNIOR, 2011, p.7). Alguns autores ressaltam determinadas nuances entre esses conceitos, outros consideram todos como sinônimos (SALES, 2014, p.44, grifo do autor)

Com o mesmo objetivo, Ferreira e Silva (2013, p. 3-4) evidencia o que há de particular nesses conceitos:

Diferente do termo *e-Science* que é relativamente novo, o termo “*cyberinfrastructure*”, mais usado nos Estados Unidos, tem mais tempo de aparecimento, sendo seu emprego inicial atribuído a Richard A. Clarke e Jeffrey Hunkeyr do *National Science Foundation / NSA*, em uma entrevista nos Estados Unidos em 1998. (CLARKE; HUNKER, 1998). Eles são identificados em alguns casos, como sinônimos e se referem ao uso de tecnologias de computação em rede para melhorar a colaboração e os métodos inovadores de investigação. A *e-Science* entretanto, tem um foco mais específico na pesquisa científica, ao passo que a *cybertinfrastructure* é mais abrangente em áreas fora das ciências e engenharia, como as ciências naturais e biológicas, e incorpora uma maior ênfase nos recursos de supercomputação e inovação. (ZEUDIE, 2012). São

também encontrados na literatura para descrever este fenômeno termos como, “ciência orientada por dados” (*data-driven science*); “computação fortemente orientada por dados” (*data-intensive computing*); “mineração de dados” (*data mining*); “quarto paradigma” (*fourth paradigm*); “dos dados ao conhecimento” (*from data to knowledge*); ciência com uso intensivo de dados (*data-intensive science*). Estes termos representam as mudanças em curso na ciência e resgatam termos mais antigos que foram percussores desse contexto atual como, “*Big Science*”, “*Cyberscience*” e “*e-Research*” e mais recentemente o “*Big Data*”. (CESAR JUNIOR, 2012, p. 7; JANKWSKI, 2009). (FERREIRA; SILVA, 2013, p. 3-4, grifo dos autores).

Segundo Medeiros e Caregnato (2012) é determinante constatar aspectos que tornam específicas as modificações no ambiente de pesquisa científica ao despontar da e-Science:

- ✓ Grupo de colaboração vasto e multidisciplinar;
- ✓ Reunião, representação e avaliação dos dados conduzidos pelo computador;
- ✓ Integração *end-to-end* – facilita a análise de uma parte que se acredita ser substancial ao estabelecimento e ao progresso da e-Science. Assim, é necessário ponderar o alicerce físico e tecnológico, como também o elemento humano e teórico que compõe esse curso.

A expectativa da comunidade acadêmica em torno da e-Science é justificável em razão do comprometimento desta em reduzir os esforços daquela nas atividades complexas, tornando-as mais simplórias. Na verdade, e-Science é uma designação das tecnologias de informação em rede que contribuem com as tarefas de pesquisa científica, como por exemplo: auxiliar na distribuição e publicação do produto final. Os dados podem perpassar entre as instituições sem a obrigatoriedade de ajustar os acordos entre estas nas quais as fases de computação precisam estar disponíveis através da solicitação e a visualização do conhecimento deve favorecer a compreensão e o estudo do saber científico. Esses são os tipos de automação, mais especificamente, trata-se de uma reorganização das tarefas entre humanos e recursos tecnológicos (MEDEIROS; CAREGNATO, 2012).

Com o propósito de intensificar as singularidades da e-Science, segue uma tabela com a aferição dos perfis da pesquisa tradicional e da e-Science:

Quadro 4: E-Science x Pesquisa Tradicional

Característica	e-Science	Pesquisa tradicional
Participantes	Diversamente qualificados, equipe de pesquisa distribuída	Pesquisador individual ou pequena equipe local de pesquisa
Dados	Gerados, armazenados e acessíveis de localizações distribuídas	Gerados, armazenados e acessíveis apenas localmente
Computação e Instrumentação	Larga escala ou sob demanda ou acesso à informação compartilhada	Emprego da computação em lote ou emprego do computador ou instrumentos do próprio pesquisador
Rede	Confiança na internet e em mediadores	Não confiante na internet
Disseminação da pesquisa	Via websites e portais especializados	Via publicações impressas ou apresentações em conferências

Fonte: MEDEIROS; CAREGNATO, 2012, p. 316 – 317

Mediante a análise do quadro, é possível certificar que as atividades da *e-Science* são particularizadas por sua capacidade de suportar essas exigências, como: o grupo de dados utilizados e os métodos de análise de dados devem estar acessíveis; o grupo de dados utilizados deve ser documentados com metadados; já os métodos de análise de dados devem ser somente documentados e; o grupo de dados utilizados deve ser devidamente referenciado em uma publicação, uma vez que permite que os dados sejam capturados em um ambiente, armazenados em outro, e processados em um terceiro. A *e-Science* é apoiada por uma rede de computação que fornece redes de alta velocidade, infraestruturas que controlam o acesso e o pagamento de serviços e dados e apresenta uma descrição dos dados e serviços. (ERBACH, 2006).

Na próxima seção tratará sobre a infraestrutura na qual a *e-Science* se expande continuamente, a ciberinfraestrutura.

3.3 CIBERINFRAESTRUTURA

Como enfatizado desde a introdução da dissertação, os avanços tecnológicos transformam continuamente a pesquisa científica. Por este motivo, várias tendências convergem na aceleração dos limites de maneira que demonstram a expectativa de evolução na maneira de criar, divulgar e preservar o conhecimento científico. Dados, informações e as redes estão sendo utilizados para compartilhar e aumentar os esforços tradicionais na ciência e na pesquisa, ou seja, na verdade para criar novas disciplinas. As duas abordagens clássicas para pesquisa científica, teórica / analítica e experimental / observacional, ocorreram para examinar um maior número de possibilidades em novos níveis de tempo e fidelidade espacial. As redes avançadas permitem que pessoas, ferramentas e informações possam ser ligadas de maneira a diminuir os empecilhos de localização, tempo, instituição e disciplina. Em muitos novos campos são distribuídos conhecimentos em ambientes que estão se tornando essenciais para dar prosseguimento a novas etapas das pesquisas. Portanto, pesquisadores que estão em níveis avançados, tanto na produção quanto na exploração de pesquisa, percebem um novo modelo de estrutura de pesquisa em evolução (ATKINS, 2003).

A ciberinfraestrutura desponta, baseando-se em novos recursos computacionais, versando sobre novas maneiras de organizar a prática da ciência e, com isso, permitindo novas formas de colaboração, de organização, de pesquisa e de aprendizagem. Tal infraestrutura aborda como obter, compartilhar, armazenar e aproveitar os dados para os principais produtos produzidos pela ciência. E, mediante a divergência sobre a concepção de dado, a ciberinfraestrutura não é um conceito singular ou unificado (JACKSON et al., 2007).

É possível afirmar que o termo ciberinfraestrutura transmite o conceito de distribuição em uma rede integrada, ainda é capaz de incluir a totalidade da infraestrutura digital e do papel crítico dos curadores digitais em sua manutenção. É denominada também como a infraestrutura indispensável para capitalizar os grandes avanços na tecnologia da informação (Atkins et al., 2003). A ciberinfraestrutura é capaz de realizar a integralização do hardware para computação, redes de dados e sensores habilitados de maneira digital, observatórios e instalações experimentais, e um grupo de softwares interoperáveis e serviços e ferramentas de middleware (NTF, 2007).

De acordo com Sayão e Sales (2015a), este conceito de ciberinfraestrutura aparece para caracterizar a estrutura tecnológica oculta. Este termo remete à infraestrutura alicerçada em computadores interligados, informações e tecnologia de comunicação. Essa estrutura está entre duas camadas: uma subjacente e outra superior. As tecnologias inferiores são os elementos da computação, armazenamento integrado electro-ópticos, e comunicação. Já os componentes do superior à ciberinfraestrutura são programas de software, serviços, instrumentos, dados, informação, conhecimento e práticas sociais aplicáveis aos projetos específicos, disciplinas e comunidades de prática. A ciberinfraestrutura é a camada que habilita o hardware, os algoritmos, o software, as comunicações, as instituições e as pessoas. Esta camada tem o objetivo de proporcionar um site eficaz e eficiente, no intuito do fortalecimento das comunidades específicas dos pesquisadores e da inovação e revolução de sua produção científica e de quem participa. (ATKINS, 2003). Em síntese, a ciberinfraestrutura é composta por três camadas principais interligadas: duas de rede (uma física e outra de lógica) e uma social. (GOMES, 2013)

É possível certificar que, no âmbito da pesquisa científica, esse termo:

[...] descreve ambientes de pesquisa que dão suporte a métodos avançados de aquisição, armazenamento, curadoria, mineração, visualização de dados, além, de outros serviços de computação e processamento de informação distribuídos em rede. Esses ambientes tecnologicamente avançados criam pontos de inflexão na trajetória histórica da ciência, redesenhando seus métodos, suas formas de socialização e seus resultados, tornando os dados elementos determinantes (SAYÃO; SALES, 2015a, p. 17)

Ray (2012) afirma que esse termo foi utilizado pela primeira vez em 2003, em um relatório de pesquisa científica produzido pela National Science Foundation (NSF), liderado por Dan Atkins. Este documento, posteriormente, passou a ser denominado de “Relatório Atkins” e já previa a importância da ciberinfraestrutura ao mencionar que esta seria tão fundamental e importante tal como um facilitador para uma empresa, tão fundamental como sala de aula, e fundamental como o sistema de conferências e revistas de divulgação dos resultados de pesquisa.

É importante enfatizar que o termo ciberinfraestrutura é um termo aplicado nos EUA proporcional à e-infraestrutura na União Europeia e à *e-Science* no Reino Unido (Atkins et al., 2003; GOMES, 2013).

A próxima seção se dedicará à gestão da informação do conhecimento no intuito de demonstrar a importância de tais para a pesquisa científica e enquadrar o tema da dissertação na linha de pesquisa proposta - organização e representação do conhecimento.

4 GESTÃO DA INFORMAÇÃO E DO CONHECIMENTO

Esta seção tratará sobre os conceitos da gestão da informação e gestão do conhecimento que estão correlacionadas com a organização do conhecimento; por isso, para a comunicação científica e para a pesquisa científica são avaliados como imprescindíveis. Para prosseguir nessa linha de pensamento, é de suma importância que se dê continuidade a esta dissertação, a menção de alguns pressupostos da gestão da informação e gestão do conhecimento.

Na comunidade científica, a informação e o conhecimento são imprescindíveis para as instituições de pesquisa. A informação e o conhecimento operam como diretrizes com a finalidade de que as instituições sejam eficazes em sua missão. É indispensável conduzir a instituição ao alcance dos patamares da gestão informacional e da gestão do conhecimento que correspondam às expectativas de crescimento e de concorrência.

Faz-se necessário que os gestores administrem da melhor forma seus pontos positivos com o propósito de se manterem participantes do ambiente científico (AMORIM; TOMAEL, 2011). Assim, os gestores da informação procuram estabelecer técnicas sempre atualizadas de gestão da informação, independente do ambiente. Lopes e Valentim (2013) advogam que a informação é a matéria-prima para a consumação das práticas institucionais e, por isso, requer gestão no intuito de assistir ao consumidor final.

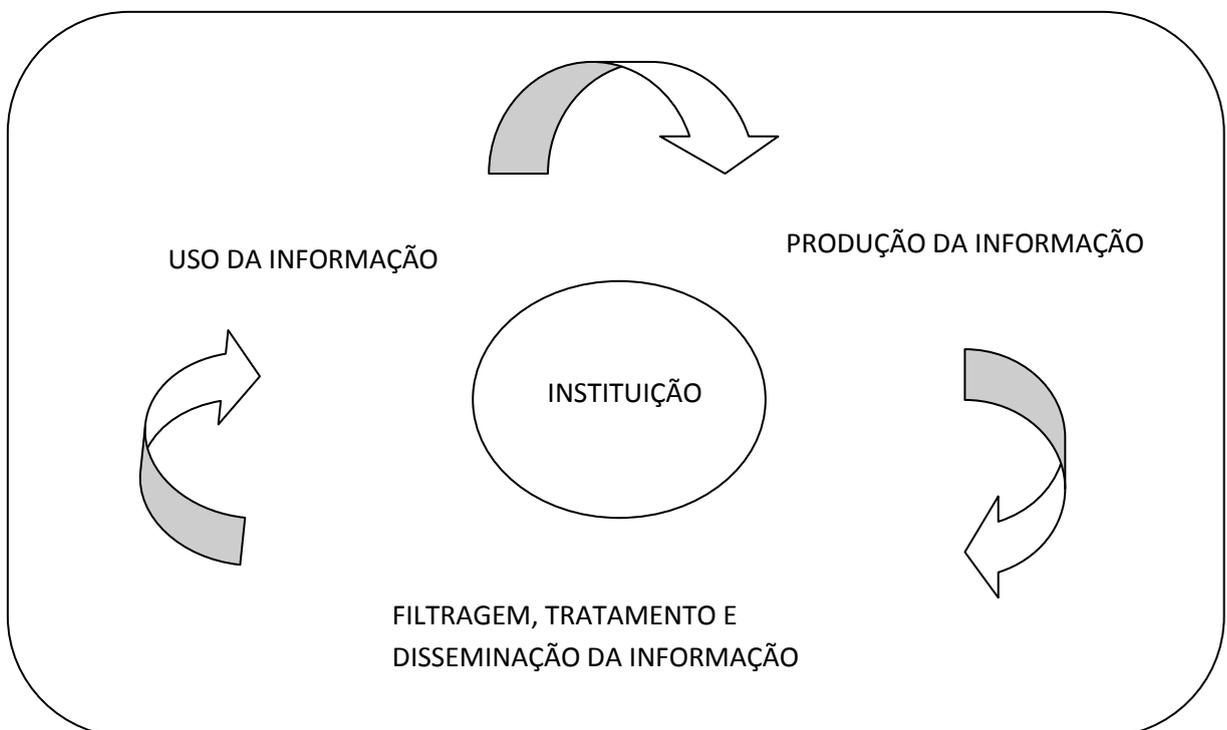
Como benefício, o processo de gestão da informação assegura metodologias que objetivam a eficiência na difusão e recuperação de informações. Desta maneira, a gestão da informação é considerada como um recurso para a procura pela inovação. Através dessa gestão é possível o monitoramento do volume de informações, de maneira a ponderar a diversidade progressiva de decisão no intuito de aumentar o fluxo, gerência, estudo e sumarização da informação para pesquisadores (LOPES, VALENTIM, 2013; AMORIM; TOMAEL, 2011).

Os principais integrantes que financiam o fazer científico são as agências de fomento consideradas essenciais para que haja esse canal de comunicação. Sobretudo para que haja busca, reunião, estudo, exame e disseminação das informações que serão utilizadas pelos pesquisadores e estudiosos. Nesse sentido, ressalta-se que é mediante a análise da informação que os pesquisadores produzem o conhecimento para intervir no ambiente de pesquisa. “Essa situação [análise da

informação] se configura como uma forma de gestão estratégica da informação, realizada a partir da necessidade de cada agente, cujo objetivo é criar conhecimento” (LOPES, VALENTIM, 2013, p.159).

A figura abaixo representa a informação durante a mediação da informação. Com destaque para a produção, recuperação, tratamento, disseminação e uso da informação e seus principais responsáveis envolvidos nesse seguimento:

Figura 1: Fluxo da informação



Fonte: Baseado em LOPES, VALENTIM, 2013.

No panorama da comunicação científica, toda instituição de pesquisa é vista como uma organização processadora de informação, à proporção que os fluxos informacionais fazem parte dos processos inerentes à geração de qualquer produto ou serviço produzido. Assim, a informação é julgada como essencial para as operações realizadas em uma instituição; entretanto, pondera que é elementar a existência do entendimento dos procedimentos e da movimentação que convertem a informação em conhecimento e que, por consequência, se transforma em uma atividade. Por isso, a instituição deve se atentar às perspectivas que mediam a

gestão da informação, em especial a produção da informação em si; desta maneira, é possível a determinação da melhor maneira de efetivar as estratégias e metas que se almeja atingir. Analisar as perspectivas que impulsionam a produção do conhecimento através da gestão de conteúdos informacionais, permite que a instituição sobressaia na comunidade científica (LOPES, VALENTIM, 2013).

Isto posto, é viável perceber que a construção da informação deve se tornar operacional mediante métodos de práticas bem definidos, obviamente, baseando-se em determinado procedimento de mudança tomando por norte uma racionalidade técnica, desempenhando, assim, atividades que são associadas ao agrupamento, à seleção, codificação, diminuição, organização e conservação de informação. Desta maneira, tais procedimentos devem ser conduzidos para a organização e controle de depósito da informação, quer seja para utilização instantânea, quer seja para utilização. Tal suporte informacional é o que retrata o conhecimento substancial no campo da troca de informações (LOPES, VALENTIM, 2013).

A gestão da informação, especificamente, envolve os estudos e as práticas gerenciais que permitem a construção, a disseminação e o uso da informação. Esse processo engloba a gestão de recursos informacionais e de conteúdos, a gestão de tecnologias da informação e a gestão das pessoas envolvidas nesses sub-processos. Esse entendimento se encontra fundamentado em Barbosa (2008) para quem a gestão da informação e a gestão do conhecimento focam dois importantes fenômenos organizacionais: o primeiro corresponde à informação ou ao conhecimento registrado e o segundo, por sua vez, refere-se ao conhecimento pessoal, muitas vezes, tácito, que precisa ser descoberto e socializado para ser efetivamente utilizado. Nesse conjunto, a Ciência da Informação se apresenta como base estrutural porque, na qualidade de ciência social aplicada, dedica-se, ao mesmo tempo, ao estudo das propriedades gerais e das condições da informação, e aos processos que possibilitam seu processamento, sua disponibilização e seu uso efetivo. A Ciência da Informação, diferentemente das demais disciplinas, ciências ou áreas do conhecimento que se dedicam, de alguma forma, ao estudo da informação, tais como Administração, Ciência da Computação, Comunicação, Engenharia de Produção, deve-se dedicar aos fundamentos teóricos do fenômeno informacional e, concomitantemente, aos processos e às práticas que possibilitam o fluxo informacional e o uso da informação. (SOUZA, DIAS, NASIF, 2011, p.59)

É interessante mencionar que a gestão da informação foi desenvolvida antes da gestão do conhecimento e, de maneira alguma, deve ser vista como uma de suas fases, pois, na verdade, é sua antecessora, “uma vez que seu foco é a informação, e

porque não se extinguiu com o surgimento da gestão do conhecimento”. (BETTENCOURT; CIANCONI, 2012, p.17). Estas estão relacionadas, mas possuem as seguintes disparidades: a gestão da informação foca nos fluxos formais da organização; isso que dizer, evidencia o conhecimento registrado, o que está sistematizado em qualquer tipo de suporte. Já a gestão do conhecimento evidencia o conhecimento do indivíduo somado ao conhecimento construído mediante a socialização e supre a gestão da informação que fornece o feedback à gestão do conhecimento. Isto tudo ocorre em um processo cíclico (POREM; GUARALDO, 2013; VALENTIM, 2008).

Logo, a gestão do conhecimento aborda a idealização e a administração de procedimentos que conduzem o fluxo do conhecimento, tanto em seu caráter explícito em que reúne práticas da gestão da informação, quanto em sua vertente tácita. A idealização e a administração dos procedimentos implicam no diagnóstico, obtenção, preservação, disseminação, produção e consumo do conhecimento tácito e explícito com o propósito de potencializar os métodos de organização em qualquer circunstância. Essa sequência só é possível em razão da comunicação; por isso, os questionamentos acerca da gestão do conhecimento e do conhecimento em si sugestionam e admitem intervenções da comunicação científica. (LEITE; COSTA, 2007)

Sob essa perspectiva, os autores Leite e Costa (2007) afirmam que a gestão do conhecimento relaciona técnicas e métodos da gestão da informação enquanto se relaciona aos mecanismos de captação, guarda, recuperação de um fragmento do conhecimento tácito que nesta fase é reduzido a suporte da informação. Entretanto, o conceito de gestão do conhecimento não deve ser tomado como gestão da informação.

Vale ressaltar que a comunicação científica é elemento essencial para o andamento da gestão do conhecimento científico. Inclusive, o ato de registrar é a atribuição substancial da comunicação científica. Isto é, a gestão do conhecimento e a comunicação científica são relacionadas por meio do vínculo que apresentam com o registro da informação. Por outro lado, no que tange à diferença conceitual do termo em questão, pode-se encontrar na literatura autores que divergem com os teóricos antepostos, pois certificam que o conhecimento uma vez registrado, torna-se informação (LEITE; COSTA, 2007).

Após a menção da importância da gestão da informação e do conhecimento para as instituições de maneira geral, a seguir será apresentada a seção na qual o tema proposto será de fato discutido – sistemas de gestão de informações sobre pesquisa, onde serão tratados o histórico, os conceitos de bases de dados, repositórios digitais, dados de pesquisa, repositório de dados de pesquisa, curadoria digital, CRIS/RIM, seus tipos e respectivos padrões e, por fim, o papel das bibliotecas em tal contexto.

5 BASES DE DADOS, REPOSITÓRIOS E BIBLIOTECAS DIGITAIS

Essa seção envolverá sistemas que também fazem parte do contexto da pesquisa científica e, ao mesmo tempo, analisará questões referentes aos dados de pesquisa. Serão abordadas as bases de dados, repositórios digitais, bibliotecas digitais, dados de pesquisa, repositório de dados de pesquisa e curadoria digital.

5.1 BASES DE DADOS, REPOSITÓRIOS DIGITAIS, BIBLIOTECAS DIGITAIS

A internet revoluciona a maneira de se comunicar, ao mesmo tempo em que inova no acesso aos sistemas de informação. Com o advento da internet, o acesso ao universo digital informacional se popularizou. E, no intuito de recuperar essas informações de maneira mais ágil, organizada e eficaz as bases de dados se difundiram.

Bases de dados podem ser definidas como um conjunto de itens reunidos por alguma característica em comum que podem ser recuperados no intuito de informar o usuário sobre determinado assunto e que são preservados eletronicamente.

A estrutura padronizada, congregando tipologias de bases de dados bem definidas, possibilita que os sistemas de informações gerenciem uma grande quantidade de bases de dados de diferentes tipos e, ao mesmo tempo, disponibiliza ao usuário um serviço de qualidade quanto ao acesso e quantidade de informações (VALENTIM, 2001, p. 70).

Existem alguns tipos de bases de dados (ROWLEY, 2002):

Bases de dados referenciais – são fontes secundárias de informação:

- ✓ Bases de Dados Bibliográficos;
- ✓ Bases de Dados Catalográficos;
- ✓ Bases de Diretórios ou Referenciais.

Bases de dados de fontes - contém os dados originais:

- ✓ Bases de Dados Numéricos;
- ✓ Bases de Dados de Texto Completo;
- ✓ Bases de Dados Textuais e Numéricos;
- ✓ Bases de Dados Gráficos ou Multimídia.

Originado pela carência de estruturas e fluxos de comunicação mais satisfatórios entre os pesquisadores emergem os repositórios digitais como mecanismos para a difusão e facilitação da comunicação científica.

Os repositórios tiveram origem na era digital e podem ser divididos em três grupos (RODRIGUES, et. al., 2011):

- ✓ Os repositórios que armazenam conteúdo no local;
- ✓ Os repositórios que armazenam somente metadados;
- ✓ Os repositórios híbridos que armazenam conteúdo e metadados no local.

Os repositórios digitais, em sua grande maioria, guardam e disseminam materiais revisados, garantindo assim mais crédito aos materiais que são acessados pelos usuários. Os repositórios institucionais que também são digitais abrigam dissertações, teses e outros tipos de pesquisas; dados científicos estruturados ou não; periódicos; documentos administrativos; memória cultural, artística e histórica da instituição; neste caso, agem como arquivo também. Existe, ainda, outro grupo de repositórios de objetos informacionais que é exibido em espaços abertos ao público como Youtube, SlideShare dentre outros, que da mesma maneira dos referidos anteriormente estabelecem limites em sua abrangência a um único tipo de recurso (MARCHIORI, 2012).

Os repositórios podem ser categorizados de acordo com o propósito para o qual foi criado. Os repositórios são classificados em Repositórios Institucionais – criados com o intuito de gerenciar a produção científica da instituição, - e Repositórios Temáticos, que reúnem a produção técnico-científica sobre um tema, independente da instituição à qual o autor é vinculado. Outra classificação relevante quanto ao tipo de material, os repositórios de e-prints tratam da versão eletrônica de

uma publicação científica e os repositórios de dados tratam de dados de pesquisa que serão tratados com mais cautela a seguir (SALES, 2014).

Os repositórios institucionais e os repositórios temáticos incluem um novo conceito de acesso livre, associados à tecnologia de *open archives* são usados pelas comunidades científicas na disseminação dos resultados das pesquisas, assim como na potencialização do seu impacto. Então, criam-se mecanismos para ratificar e incentivar a disseminação de pesquisas produzidas (MORENO, LEITE, ARELLANO, 2012).

[...] as bibliotecas digitais, os repositórios institucionais e os repositórios públicos da *web* oferecem itens/peças de informação que são razoavelmente “estáticas”. Isto é, logo de sua armazenagem se tornam disponíveis para recuperação e uso e não são mais submetidas a mudanças de conteúdo ou formato (MARCHIORI, 2012, p.14-15)

Já a Biblioteca digital é capaz de reunir versões eletrônicas da literatura já publicada e são disponibilizadas através de uma rede de computadores conectados, armazenam artefatos de informação únicos e de características únicas no mínimo em duas perspectivas: a tipológica e os critérios de busca. Entretanto, alguns autores evidenciam que as bibliotecas digitais ou virtuais possuem conexões frágeis com as carências dos pesquisadores no âmbito de seus anseios informacionais (MARCHIORI, 2012).

Contudo, é importante acentuar que a biblioteca digital não é uma coleção digitalizada favorecida por instrumentos de gestão da informação. Na verdade, essa ideia de biblioteca digital foi substituída por uma concepção mais complexa. Atualmente, o conceito de biblioteca digital remete a um meio que agrega coleções, serviços e pessoas para apoiar o ciclo de vida de produção, disseminação, uso e preservação de dados, informação e conhecimento (SAYÃO, 2008-2009).

As bibliotecas digitais são o complemento das bibliotecas tradicionais e são compostas por objetos eletrônicos e objetos impressos e outros tipos de objetos, como áudio, vídeo etc. A biblioteca digital dá acesso à informação satisfazendo a ideia de onipresença, a qualquer hora e em qualquer lugar, o que não pode ser confundido como se ela só indicasse para a informação. Tal tipo de biblioteca possui uma estrutura de organização consolidada com pontos sólidos para acessar os dados e ainda é capaz de oferecer acesso não só ao seu acervo, como também,

acesso ao acervo de outras bibliotecas digitais. Desta maneira, dá suporte ao acesso à informação de forma rápida e eficiente em uma grande extensão de fontes de informações e são conectados por links. Ela abarca os processos e trabalhos oferecidos pelas bibliotecas tradicionais, ainda que esses métodos precisem ser verificados no intuito de dispor diferenças entre mídias digitais e impressas (SAYÃO, 2008-2009).

Baseado nesse estudo, é possível afirmar que:

“Bibliotecas digitais são organizações que disponibilizam os recursos, incluindo pessoal especializado, para selecionar, estruturar, oferecer acesso intelectual, interpretar, distribuir, preservar a integridade e assegurar a persistência ao longo do tempo de coleções de trabalhos digitais, de forma que eles estejam pronta e economicamente disponíveis para uso de uma comunidade definida ou um conjunto de comunidades” (SAYÃO, 2008-2009, p.15)

Por meio dessas bases de dados, um enorme volume de dados foi criado e precisa ser tratado e disseminado. À vista disso, emergiram novos conceitos que norteiam a organização e a disseminação desses dados que serão apresentados a seguir.

5.2 DADOS DE PESQUISA, REPOSITÓRIO DE DADOS DE PESQUISA, CURADORIA DIGITAL

É importante ressaltar que a visão centrada em dados de pesquisa não é apenas um ponto de vista de informações sobre a pesquisa, mas sim, a apreciação de um status especial devido à supremacia dos dados em pesquisas científicas. Afinal, o esforço científico se dedica à análise, interpretação, modelagem e compreensão dos grupos de dados observáveis (ERBACH, 2006).

Alguns dados ocorrem naturalmente (clima, eventos astronômicos); outros são os resultados incidentais de atividade humana (dados sócio-econômicos), e outros ainda são produzidos no interesse da investigação científica (dados experimentais). Independentemente da origem dos dados, ele deve ser documentada através de metadados com uma explicação de quando, onde e como eles foram capturados, com o qual os instrumentos, por quem, e como os dados são codificados. Tais metadados são aplicáveis a conjuntos

inteiros de dados, bem como os pontos de dados individuais (tempo e localização). Frequentemente, os dados não são capturados e representados de modo neutro teoricamente, mas a codificação dos dados são baseadas em certas suposições teóricas e os interesses dos pesquisadores (ERBACH, 2006, p. 220, tradução nossa).

Os dados de pesquisa são registros de fatos, quer sejam números, documentos textuais, imagens ou sons, usados como fontes primárias para a pesquisa científica, e comumente aceitos na comunidade científica para validação de resultados da pesquisa. Um conjunto de dados de pesquisa constitui uma representação parcial sistemática do assunto a ser investigado (OECD, 2007).

Esses dados são objetos complexos, por isso seus critérios de compartilhamento, propósitos e métodos variam de acordo com o campo científico no qual estão inseridos (BORGMAN, 2010). Apresentando particularidades que implicam na forma de arquivar e preservar, sendo classificados como:

- ✓ Observacionais – Reunidos mediante análise empírica e, “devem ser difíceis ou impossíveis de se obter novamente por serem próprios de um momento ou circunstâncias irreplicáveis de uma forma exatamente igual” (TORRES-SALINAS, ROBINSON-GARCÍA E CABEZAS-CLAVIJO, 2012, p. 175 apud MEDEIROS; CAREGNATO, 2012, p.318).
- ✓ Computacionais – Estabelecidos através de simulação ou modelos / modelagens de computadores e;
- ✓ Experimentais – Mensurados por intermédio de medições de experimentos laboratoriais (NATIONAL SCIENCE BOARD, 2005 apud MEDEIROS; CAREGNATO, 2012).

Com relação a sua estrutura podem ainda ser classificados como:

- ✓ Analógicos ou digitais;
- ✓ Livros de laboratórios ou software;
- ✓ Criados na forma digital (“born digital”);
- ✓ Convertidos para o formato digital (digitised”).

À vista disso, Sayão e Sales (2014) afirmam que as informações contextuais (semânticas e estruturais) devem acompanhar os dados digitais, a fim de que estes se tornem autodescritos. Assim, os metadados registram o que é interesse do pesquisador.

Esse verdadeiro dilúvio de dados vem sendo desencadeado principalmente pelo avanço extraordinário de instrumentos, sensores e escalas, que aumentaram exponencialmente a capacidade de obtenção de dados pela realização de observações e medições de fenômenos, somados às informações geradas artificialmente por simulações e por software (SAYÃO; SALES, 2012, p. 181).

O crescimento do fluxo de geração de dados de pesquisa científica tem agradado a comunidade científica, pois, por meio desses dados é possível obter uma melhor compreensão dos problemas científicos (BORGMAN, 2010). E, “providenciam as evidências necessárias para conferir veracidade, autenticidade e capacidade de reprodutibilidade ao corpo de conhecimento publicado nos periódicos, [...]” (SAYÃO; SALES, 2014, p. 80). Diante do exposto, cientistas de todo o mundo discutem a ampliação do acesso aos dados da pesquisa, uma vez que a ciência é internacional e, conseqüentemente, sua cooperação mundial é fundamental, tornando o acesso aos dados de pesquisa imprescindível. Em todo o mundo, as organizações que financiam a pesquisa científica estão trabalhando com a conscientização sobre a importância e a usabilidade dos dados da pesquisa (BRASE; FARQUHAR, 2011).

Nesta direção, tornou-se necessário um mecanismo que gerisse todos esses dados, que é chamado de Curadoria Digital.

Para Abbott (2014), a Curadoria Digital surge para gerar e preservar os dados digitais em longo prazo e envolve a gestão de conjuntos de dados para o uso diário. Sendo assim, no futuro, atividades que se tratam de gerenciamento de dados, de planejamento para sua criação, melhores práticas em digitalização e documentação, farão parte da curadoria digital.

Segundo Sayão e Sales (2014), curadoria digital é um conceito em evolução, que abrange o conhecimento e as práticas em preservação e acesso a recursos digitais adquiridos na última década. “Curadoria digital estende-se além do controle do repositório que arquiva os recursos e envolve a atenção do criador do conteúdo e

dos usuários futuros” (ABBOTT, 2008 apud SAYÃO E SALES, 2014, p. 14). Vale lembrar que, atrelados ao conceito de curadoria digital estão os processos de arquivamento digital, de preservação digital, curadoria de dados, gerenciamento de registros eletrônicos, e gerenciamento de ativos digitais. No intuito de evitar confusões entre alguns desses termos, conceituá-los e hierarquizá-los em suas devidas posições, Yakel (2007) cita o Relatório de Curadoria de e-Science¹, onde faz a diferenciação dos conceitos de curadoria, arquivo e preservação.

Com base nesse material, criou-se o esquema abaixo que conceitua curadoria, arquivamento e preservação e demonstra as respectivas atividades de arquivamento e preservação dentro do contexto de curadoria digital.

Quadro 5: Curadoria, Arquivamento e Preservação

Curadoria - Atividade que se responsabiliza pela gestão e incentivo do uso de dados, partindo de sua criação. Garantindo que este dado esteja disponível para reutilização e acesso. Para conjuntos de dados dinâmicos, isso significa contínuo enriquecimento ou atualização, mantendo-o adequado à sua finalidade.

Arquivamento - Uma atividade de curadoria que garante que os dados sejam selecionados e armazenados. E que mantém a integridade física, segurança, autenticidade dos dados ao longo do tempo, podendo esses ser acessados.

Preservação - Uma atividade de arquivo em que os itens específicos de dados são mantidos ao longo do tempo, de modo que eles ainda possam ser acessados e compreendidos por meio da mudança sucessiva e obsolescência de tecnologias.

Fonte: A autora

Neste mesmo artigo, Yakel (2007) revela que vários conceitos surgiram a partir dessas definições como: gestão do ciclo de vida, envolvimento ativo ao longo do

¹ Disponível em: < www.jisc.ac.uk/publications/publications/pub_escience.aspx >

tempo dos criadores de registros e potencialmente dos curadores digitais, avaliação e seleção de materiais, desenvolvimento e fornecimento de acesso e a garantia da preservação dos objetos, no âmbito da usabilidade e da acessibilidade. Chegando a idealizar curadoria como “um termo abrangente para a preservação digital, curadoria de dados, e de ativos e registros eletrônicos de gerenciamento digital” (YAKEL, 207, p. 338)

Para Sayão e Sales (2012), o maior desafio da curadoria digital é manter no dado a capacidade de transmitir conhecimento no futuro, permitindo que os novos pesquisadores analisem novamente estes dados dentro de novas pesquisas. Para isso, é necessário que as “informações estejam estruturadas por modelos de informação e traduzidas por esquemas de metadados.” (SAYÃO e SALES, 2012, p. 180)

A curadoria digital possui benefícios e valores, como:

- ✓ O acesso persistente para dados digitais confiáveis;
- ✓ A melhoria da qualidade dos dados em si e seu contexto de pesquisa;
- ✓ A utilização de normas comuns em diferentes conjuntos de dados;
- ✓ As verificações de autenticidade;
- ✓ Avaliação de dados seguros como um registro formal;
- ✓ A exploração de investimento inicial;
- ✓ A maior velocidade e amplitude de acesso, compartilhamento de dados e análise de oportunidades;
- ✓ Preservação de dados e proteção contra perda e obsolescência;
- ✓ Permissão de acesso contínuo aos dados;
- ✓ Incentivo à reutilização de dados;
- ✓ Maximização da exploração de materiais digitais cumulativamente ao longo do tempo;
- ✓ Fornecimento de informações sobre o contexto e proveniência dos dados;
- ✓ O uso de ferramentas e serviços para migração de dados, metadados e outras informações
- ✓ Representação em novos formatos para assegurar que a informação permaneça significativa para os usuários;
- ✓ A infraestrutura de gestão para a preservação e difusão de dados em perpetuidade (ABBOTT, 2014).

Os dados de pesquisas são tão importantes quanto qualquer recurso informacional, visto que “todos eles são resultados que precisam ser considerados como parte da infraestrutura mundial de informação científica” (SAYÃO e SALES, 2014, p.3).

O acesso aos dados de pesquisa possui reflexos globais, visto que os pesquisadores utilizam dados cooperativos internacionais. Entretanto, esses pesquisadores estão inseridos em contextos, estruturas políticas e organizacionais diferentes. Essa transferência de dados permite a verificação da confiabilidade dos resultados das pesquisas e permite que novas descobertas sejam desenvolvidas através de informações pesquisadas anteriormente. Desta forma é possível que se encurte “o ciclo clássico de comunicação científica e, [que sejam desenvolvidas] novas formas de interlocução e de socialização no mundo científico, além de contribuir para a racionalização dos recursos financeiros públicos aplicados na pesquisa científica” (SAYÃO e SALES, 2014, p.3).

“A curadoria digital envolve a manutenção, a preservação e a agregação de valor a dados de pesquisa durante o seu ciclo de vida” (Data Curator Center apud SAYÃO E SALES, 2014, p. 14). Comumente, os dados tinham somente sua estrutura final levada em conta e, muitas vezes, eram descartados. E o mais importante fator para sua deterioração era o tempo. Entretanto, uma boa gestão, minimiza o risco do dado se tornar obsoleto. Outro fator interessante é o de que ao reutilizar o dado em novos trabalhos, aumenta sua vida útil e implica na rapidez da nova pesquisa realizada. Lembrando que os dados tratados pela curadoria digital fornecem seguridade, precisão, integridade e autenticidade.

Vale lembrar que curadoria digital é um processo contínuo e em longo prazo, consistindo em uma série de atividades. Ou seja, curadoria digital não é um back up (ABBOTT, 2014).

No novo ambiente de pesquisa redesenhado pelas práticas da *e-science*, o ciclo de vida da curadoria digital incorpora-se como uma peça-chave no fluxo tradicional de comunicação científica baseado tradicionalmente em artigos de periódicos. A curadoria digital, no momento em que gerencia e preserva os dados de pesquisa para que sejam acessados e compreendidos por outros pesquisadores estabelecendo um diálogo com o futuro, cria a possibilidade de se criar conceitos inovadores de documentos de registros de pesquisa, rompendo com o paradigma unidimensional e absoluto do artigo de periódico. (SAYÃO E SALES, 2014, p. 17-18)

Visto que as áreas de pesquisa possuem normas diferentes para inserção de dados, é preciso a criação de estruturas técnicas, gerenciais e sociais que unam os diferentes conjuntos de dados e, sendo criados assim, canais de colaboração de dados. (SAYÃO E SALES, 2014). Desse modo, várias organizações começam a perceber a importância do trato dos referidos dados, pois toda essa informação poderá ser reusada e compartilhada, caso haja um padrão entre estas.

Outra questão que nos inspira cuidado é o acesso aberto aos dados científicos, pois estas informações, além de constituírem grandes descobertas dos pesquisadores, tratam dos dados abertos que facilitam a conferência da veracidade dos trabalhos publicados, evitando até possível duplicação dos mesmos.

Portanto, quanto maior a capacidade dos sistemas de informação de oferecer dados de pesquisa livremente e tratado por metadados de forma que possam ser interpretados e reutilizados pelo maior número possível de pesquisadores de diversas áreas, maior será o grau de transparência, de reprodutibilidade e de eficiência do processo de geração de conhecimento científico, e maior será a amplitude de aplicação dos projetos de pesquisa para a sociedade (SAYÃO; SALES, 2014, p. 7)

O reuso dos dados de pesquisa se caracteriza como um desafio para a curadoria digital e para e-science. Desta forma, a curadoria de dados surge com o intuito de que os dados mantenham o seu dever de repassar a informação desejada. Devido à grande variedade de qualidade de dados, deve ser feita uma política de curadoria de dados ampla, em que todas ou quase todas essas qualidades de dados científicos sejam reunidas.

Portanto, verifica-se um deslocamento no padrão de arquivamento estático e inacessível promovido pelos *dark archives*, repositórios de acesso restrito voltados para garantir integridade e autenticidade. O foco da curadoria digital está na gestão por todo o ciclo de vida do material digital, de forma que ela permaneça continuamente acessível e possa ser recuperado por quem dele precise. Ampliando a capacidade dos dados serem recuperados e acessados estão os modelos de informação, expressos por metadados; além do mais, os metadados são também ferramentas importantes para os procedimentos de controle de autenticação (SAYÃO e SALES, 2012, p. 184)

O compartilhamento e a reutilização dos dados que proporcionam novas análises das informações valorizam e protegem o capital utilizado na pesquisa primária. Mas, infelizmente, as poucas instituições envolvidas nesse projeto de gestão desses dados não estabeleceram padrões e não conseguiram financiamento que assegurem o completo sucesso do projeto.

Uma base de dados que tem em vista organizar e publicar esses dados de maneira ordenada, com o propósito inclusive de reutilizá-los em futuras pesquisas é o Repositório de Dados.

Os repositórios de dados são bancos de dados digitais que possibilitam acesso dos usuários aos resultados das pesquisas, priorizando o acesso aberto a esses materiais. Manter o acesso aberto certifica que esses dados não ficaram presos em portais comerciais. Assim, Sales (2014, p. 63-64) ressalta os benefícios de repositórios com padrões e interoperabilidade:

- ✓ Amplia a visibilidade dos resultados de pesquisa, posto que, via de regra só é formalmente disseminada a fração que está registrada nas publicações acadêmicas.
 - ✓ Proporciona mecanismos de preservação de longo prazo, em termos de preservação digital, de arquivamento seguro e de curadoria digital.
 - ✓ Permite que o material depositado esteja disponível *on-line* continuamente para ser consultado e citado mais frequentemente.
 - ✓ É um instrumento chave para os processos de reformatação e recriação de dados proporcionados pela curadoria digital.
- Abre a possibilidade de criação de novos serviços de informação para pesquisadores e gestores a partir da análise dos dados arquivados e a integração de dados e publicações acadêmicas.
- ✓ Permite a criação de redes de repositórios interoperáveis.
 - ✓ Aumenta o grau de reuso dos dados minimizando a duplicação de esforços e otimizando os investimentos na geração de dados.

O repositório de dados se diferencia das outras naturezas de repositórios, pois cada grupo de dados possui atributos exclusivos e requerem análises diferenciadas, como exemplo de tratamento peculiar de um grupo dados, a linguagem de notação XML. Devido às peculiaridades dos dados, as ações de curadoria no repositório de dados devem atuar em 3 níveis: de armazenamento, de representação do conjunto e de representação do item. O armazenamento certifica que os dados sejam guardados de maneira segura em um sistema de constante manutenção, representação do conjunto que garante a descrição dos dados, metadados fidedignos a produto e, por fim, a representação do item que trata do assunto do

modelo e da descrição para itens individuais, sendo dependente da natureza do grupo de dados (RODRIGUES, 2010).

Não obstante, Sales (2014) observa que os repositórios de dados se particularizam por possuírem a estrutura de banco de dados. Mas, através das variantes de cada disciplina, a pluralidade e complexidade dos resultados das pesquisas, os repositórios de dados vêm se tornando excessivamente heterogêneos no que tange ao conteúdo, à concepção, ao tratamento de dados e à gestão. Desta maneira, a escolha da tecnologia para a criação do repositório de dados está relacionada ao modo com que os dados são gerenciados (curadoria de dados), quais os tipos de dados serão armazenados, quais os métodos de pesquisa que darão resultados aos dados e suas possíveis reutilizações.

Por isso, nos repositórios de dados são exploradas muitas soluções tecnológicas que são percebidas como plataformas desenvolvidas de modo genérico para repositórios; outras soluções, no entanto, são produzidas notadamente para um caso exclusivo (RODRIGUES, 2010).

Assim, os repositórios de dados de pesquisa possuem a missão de edificar conexões de dados entre disciplinas que particularizam o estado da arte da pesquisa científica. Pode-se dizer que os pilares para isso ocorrer são o compartilhamento e o trabalho em conjunto e tecnologias e padrões que apoiem todo esse processo de banco de dados de pesquisa. (SALES, 2014)

A seção seguinte dará início à conceituação do CRIS. As seções posteriores tratarão de seus tipos e padrões e suas principais experiências no mundo.

6 CRIS/RIM (RESEARCH MANAGEMENT SYSTEM)

As informações sobre pesquisas muitas vezes são fragmentadas em vários sistemas ou planilhas em uma mesma instituição o que sobrecarrega os sistemas, pesquisadores e funcionários envolvidos neste seguimento e gera dados inconsistentes, incompletos e relatórios não confiáveis, os quais afetam negativamente as decisões da instituição. Para resolver essas questões, surgem sistemas que envolvem todas as informações necessárias para todo o processo de pesquisa, os Sistemas de Gestão de Informações para a Pesquisa.

Nessa seção serão tratados o histórico, o conceito, padrões e softwares que corroboram para a sustentação desse sistema. Ainda, serão mencionadas as principais iniciativas de CRIS no mundo.

6.1 HISTÓRICO

Os passos elementares na padronização de modelos de dados são datados do início da década de 1980, logo após os primeiros sistemas para permuta de pesquisas terem sido criados. Devido ao surgimento desses sistemas, cresce a demanda de coleta, integração e análise de informações de pesquisa, o que torna a gestão de informação de pesquisa um assunto de extrema importância (QUIX; JARKE, 2014).

Durante todo o processo de pesquisa, a integração da informação é um recurso bastante requisitado pelos cientistas, pois as informações de pesquisa são compiladas a partir de várias fontes, o que depende da grande quantidade de esforço e tempo. Como os relatórios sobre os resultados das pesquisas são solicitados por instituições diferentes, resultam em execuções repetidas e, portanto, consomem mais empenho do pesquisador. Desta forma, um modelo uniforme de dados, ou um modelo padrão de dados para informações de pesquisa, torna mais simples a vida do pesquisador (QUIX; JARKE, 2014).

Nesse cenário, surgem os sistemas de informações que têm como finalidade tratar, armazenar e recuperar esses dados produzidos e gerenciar as fases dos processos de pesquisas científicas. Da mesma forma, apresentam o contexto dos procedimentos científicos, aprimoram a qualidade do trabalho e se valem de padrões

que permitem a avaliação de um bom procedimento de pesquisas, assim como a reutilização de dados que resultam em novas descobertas (SAYÃO; SALES, 2014). Estudos como o UKRISS², sobre a gestão de informações de pesquisa ratificam esse pensamento ao mencionar que os benefícios de uma base de dados integrada sobre informações de pesquisa é um dos principais motivos para a implementação de um CRIS (QUIX; JARKE, 2014).

Pela importância já mencionada na esfera da comunicação científica, é relevante mencionar iniciativas de sistemas de informação para pesquisas. Então, os EUA foram um dos pioneiros na criação de um sistema de pesquisa corrente. Tal sistema foi desenvolvido em 1950 e abrangia pesquisas na área médica. Os EUA elaboraram um sistema de âmbito nacional, que abordava as pesquisas básicas e aplicadas chamado de SSIE - Smithsonian Science Information Exchange. Desde então, surgiram outras iniciativas como o Directory of Federal Technology Transfer, o Experimental Contract Highlight Operation Project (ECHO) entre as décadas de 1960 e 1970. Em seguida, outros sistemas apareceram em outros países tais como: Canadá, União Soviética, Israel, Hungria e Holanda (BARRETO, 1981).

O aparecimento desses sistemas incentivou a realização do Simpósio Internacional sobre Sistemas de Informação e Serviços em Pesquisa Corrente em Ciências pela UNESCO (Organização das Nações Unidas para Ciência, Educação e Cultura) em 1975, na França, através do UNISIST, um Programa para Cooperação Internacional em Informação Científica e Tecnológica.

No Brasil, a primeira iniciativa de um sistema de informação de pesquisa corrente que se tem conhecimento é o BRACARIS (Sistema Brasileiro de Informação sobre Pesquisa Agrícola em Andamento), que se tratava especificamente da área de pesquisa agrícola, ou seja, era um CRIS temático, coordenado pelo CENAGRI (Centro Nacional de Informação Documental Agrícola) e interligado ao CARIS (Current Agricultural Research Information System). O BRACARIS surgiu com o intuito de conservar e disseminar um sistema sobre onde, o que e em que estão abrangidos os projetos de pesquisa agrícola ou agro-industrial de pesquisa, quer sejam pesquisadores, quer seja o público em seu todo (BARRETO, 1981). O sistema possuía objetivos como:

² Disponível em: < <https://ukriss.cerch.kcl.ac.uk/> >

- ✓ Manter atualizado um cadastro de instituições, pesquisadores e projetos de pesquisa;
- ✓ Intensificar a comunicação entre as instituições e seus pesquisadores;
- ✓ Disponibilizar dados básicos para a avaliação dos esforços de pesquisa, identificando as lacunas e as áreas carentes de recursos;
- ✓ Oferecer auxílio financeiro às instituições para a tomada de decisões referentes à política científica em agricultura, quer em nível nacional ou estadual (SOUZA, 1983).

Aos poucos começaram a surgir outros sistemas de informação em pesquisa corrente em agricultura pelo mundo, sendo precursores:

O Current Agricultural Research Information System - CARIS, implantado pela Organização das Nações Unidas para a Alimentação e Agricultura (FAO); o Permanent Inventory of Agricultural Research Projects in the European Community - AGREP, criado pela Comissão da Comunidade Européia; o Current Research Information System - CRIS, desenvolvido pelo Departamento de Agricultura dos Estados Unidos; o Inventory of Canadian Agricultural Research - ICAR, desenvolvido pelo Conselho de Pesquisa Agrícola do Canadá e o Retrieval System for Current Research in Agricultural Sciences - RECRAS, pertencente ao Ministério da Agricultura, Floresta e Pesca do Japão (SOUZA, 1983, p.87)

Atualmente, esses sistemas expandiram-se no âmbito temático e dominaram quase todas as áreas de conhecimento. Entretanto, nacionalmente, as iniciativas de desenvolvimento de um CRIS são escassas. Acredita-se que apenas dois sistemas poderiam interoperar em um futuro projeto de elaboração de um CRIS: a Plataforma Lattes³ e o Sucupira⁴.

6.2 CONCEITOS E DEFINIÇÕES

O interesse pelas informações sobre pesquisa cresce constantemente e, como consequência, aumenta o interesse pelo CRIS. Ele é utilizado por vários tipos de pessoas: pelos pesquisadores, pelos gestores de pesquisa, pelos empresários e pela mídia em geral. Esse sistema se tornou fundamental para o apoio à tomada de

³ Disponível em: <<http://lattes.cnpq.br/>>

⁴ Disponível em: <<https://sucupira.capes.gov.br/sucupira/>>

decisão estratégica e o estímulo a melhor qualidade de vida. É o maquinário da sociedade do conhecimento.

Para Zimmerman (2002 apud KOVACEVIC, 2011), os Research Information Management (RIM) têm sido de profunda importância no desenvolvimento da ciência, devido ao serviço que prestam à comunidade de pesquisadores. Esses sistemas coletam, preservam e divulgam dados sobre organizações, pesquisadores, projetos de pesquisa, produtos, dados de resultados obtidos nas pesquisas e diversas outras informações do processo de pesquisa científica.

Basicamente, CRIS e RIM tratam-se do mesmo sistema. O sentido de comunidade profissional para o RIM é significativamente menor do que em outras áreas, como o Repositório Institucional. Isso pode ser uma consequência da natureza específica da atividade de apoio à pesquisa realizada pelas equipes envolvidas nas atividades relacionadas com o CRIS, que é tanto mais interna e orientada comercialmente do que a média da iniciativa de gestão de informação institucional (CASTRO; MENNELLI, 2014).

Outros autores como Ivanovic, Ivanovic, Surla (2012) consideram que os sistemas CRIS devem ser compatíveis com os CERIF (Common European Research Information Format). Já Joint (2008) afirma que embora conceitos como " ambientes virtuais de pesquisas " e " sistemas de gestão de pesquisas " também remetem à ideia de gestão informações de pesquisa e são, de certa maneira, sinônimo de ou inclui o conceito de um CRIS.

De acordo com Bevan e Harrington (2011), o CRIS é um sistema que é baseado na tecnologia de banco de dados estruturado com uma exímia atividade de atualização, bem como excelentes atividades de recuperação e comunicação. Asserson e Jeffery (2009) complementam, ao afirmar que o CRIS oferece uma básica referência de informação para dirigir os fluxos de trabalho - para a própria pesquisa, para os processos administrativos associados e para os processos de gestão - como compras, contabilidade, recrutamento, gestão de pessoal, geração de proposta de projeto, gerenciamento de projeto, publicação, registro de patente e outros. Já interligado com outros sistemas dentro de uma organização, o CRIS se torna o núcleo que une oferecendo informação homogênea a partir de fontes heterogêneas.

O CRIS pode ser caracterizado como um sistema que comporta informações sobre as atividades de pesquisa sobre diversos aspectos. Tratando, principalmente,

de informações que se referem aos autores (pessoas, agências de fomento), às atividades de pesquisas (projetos), às ferramentas (sistemas de software), aos recursos (dados), aos conceitos (teorias, tecnologias), à produção (publicações) e às relações entre essas entidades. Esse sistema é voltado para integração, curadoria, acesso, compartilhamento de dados, informações de pesquisa e é composto por um modelo de dados que descreve objetos de interesse da atividade de pesquisa e de seu desenvolvimento e de uma ferramenta ou conjunto de ferramentas que gerencie dados. O modelo de dados inclui entidades como: pessoas, projetos, organizações, resultados de publicações, resultados de patentes, resultados de produtos e serviços (ERBACH, 2006; ZIMMERMAN, 2002).

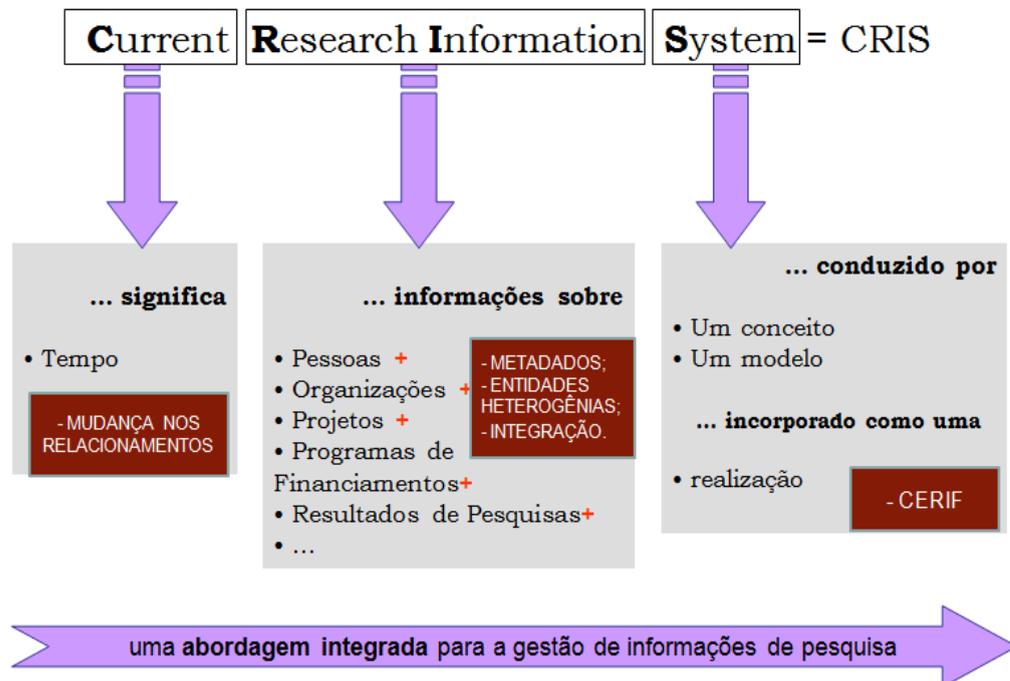
“[...] CRIS é um modelo de dados que descreve um conjunto de objetos de interesse para as atividades de pesquisa e uma série de ferramentas que possibilitam ao usuário (pesquisador, gestor etc.) a gestão de seus dados de pesquisa em todos os processos, incluindo alocação de recursos, avaliação de projetos, identificação de novos mercados para produtos de pesquisa, análise de tendências entre outros serviços.” (SAYÃO; SALES, 2014, p. 86).

Essa estrutura tecnológica pode ser enquadrada como qualquer ferramenta de informação dedicada a fornecer o acesso e a divulgação de informações sobre pesquisas. O seu principal objetivo é gerenciar as informações produzidas em uma instituição ou país⁵.

Segue um esquema que estrutura de maneira organizada o conceito de um CRIS:

⁵ Disponível em: <<http://www.eurocris.org/>>

Figura 2: Conceituando um CRIS



Fonte: <<http://www.eurocris.org/Index.php?page=featuresCERIF&t=1>>

Alguns autores adotam uma definição mais simples para o CRIS, definindo-o como um sistema de gestão de dados que possui a capacidade de integrar dados de vários outros sistemas, por exemplo, extrair dados pessoais de um sistema de recursos humanos; já os dados financeiros dos projetos são captados de um sistema de gestão financeira (QUIX; JARKE, 2014). Resumidamente, o CRIS apresenta a capacidade de retirar as informações específicas de cada sistema com o intuito de reunir todas as informações relevantes ao procedimento de pesquisa em uma única base de dados.

Joint (2008) corrobora a definição acima, pois descreve esse sistema de pesquisa científica como um sistema de informação que gerencia todas as informações relevantes de uma pesquisa, desde as oportunidades de financiamento, da escrita e da apresentação das etapas de propostas até a conclusão do processo de pesquisa. Esse sistema sobre pesquisa científica trata de materiais de manipulação de contratos, projetos, publicações, planos de estudos. Contém dados / metadados ou informações sobre os gerentes de projetos, projetos em andamento e concluídos, departamentos de pesquisa, programas de financiamento e fundos, pesquisadores, resultados de pesquisas

(publicações, patentes, produtos), eventos, instalações, serviços e equipamentos e suas relações e oferece uma abordagem integrada para a gestão de informações de pesquisa. Contudo, estes sistemas focam nas informações curriculares mais do que nas informações pessoais, devido ao meio formal no qual estão inseridos (PINTO; SIMÕES; AMARAL, 2014).

Além de ser utilizado na tomada de decisões em todos os níveis da instituição, o CRIS deve ser utilizado inclusive para a gestão das atividades de pesquisa e para a divulgação dos resultados. É fundamental para facilitar os processos de criação e gestão do conhecimento e, por consequência, o crescimento econômico (ZIMMERMAN, 2012).

No intuito de uma instituição desenvolver um CRIS são necessários alguns pontos, como: melhorar o quadro político-institucionais nos quais os sistemas de informações da universidade operam atualmente; melhorar o momento do acesso dos depositantes, não somente melhorando a interface, mas também pela informação que estará disponível, principalmente aos pesquisadores; desenvolver um “gerador” de perfil de pesquisa que proporcione uma lista completa de autores e suas publicações sendo combinada com a informação relacionada, tais como indicadores, pesquisa e perfis de interesse.

Já foi possível estudar a fundo o conceito de um CRIS. Mas, afinal, qual é a serventia de um CRIS para a instituição e para a comunidade no qual está alocado?

As funções do CRIS são variadas o que possibilita que sejam utilizados para detectar e mapear tendências em ciência, identificar especialistas de disciplinas, encontrar equipamentos ou instalações especializadas, apontar inovações e resultados, com o propósito de minimizar a duplicação de trabalhos, gerenciar o processo de concessão, produzir estatísticas e relatórios, avaliar projetos, promover a ciência na sociedade e angariar fontes de financiamento (ZIMMERMAN, 2012).

O CRIS permite o compartilhamento de informações entre os pesquisadores e mantém a comunicação entre eles, o que facilita a pesquisa interdisciplinar. Um CRIS planejado de maneira inteligente é capaz de fornecer um contexto para a ciência. Ainda, proporciona uma estrutura de apoio às decisões e análises estatísticas, utilizando o CRIS para gerar orçamentos e relatórios em todos os níveis da instituição. Além disso, a tomada de decisões estratégicas e

operacionais em todas as organizações é baseada na boa informação. Esta informação vem de fontes variadas, mas muitas vezes um único indivíduo toma as decisões. Por isso, os dados do CRIS podem abastecer os sistemas de suporte necessários para orientar os tomadores de decisão.

Zimmerman (2012) afirma que o CRIS pode ser usado para dar suporte à gestão de conhecimento institucional, à administração do processo de desenvolvimento da proposta, ao processo de submissão e revisão por pares, à supervisão de subsídios e à publicação dos resultados. Neste sentido, ao trabalhar como um todo integrado, CRIS administra o processo de pesquisa e possibilita a publicação e o acesso à informação de pesquisa. Pode ser usado para publicar resultados de pesquisas em revistas eletrônicas com revisão por pares, assim como os *preprints*, fazendo pleno uso de multimídia e capacidades de hipertexto. O sistema CRIS pode ser utilizada como a espinha dorsal dos repositórios digitais e bibliotecas digitais.

Com a grande quantidade de informação disponível é cada vez mais difícil encontrar as informações relevantes. A produtividade não pode diminuir devido ao tempo necessário para peneirar a grande quantidade de dados e informações de qualidade variável; por isso, o CRIS gerencia o conhecimento produzido. A gestão do conhecimento tem duas funções principais: a construção da comunidade de práticas e o aumento na estrutura de um corpo de conhecimento rico, de maneira que ele possa ser usado e reutilizado. A gestão do conhecimento é fundamental para que novos conhecimentos sejam transferidos para a resolução de um problema novo (ZIMMERMAN, 2002).

Para as agências de fomento, o CRIS auxilia cooperando com o planejamento ou com os programas de competição de pesquisas. Colabora para o ajuste das prioridades de pesquisa de maneira com que coincidam com a programação da instituição. Também encontram colaboradores adequados para as propostas de pesquisa e comparam com o desempenho de pesquisa estatisticamente. Da mesma maneira, em instituições de pesquisa, os gestores de pesquisa podem acompanhar seu desempenho e planejar suas estratégias. Já os estrategistas podem analisar a pesquisa de financiamento de outros países e, por consequência, ajustar seus planos, possibilitando aos empreendedores poder encontrarem ideias para comercialização (ASSERSON; JEFFERY, 2009).

Ao mesmo tempo, os pesquisadores podem encontrar outros pesquisadores

que trabalham nos mesmos temas de suas pesquisas ou em tema correlacionado e rastrear seus projetos, métodos, resultados, com o propósito de estimular a concorrência ou a cooperação. Isto é um diferencial nas pesquisas interdisciplinares em que o pesquisador não está trabalhando com a comunidade de pesquisadores na qual está familiarizado. No que se refere à mídia, ela se torna capaz de acessar inúmeras informações de pesquisa que subsidiam "histórias de pesquisa" de interesse para o público em geral (ASSERSON; JEFFERY, 2009).

Na instituição do qual faz parte, o CRIS serve como instrumento na formulação de políticas, gera um "inventário" das pesquisas em andamento e avalia as pesquisas através do registro das atividades se baseando nas produções dos trabalhos. Igualmente, auxilia seus pesquisadores em seus apontamentos, nas apresentações de relatórios e nas tomadas de decisões a respeito da pesquisa a ser realizada. Já para os usuários "consumidores" é essencial na avaliação das oportunidades de financiamento da pesquisa, evita a duplicação de atividades de pesquisa, analisa as tendências, compila referências e links para um texto completo ou publicações acadêmicas multimídias, localiza novos contatos e identifica novos mercados para os produtos da pesquisa.

Por isso, o principal motivo para implementação de um CRIS é o gerenciamento de informações produzidas em uma instituição. De acordo com Bevan e Harrington (2011), o CRIS é um sistema que visa proporcionar uma única fonte de informação de pesquisa e que permite o planejamento, monitoramento, relatórios, comunicação e modelagem futura de atividades de pesquisas. Eles mencionam que, para desenvolver um CRIS em uma universidade (ou instituição) é necessário seguir as seguintes etapas:

- ✓ Aprimorar os quadros políticos institucionais nos quais os sistemas de informações da universidade operam atualmente;
- ✓ Melhorar o acesso dos autores, não somente na interface, mas, também, aperfeiçoar a informação que estará disponível, principalmente, aos pesquisadores e,
- ✓ Desenvolver um buscador de perfis de pesquisas que proporcione uma lista completa de autores e suas publicações, combinando com a informação relacionada, tais como indicadores, pesquisa e perfis de interesse.

O CRIS deve tornar evidente a importância das informações que disponibiliza, pois os usuários externos geralmente têm interesses muito específicos de informação. Deve ser capaz de suportar uma busca de especialistas ou instituições especializadas relacionadas a questões específicas, fornecendo informações estruturadas para os processos de avaliação (HORNBOSTEL, 2006).

Obviamente, as informações que constam em um CRIS devem ser sempre atualizadas e devem ser válidas, mas no caso dos rankings, os editores são obrigados a verificar a veracidade das informações. A propagação da responsabilidade ocorre muito rapidamente no caso de sistemas descentralizados; por isso, o melhor método é ter o sistema de validação cruzada⁶.

Não obstante, é necessário que o CRIS seja estruturado de tal maneira que as informações sobre as pessoas, projetos, publicações, patentes possam ser utilizadas em contextos diferentes. O CRIS deve ser tão útil para análises internas quanto para os relatórios dirigidos ao público em geral, permitindo o acesso a informações detalhadas a um nível de usuário, enquanto usuários externos fornecem agregações dependentes de demanda. Assim, a qualidade de um CRIS depende dos benefícios que fornecem aos cientistas, tanto no fornecimento de dados, quanto no acesso e na utilização de dados. A principal qualidade de um CRIS deve ser facilitar o trabalho do pesquisador, como, por exemplo, evitar a gravação múltipla de dados. Isto pode ser conseguido ligando-se com outros recursos de informação. Contudo, essa simplicidade só é possível quando o CRIS é sistematicamente incorporado na gestão da pesquisa, nos processos de tomada de decisão, avaliação e elaboração de relatórios de pesquisa a nível institucional, o que fará com que os cientistas tenham vital interesse no fornecimento de informações regularmente válidas. A vantagem que CRIS oferece é a estrutura quando segue determinados padrões, como, por exemplo o CERIF. No entanto, para isso é necessário um alto nível de interoperabilidade semântica e técnica (HORNBOSTEL, 2006).

Assim como suas qualidades, as funções do CRIS são variadas como as funções dos sistemas de informação. Eles podem ser usados, entre outras

⁶É uma ferramenta de análise e é de grande importância no auxílio de desenvolvimento de e ajuste de modelos de mineração de dados. É utilizada após a criação de uma estrutura de mineração e os modelos de mineração relacionados para assegurar a validade do modelo.

coisas, para detectar e mapear tendências da ciência, identificar a área dos especialistas, encontrar equipamentos ou instalações especializadas, reconhecer inovações e resultados (para evitar a duplicação de esforços das pesquisas), gerenciar o processo de concessão, produzir estatísticas e relatórios, avaliar projetos e avaliar ciência, promover a ciência na sociedade, e para localizar fontes de financiamento (ZIMMERMAN, 2002). Nessas linhas que estão sendo examinadas durante a exploração do tema, concebe-se que os Current Research Information Systems apresentam diferentes níveis de abrangência de informação, que atingem pesquisadores com interesses diversos em ambientes diversificados. Por isso, existem tipos de CRIS que reúnem as principais características de um conjunto de CRIS. Desta maneira, o próximo tópico tratará dos tipos de CRIS, suas respectivas características e exemplos de cada tipo de CRIS.

6.3 TIPOS DE CRIS

O CRIS pode ser aplicado em algumas esferas: **Institucional; Nacional; Regional e Internacional**. E, também existem CRIS que são caracterizados por **área de conhecimento ou por disciplina** (PINTO; SIMÕES; AMARAL, 2014).

1 – Institucional

Possui informações de apenas uma instituição e o seu acesso é restrito.

Seguem exemplos de CRIS institucionais:

- ✓ HKU Scholars Hub
- ✓ Sistema de Informação para a Gestão Agregada dos Recursos e dos Registos Acadêmicos

2 – Nacional

Trata da gestão das informações de pesquisa de várias ou todas as instituições de determinado país. A infraestrutura de um CRIS nacional é constituída pelas seguintes partes: banco de dados central com a

apresentação do seu conteúdo web e o sistema de coleta de dados⁷.

Seguem exemplos de CRIS nacionais:

✓ Slovenian Current Research Information System (SICRIS)

Esse sistema possui o diferencial de ser interoperável com outros sistemas, como, Scopus, Google Scholar or SciELO e, esta integração é capaz de evitar o retrabalho.

✓ National Academic Research and Collaborations Information System (NARCIS)

✓ Russian Research Community CRIS (Socionet)

3 – Regional ou Internacional

Para que um CRIS seja classificado como internacional deve envolver mais de um país e apresentar outros itens primordiais como: discernir dado, informação, conhecimento e suas respectivas interações; contar com tecnologia que trate das linguagens naturais e de suas variações; incentivar que autores, avaliadores e editores ingressem e participem do sistema; gerar a menor quantidade possível de formulários a serem preenchidos pelos pesquisadores; distinguir informação validada de informação duvidosa e, desde o início, possuir uma vasta quantidade de parceiros influentes na área (MONS; ET AL., 2014).

Segue exemplo de CRIS regional:

✓ IST World

4 – Por Área de Conhecimento ou Disciplina

Possui uma estrutura temática ou empenham-se em uma área de pesquisa específica.

Seguem exemplos de CRIS por áreas de conhecimento:

⁷ É um serviço de verificação de integridade dos dados que são públicos e um conjunto de ferramentas que apoiam o processo de coleta de dados.

- ✓ Human Nutrition Research Information Management - Nutrição
- ✓ CRIS of National Institute of Food and Agriculture (NIFA) – Agricultura.

6.4 CRIS X REPOSITÓRIOS INSTITUCIONAIS

Esta seção fará um comparativo entre as principais características do CRIS e dos repositórios institucionais, com o propósito de disseminar conhecimento de ambos os assuntos.

Conforme mencionado anteriormente, devido à aceleração no contingente da produção de informações de pesquisas, emerge a necessidade de criar maneiras que facilitem o trabalho do profissional na gestão dessas informações. Por isso, são concebidos sistemas que visam e facilitam o acesso dos pesquisadores a esses materiais, tais como os Repositórios Institucionais (RIs) e os CRIS.

Os Repositórios Institucionais são os primeiros a divulgarem os materiais produzidos na instituição. Eles são abertos e oferecem meios de busca, identificação e recuperação. Possuem a atribuição de aumentar a visibilidade da instituição e do material produzido nesta, de modo informal e ágil, além de disporem da função de memória institucional (MULLER, 2006).

Acentua-se que o conceito de repositório digital está em constante mutação, “podendo abrigar a produção intelectual de uma instituição, denominado repositório institucional, ou repositório temático que coleciona e preserva material de determinada área do conhecimento” sendo possível afirmar que, de maneira resumida, o repositório “pode ser descrito como um arquivo em formato digital [sistema digital de arquivamento] que congrega a produção acadêmica de uma instituição. É um conceito aderente aos movimentos OAI [Open Achieves Initiative] e arquivos abertos” (CARVALHO, 2011, p. 25). Neste contexto, propriamente dito, a palavra arquivo é aplicada como espaço controlado de depósito de informações digitais. Além do que já foi explicitado, “os RI abrigam um conjunto da literatura que está no formato digital, *on-line*, isenta de custos, e livre de empecilhos no que diz respeito aos direitos autorais. [...], o conteúdo deve estar como texto completo, e não somente resumos e referências” (SIMPSON, 2005 apud CARVALHO, 2005).

Björk (2005 apud Muller, 2006) relata que os repositórios institucionais estão entre os quatro tipos de canais abertos mais importantes, além de periódicos científicos eletrônicos com avaliação prévia pelos pares, servidores de e-prints para áreas específicas – repositórios para assuntos específicos e autoarquivamento em páginas pessoais dos autores. Entretanto, Muller (2006, p. 33) afirma que dentre esses canais, o repositório institucional é o que possui o maior potencial de desenvolvimento, pois “disciplinam e sistematizam a atividade do arquivamento individual”.

De acordo com Ferreira (2008, p117-118 apud CARVALHO, 2011, p. 26),

“ [...] o repositório assume características que o tornam único e peculiar [...] e [...] utiliza as TIC oriundas da internet tanto para a atividade de depósito como para divulgação e o compartilhamento dos conteúdos [...]. Estes sistemas [...] antecipam, assim, a disseminação dos conhecimentos recém-gerados, reforçam o acesso gratuito e o direito do autor em depositar seu trabalho a sua escolha e assumem, ainda, a responsabilidade de distribuição do material ”

No intuito de facilitar e agilizar o acesso à produção do conhecimento da Universidade de Cornell, surge em 1991, o ArXiv – sistema constituído na comunidade acadêmica, mas especificamente na escola de física. Por causa de sua constante evolução os repositórios, hoje, podem ser classificados nos seguintes grupos: institucionais, nacionais, internacionais, mantidos por consórcios de pesquisa, por agências de financiamento, dentre outros (CARVALHO, 2011).

No que tange à usabilidade para os usuários, ressalta-se que para recuperar a informação é necessária a compilação de perspectivas intelectuais da descrição da informação e sua especificação para a busca, do mesmo modo os sistemas de recuperação da informação devem seguir estas primícias. Desta maneira, pela importância dos RIs na reestruturação do modelo de publicação científica tradicional e, por possuírem como principal característica a preservação digital, esses sistemas devem acompanhar tais princípios (DIAS, 2003 apud MACHADO, 2012).

Comumente, os repositórios eram e/ou são estruturados utilizando o Dublin Core como padrão de metadados. Por isso, algumas instituições desenvolvem recursos fundamentados em pacotes desse software que originalmente eram destinados à gestão de repositórios institucionais para a criação de seus CRIS (SAYÃO; SALES, 2015a)

Com a literatura analisada, pode-se notar alguma diferenças entre os CRIS e os Repositórios Institucionais (SAYÃO; SALES, 2015a):

- ✓ Possuem objetivos distintos que por diversas vezes se sobrepõe;
- ✓ O CRIS abrange os processos de pesquisa científica com caráter mais gerencial e abrangente, servindo à instituição a qual é subordinada.
- ✓ O RI possui uma perspectiva acadêmica, disponibilizando publicações acadêmicas por completo, no intuito de disseminar a memória da instituição através de protocolos.
- ✓ O CRIS registra os dados bibliográficos e documenta as informações ligadas ao fomento.
- ✓ Já o RI arquiva os textos completos das publicações acadêmicas.

Através do exposto acima, pode-se concluir que o CRIS centra-se no acompanhamento das atividades de pesquisa, enquanto o RI está focado na preservação do material e maximização do impacto destes, e não, somente em dados bibliográficos. Isso se deve ao fato de que muitas vezes esses sistemas estão vinculados a diferentes unidades organizacionais em várias organizações acadêmicas e de pesquisa. Por exemplo: RI é subordinado à biblioteca e outras unidades de informação, enquanto o CRIS é subordinado às instâncias administrativas de instituições de pesquisas (SALES, SAYÃO, 2015a).

Mas é a superioridade dos CRIS no fornecimento das informações contextuais e dos metadados mais completos, que são formalmente estruturados, de modo que sejam compreensíveis e legíveis por computador. Além disso, possibilita a formação de uma rede de publicações para pesquisar grupos de dados e software, proporcionando um ambiente mais rico para o pesquisador (ASSERSON; JEFFERY, 2009).

Pode-se notar a melhor adequação de um CRIS no âmbito da homogeneidade de informações de pesquisas, quando Zimmerman e Jeffery (2002) mencionam que para definir uma plataforma de descrição de informação uniforme é preciso:

- ✓ Definir um modelo de dados CRIS completo que irá cobrir as estruturas de banco de dados da maioria dos CRIS existentes;

- ✓ Caracterizar um conjunto de modelos de dados que poderia fornecer exemplos para a troca de dados;
- ✓ Criar um modelo de dados de metadados para fornecer uma ideia de uniformidade de resumos para fontes de informação heterogêneas.

A subseção abaixo analisará algumas normas para o CRIS. Algumas dessas normas podem ser utilizadas no CRIS. Essas normas são: protocolo Z39.50, OAI-PMH, OAI-ORE, Dublin Core (DCMI), CERIF, NL CERIF, CASRAI e ORCID.

6.5 NORMAS/PADRÕES

Através do progresso das tecnologias emergiram novos protocolos de comunicação a fim de possibilitar a interoperação entre sistemas heterogêneos. Isso possibilita a troca de informação entre eles, torna possível o acesso aos serviços e informações desejados a partir de um ponto de contato, ao mesmo tempo que agiliza a recuperação da informação.

PROTOCOLO Z39.50

A ideia de interoperação teve início através da NISO (National Information Standards Organization) que formou um comitê com o intuito de criar um protocolo de recuperação de informação. A partir das análises feitas por este comitê, nasceu a primeira versão do protocolo Z39.50. Este protocolo proporciona a comunicação entre computadores com objetivo de facilitar a pesquisa e a recuperação da informação em redes de computadores alocados em diferentes lugares. Está embasado na estruturação cliente/servidores e atua sobre a internet (RODRIGUES et al, 2011).

Após essa versão, criou-se um protocolo de interoperação que é capaz de ser implementado em qualquer site ou sistema operacional que não for compatível com o sistema que gerencia o banco de dados. Além disso, sua interface conecta-se a vários sistemas e, desta maneira, o pesquisador pode ter acesso a outro sistema sem precisar se desconectar do sistema ao qual está acessando. Ou seja, por mais que o pesquisador tenha acesso a vários sistemas, o resultado da busca é

apresentado no sistema local. Contudo, este padrão está contido em um ambiente de constantes atualizações tecnológicas e, por isso, é avaliado e inovado continuamente.

O projeto de reestruturação foi denominado pelo acrônimo ZIG/ZING (*Z39.50 Next Generation*). Uma evolução do ZING são os padrões SRU (*Search/Retrieve via URL*) e SRW (*Search/Retrieve via Web Services*), voltados para o contexto da Internet (MOREIRA; RIBEIRO, 2008). SRW e SRU definem uma forma padrão para pesquisas na Internet, bem como a estrutura das respostas usando como transporte o protocolo HTTP (RODRIGUES et al, 2011, p. 352).

OAI-PMH E OAI-ORE

Outra iniciativa importante de ser mencionada é o movimento do acesso livre que trouxe consigo a necessidade da interoperabilidade entre os repositórios e/ou periódicos. Posteriormente viria a desencadear o aparecimento de alguns padrões como o OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) que incentiva padrões de interoperabilidade no intuito de disseminar a informação de maneira eficiente. Não obstante, ter tido origem nas atividades de acesso aberto e dos repositórios institucionais, ampliou-se na promoção da disseminação dos recursos digitais. Vale lembrar que SRW/U e OAI-PMH são protocolos complementares, pois possuem propósitos similares e são capazes de recuperar metadados de repositórios digitais; no entanto, diferenciam-se na técnica de recuperação da informação. Algumas plataformas de CRIS comerciais desenvolveram recentemente características de repositórios Institucionais, como a conformidade com o protocolo OAI PMH.

A principal diferença entre os protocolos SRW/U e OAI-PMH está no objetivo específico de cada um, enquanto os protocolos SRW/U são uma ferramenta indicada para usar quando se deseja extrair apenas os dados que atendem a requisitos de uma consulta, o OAI-PMH é mais pertinente quando se deseja ter subconjuntos predefinidos dos dados (RODRIGUES et al, 2011, p. 352).

Desta forma, o OAI-PMH é mais recomendado para a criação de um metarrepositório de metadados, porque possibilita a extração de metadados de novos documentos depositados no sistema de maneira periódica, enquanto o

SRW/U é pertinente para questões específicas com intenção de encontrar materiais de uma consulta.

Existe ainda a norma para OAI-ORE (Object Reuse and Exchange) para um objeto digital complexo que permite a descrição das agregações de objetos de modo padronizado. Esse objeto digital pode ser conceituado como uma publicação ampliada e como qualquer outro material precisa ser descrito da maneira correta a fim de ser recuperado. Mas para que essas informações possam ser trocadas com outros sistemas é necessário que estejam descritas de forma padronizada. Para esse fim, existe a norma que tem por objetivo disseminar o conteúdo dessa reunião para ministrações que admitam sistemas de autoria, depósito, intercâmbio, visualização, reuso e preservação (SALES, 2014).

DUBLIN CORE (DCMI)

O Dublin Core (DCMI) é outro padrão comumente utilizado nos sistemas de gestão de informação, considerado um padrão de metadados que proporciona a catalogação de documentos eletrônicos na Web e oferece suporte ao compartilhamento no design de metadados e as melhores práticas através de uma ampla gama de modelos de objetivos e atividades. Fornece esse suporte através do gerenciamento da curadoria a longo prazo e desenvolvimento das especificações do DCMI e dos termos de metadados; do contínuo gerenciamento da discussão dos atuais temas de trabalho do DCMI; da configuração e gerenciamento de eventos internacionais e regionais; da criação e entrega de recursos de treinamento de metadados; e da coordenação da comunidade global de voluntários DCMI⁸.

Existem 15 elementos básicos de Dublin Core que foram apresentados no documento Dublin Core Metadata Element Set⁹ : Título, Criador, Assunto, Descrição, Publicador, Colaborador, Data, Tipo, Formato, Identificador, Fonte, Idioma, Relação, Cobertura, Direitos Autorais. Esses 15 elementos podem ser estendidos pela utilização do refinamento, que diz respeito aos qualificadores Dublin Core. Quando

⁸ Disponível em: <<http://dublincore.org/about-us/>>

⁹ Disponível em: <<http://dublincore.org/documents/2010/10/11/dces/>>

esses qualificadores são adotados geram a possibilidade de representação da informação com maior precisão (BENTANCOURT; ROCHA, 2012).

MODELO CERIF

A heterogeneidade dos CRIS faz emergir a necessidade de um formato padrão para a troca de informações de desenvolvimento e pesquisa. Essa padronização é necessária para regulamentar o desenvolvimento dos CRIS e, principalmente, para possibilitar que esses sistemas alcancem os níveis mais elevados de interoperabilidade. Percebe-se que padrões nacionais e iniciativas internacionais ou regionais não são eficazes, porque possuem abrangência limitada ou os padrões desenvolvidos estão em paralelo com as respectivas políticas (PINTO; SIMÕES; AMARAL, 2014).

A ideia do CERIF nasce quando surge a necessidade de um formato padrão para a troca de informações de desenvolvimento e pesquisa. Assim, é conceituado como um modelo que suporta a gestão de informações de pesquisa, permitindo a criação e a interoperabilidade entre Sistemas de Informação de Pesquisa (JÖRG; et al, 2009). Sendo criado como um formato de troca de dados e, baseado em registros que descrevam projetos, com pessoas e unidades organizacionais como atributos.

É criado um padrão que permite a criação e a interoperabilidade entre Sistemas de Informação de Pesquisa (JÖRG; et al, 2009). Este fornece uma base para um sistema de gestão de informações de pesquisas, onde propõe um modelo de dados formal que inclui entidades, atributos, relações entre entidades, um modelo conceitual que descreve a semântica, modelos lógicos bem documentados com mapeamentos para modelos conceituais e implementações de referência em SQL e XML disponíveis (QUIX; JARKE, 2014).

Este padrão é útil para conteúdos de informação e para indexação de assunto. Além disso, gera listas de vocabulários controlados que facilitam no recolhimento e na troca de dados, caso estas estejam adaptadas às regras comuns. Por isso, a importância do uso de vocabulário controlado padrão. O uso da padronização do

vocabulário controlado e das estruturas de dados deve ser tão disseminado quanto possível, a fim de tornar os provedores de dados e os cientistas familiarizados com as características CRIS (ZIMMERMAN, 2002).

De acordo com o site da EuroCris, o CERIF segue um padrão simples parecido com o Dublin Core Metadata Standard (DC), sendo criado como um formato de troca de dados e baseado em registros que descrevam projetos, com pessoas e unidades organizacionais como atributos.

Muitas mudanças foram necessárias, por isso houve várias versões de CERIF. No CERIF2006 foram implementadas melhorias substanciais, como por exemplo, a introdução da camada semântica, que torna o modelo flexível e escalável para aplicações em ambientes heterogêneos.

O CERIF propicia interoperabilidade entre os sistemas CRIS e outros sistemas associados com a sintaxe formal e semântica declarada para que seja confiável e escalável, o que sana, portanto, a questão da padronização.

Evidentemente, é possível afirmar que o CERIF pode ser conceitualizado como um conjunto de diretrizes projetadas para lidar com todos os sistemas de informação de pesquisa, sendo aplicado no desenvolvimento de novos sistemas de informação de pesquisa, no auxílio dos CRIS consolidados - ao considerar as respectivas extensões já existentes, e na orientação de como estruturar e indexar dados nos CRIS (ZIMMERMAN, 2002).

Do mesmo modo, Russel (2011, p.1) certifica que o CERIF é:

[...] um padrão para o gerenciamento e intercâmbio de dados de pesquisa, ou seja, informações sobre os pesquisadores, organizações, projetos, produtos e financiamento, decorrente do processo de pesquisa. Ele fornece um modelo de dados que pode ser usado para descrever o domínio da pesquisa, incluindo as relações entre as partes constituintes, e quão estas mudam com o tempo.

Apesar de existir a possibilidade de um CRIS ser implantado utilizando um subgrupo ou um supergrupo do modelo completo do CERIF com base nas funções e relações de ligação temporárias, isso não é aconselhável, pois a personalização do CERIF pode comprometer seu principal objetivo: unificar modelos de dados no intuito de garantir a interoperabilidade. Dentro deste

raciocínio, afirma-se que há maior compatibilidade entre os sistemas, caso a versão integral do CERIF seja adotada (PINTO; SIMÕES; AMARAL, 2014).

Este padrão é dividido em três níveis¹⁰:

- ✓ No nível conceitual descreve conceitos sobre entidades de pesquisa e as especificações de seus relacionamentos;
- ✓ No nível lógico descreve entidades de pesquisas e de seus modelos de relacionamentos;
- ✓ No nível físico formaliza entidades de pesquisas e seus relacionamentos de scripts de banco de dados.

Do mesmo modo, percebe-se no decorrer das leituras realizadas que estes níveis são os mesmos assinalados por Silva e Neves. Os autores mencionam que esse princípio é conhecido como “Nível de Abstração dos Dados” e é conceitualizado como “[...] uma quantidade de abstração de informações que varia conforme o tipo de usuário e modelo de dados escolhido. [...], esta característica representa a omissão de detalhes sobre o armazenamento das informações para alguns usuários” (KORTH; SILBERSCHATZ, 1995 apud SILVA; NEVES, 2012, p. 4).

Neste sentido, podemos afirmar que, as principais características do CERIF são¹⁰: ter o conceito de objetos ou entidades com atributos, como: projeto, pessoa, unidade organizacional; ser internacionalizado na linguagem e conjunto de caracteres e; ser extensivo, garantindo a interoperabilidade.

Para Bevan e Harrington (2011), é necessário que os sistemas estejam compatíveis com o CERIF, pois é o melhor padrão de compartilhamento de dados de pesquisas e proporciona a interoperabilidade entre os CRIS e sistemas associados, com sintaxe formal e semântica. Por esses e vários outros benefícios, o CERIF vem sendo cada vez mais adotado como padrão de um CRIS e é recomendado aos países membros da União Europeia.¹⁰

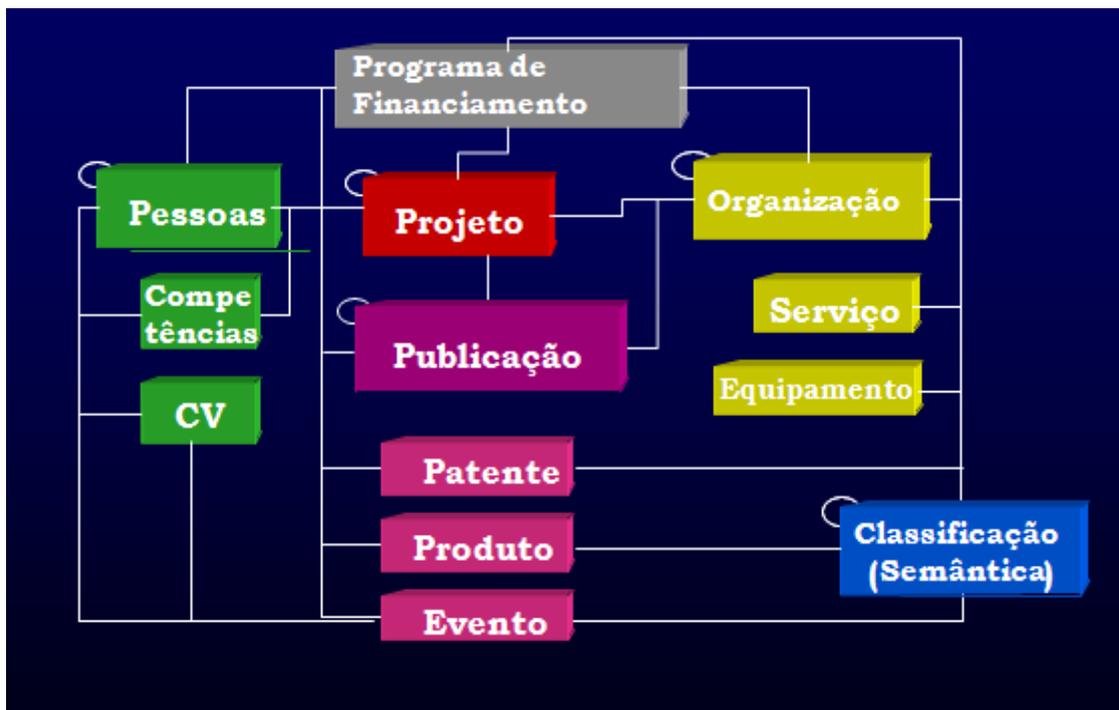
Ao fornecer um modelo de dados, o CERIF permite que os metadados descrevam entidades de pesquisas, suas atividades, interconexões e seus resultados, admitindo alta flexibilidade, com relações formais, através de sua

¹⁰ Disponível em: < <http://www.eurocris.org/>>

semântica. Da mesma forma, permite a conservação da qualidade, do arquivamento, do acesso e do intercâmbio de informações de pesquisas e suporta a transferência de conhecimento para administradores, avaliação de pesquisa, gestores de pesquisa, estrategistas, pesquisadores, editores e público em geral. Um CRIS pode ser implementado utilizando um subgrupo ou um supergrupo do modelo completo do CERIF, com base nas funções e relações de ligação temporárias. Em suma, a “grande ideia” presente no modelo de dados do CERIF é sua estrutura e a lógica de primeira ordem – sintaxe e semântica, respectivamente¹¹.

Na figura 3 existe um esquema para que se possa compreender melhor a estrutura organizacional deste modelo:

Figura 3: Modelo CERIF



Fonte: <<http://www.eurocris.org/Index.php?page=featuresCERIF&t=1>>(tradução nossa)

O CERIF possui características como expansibilidade e é compreensível e legível por máquina, mantendo a continuidade das versões anteriores para permitir a

¹¹ Disponível em: < <http://www.eurocris.org/>>

interoperação. É capaz de promover a conexão com qualquer outro sistema, utilizando as entidades de ligação. Além disso, é planejado para evitar a replicação de dados e solucionar problemas de integridade na atualização. O CERIF é implementável através da tecnologia de hipermídia para recuperação da informação semiestruturada e para sistemas baseados no conhecimento. Segue a lógica de primeira ordem, estando disponível para dedução e indução, gerando maior utilização das informações (dados)¹².

Em seu documento, JÖRG; et al (2009) retrata que o CERIF é estruturado em tipos de entidades e recursos. Baseado nesse documento, segue um quadro com as entidades CERIF e seus respectivos elementos e funções. Logo após, está estruturado um esquema que destina-se a correlacionar as relações entre as entidades CERIF. Tanto o quadro, quanto o esquema elucidam as entidades CERIF e suas respectivas ligações.

Quadro 6: Tipos de Entidades CERIF

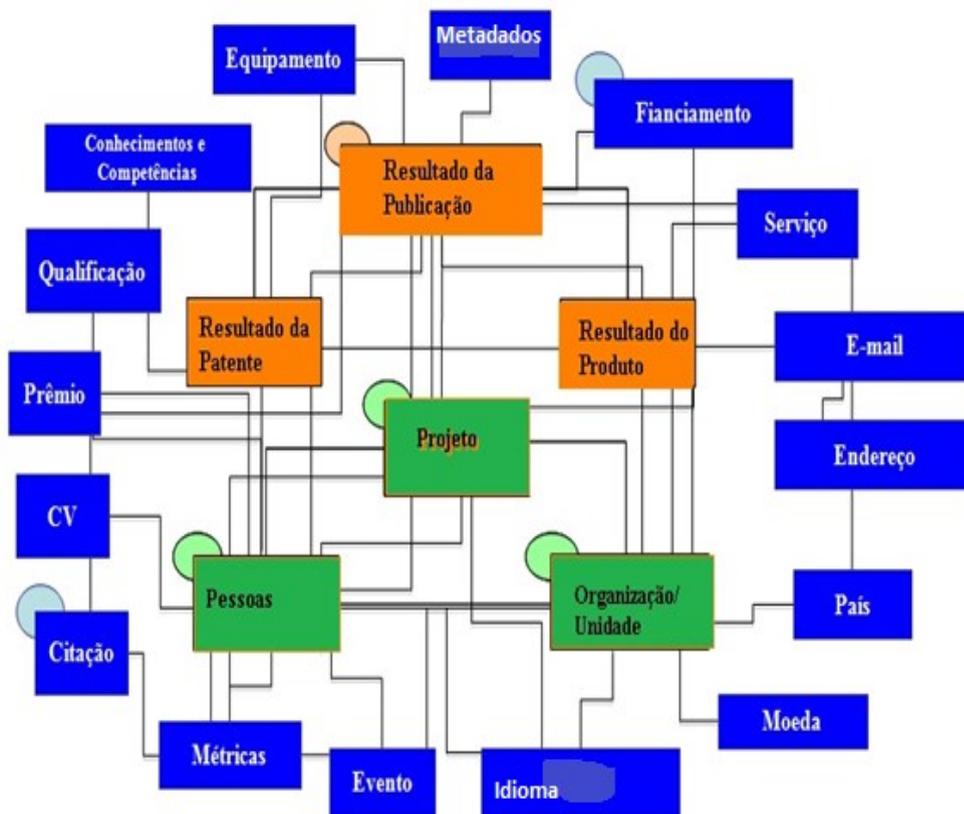
Entidades CERIF	Elementos	Funções
Entidades Centrais	Projetos, Pessoas, Organização /Unidade	Repetidamente, cada entidade central faz ligações para si e, mantém relações com outras entidades centrais. As entidades centrais permitem uma representação dos autores científicos e de suas diferentes formas de interações.
Entidades de Resultados	Resultado da Publicação, Resultado de Patente, Resultado de Produtos	Realizam ligações para si, recursivamente. Essas entidades representam os resultados das pesquisas.
Entidades de Segundo Nível	Equipamentos, Serviços, Países, Língua, Evento, Citação, E-mail, dentre outros.	Permitem a representação do ambiente de pesquisa, ligando a si próprias nas entidades centrais, nas entidades de resultados e em outras entidades de

¹² Disponível em: < <http://www.eurocris.org/>>

		segundo nível.
Entidades de Ligação	As relações ou ligações entre as entidades CERIF	São consideradas o ponto forte do modelo CERIF. Esta entidade tem a capacidade de fazer ligações entre duas entidades.

Fonte: Baseada em JÖRG; et al (2009)

Figura 4: Ligações entre as Entidades CERIF



Fonte: <<http://www.eurocris.org/Index.php?page=featuresCERIF&t=1>>(tradução nossa)

O recurso de multilinguagem existe, porque muitas informações em ambientes de pesquisa precisam ser representadas em mais de um idioma. Esse recurso multilingue é de extrema importância, sobretudo em países onde várias línguas oficiais são faladas e preservadas. Com esse recurso, o CERIF é capaz de oferecer

suporte de multilinguagem para nomes, títulos, descrições, palavras-chave, sumários e, até mesmo, para a semântica. É importante ressaltar que, além das Entidades Centrais, das Entidades de Resultados e das Entidades de Segundo Nível, as Entidades de Classificação da Camada Semântica permitem registros em múltiplas línguas, possibilitando manter esquemas de classificação em diferentes idiomas (JÖRG; et al, 2009).

A camada semântica permite a representação de tipos de relacionamentos, visualização de aplicativos, classificações de assunto e quaisquer outros esquemas de classificação ou mapeamentos entre esquemas (JÖRG; et al, 2009). Essa camada possibilita a categorização de entidades e suas correlações de acordo com as estruturas classificatórias (IVANOVIC; SURLA; RACKOVIC, 2011 apud SAYÃO; SALES, 2014).

A camada semântica do CERIF consiste em dois tipos de categorias de entidades: a classificação <cfClass> e a classificação esquema <cfClassScheme>, permitindo, também, uma representação de termos em vários idiomas <cfClassTerm> e descrições de categorias <cfClassDescr> (JÖRG; et al, 2009,).

Entretanto, Ivanovic, Surla e Rackovic (2011 apud SAYÃO; SALES, 2014) garantem que as demais entidades CERIF são ligadas na camada semântica através da entidade <cfClass>, descrevendo o nível de colaboração da pessoa na elaboração do resultado da pesquisa, categorizando o resultado, as publicações em que o resultado é disseminado, o evento em que o resultado é apresentado e o prêmio que é dado ao autor. Os autores advertem que na versão CERIF 2008 é aconselhável a utilização do tesauro multilingue Ortelius¹³ que “padronizou a indexação de assunto e os códigos utilizados para as áreas de atividades econômicas e produtos e ainda uma lista controlada de valores e atributos de determinados elementos” (IVANOVIC; SURLA; RACKOVIC, 2011 apud SAYÃO; SALES, 2014, p. 88).

Muitas mudanças estruturais foram implementadas ao longo dos anos; por isso, houve várias versões de CERIF. No CERIF2006 melhorias substanciais ocorreram, como, por exemplo, a introdução da camada semântica que torna o modelo flexível para aplicações em ambientes heterogêneos.

A título de interesse, seguem as modificações realizadas no modelo CERIF

¹³ Disponível em: < http://cordis.europa.eu/cerif/src/sum_concl.htm>

no decorrer dos anos. Tais informações foram estruturadas no formato de um quadro que demonstra as diferenças de cada atualização do padrão.

Quadro 7: Inovações nas atualizações do modelo CERIF

CERIF	Atualizações
CERIF 91	<ul style="list-style-type: none"> - Criação de redes de banco de dados; - Troca de registros; - Recomendação para os estados membros.
CERIF 2000	<ul style="list-style-type: none"> - CERIF é entregue aos cuidados do EuroCRIS; - Ferramentas: modelo de dados; indexação de assuntos (vocabulário controlado); orientações para usuários (serviço de informações; perguntas frequentes); - Multilíngue; - Recomendações da Comunidade Europeia para os estados membros.
CERIF 2004	<ul style="list-style-type: none"> - Modelo de introdução e especificação; - Ontologia CERIF.
CERIF 2006	<ul style="list-style-type: none"> - Implementado “o princípio” da camada semântica, que possibilitou a flexibilidade ao modelo para aplicação em ambientes heterogêneos; - Modelo de Introdução e Especificação de Documentos (FDM); - CERIF XML Exchange Specification Document; - CERIF XML Schemas for XML Validation; - CERIF XML Example Files.
CERIF 2008	<ul style="list-style-type: none"> - Modelo Introdução e Especificação de Documentos (FDM); - CERIF XML Exchange Specification Document, - CERIF XML Schemas for XML Validation, - CERIF XML Example Files, - CERIF Semantics.
CERIF	<ul style="list-style-type: none"> - Modelo Introdução e Especificação de Documentos

1.3	(FDM); - CERIF XML Exchange Specification Document, - CERIF XML Schemas for XML Validation, - CERIF XML Example Files, - CERIF Semantics.
CERIF 1.4	- Permite uma nova estrutura para os arquivos XML e é baseado no 1.3.
CERIF 1.5	- Atualização e ampliação substancialmente do vocabulário CERIF (atualmente cerca de 450 termos possuem descrições, definições, exemplos, referências de fontes e esquemas contextualmente rotulados); -A atualização traz uma nova entidade para identificadores federados e permite a uma pessoa estrutura de múltipla nomeação.
CERIF 1.6	- Esta versão do CERIF acomoda metadados de conjuntos de dados de pesquisa; - A extensão adiciona uma entidade multi-lingual; -A forma mais adequada de utilizar a informação, agora, está usando uma classificação ou identificadores federados.
CERIF 2.0	- Prioridades das próximas etapas: dados da pesquisa, impacto de contexto, geolocalização, classificações de assunto, vocabulários (instalações, equipamentos, serviços), mapeamentos para formatos de repositórios mais utilizados, metadados restritivos.

Fonte: Baseada em: <<http://www.eurocris.org>>.

É possível afirmar que a evolução apresentada pelo modelo de dados CERIF consta em sua estrutura de dados formais que fornecem a integridade das informações e impossibilita redundâncias dos valores de atributos; “no uso de relações *n:n* permitindo declarar o papel e a duração temporal dos projetos; na preservação das características individuais de cada sistema e em sua essência multilíngue” (SAYÃO; SALES, 2014, p. 88). Em concordância com os autores citados, vale ressaltar que o CERIF tem a pretensão de identificar as entidades e suas respectivas relações, possibilitando uma rede de informações “acopladas”.

Existem também outros padrões que estão sendo utilizados pelo CRIS em todo o mundo, como padrões de modelos de dados, padrões de identificadores únicos e padrões de dicionários de dados comuns.

NL CERIF

No início de 2012, foi solicitado o desenvolvimento de um modelo de padrão de dados holandês para informações sobre pesquisa. Logo após, um grupo de trabalho liderado pelo SURF¹⁴ com especialistas em informações de pesquisa e arquitetura de informação, desde então têm trabalhado na definição deste padrão. Este estudo resultou no NL CERIF, perfil de aplicação holandês do padrão internacional CERIF.¹⁵

As universidades e institutos de pesquisas holandeses contam com diferentes softwares de sistemas de informação sobre pesquisa, como o Converis e o Pure que serão tratados na próxima seção. Assim, o NL CERIF foi desenvolvido para acomodar diversas linhas de informações sobre pesquisas e, por isso, é recomendado para as instituições holandesas que desejarem trocar informações sobre pesquisas. Contudo, é importante ressaltar que o NL CERIF não prescreve como a informação pesquisada deve ser organizada.¹⁶

CASRAI

O Consortia Advancing Standards in Research Administration Information (CASRAI) é uma organização internacional sem fins lucrativos dedicada a reduzir a carga administrativa e aprimorar a medição do impacto. A sua abordagem consiste na melhoria do fluxo de informações dentro e entre as partes interessadas na pesquisa. Desenvolvem e mantêm acordos padrões entre as diversas partes interessadas na pesquisa. Os padrões formam pontes nos compartilhamentos dos processos de trabalho¹⁷.

¹⁴ Disponível em: < <https://www.surf.nl/en>>

¹⁵ Disponível em: < <https://wiki.surfnet.nl/display/nlcerif/Home>>

¹⁶ Disponível em: <https://www.surf.nl/en/themes/research/research-information/research-information-systems-cris/index.html>>

¹⁷ Disponível em: < <http://reconnect.casrai.org/15/#schedule>>

CASRAI está construindo uma rede global de organizações sem fins lucrativos de pesquisas que colaboram para padronizar acordos em informação portátil para um compartilhamento mais fácil. Para isso, o CASRAI reúne uma série de atividades em torno de áreas temáticas-chave de interesse para as organizações produtoras de pesquisas e dos seus parceiros.¹⁸

Os acordos de informação padrões do CASRAI são curados e mantidos em um dicionário on-line aberto disponível no link: <http://dictionary.casrai.org/>.¹⁸

ORCID

ORCID é uma organização sem fins lucrativos de âmbito internacional e interdisciplinar. Pretende resolver o problema da ambiguidade dos nomes nas pesquisas e nas comunicações acadêmicas através da criação de um registro central de identificadores únicos para os pesquisadores individuais, e de um mecanismo de vinculação aberto e transparente entre o ORCID e outros esquemas atuais de identificação pesquisadores. Estes identificadores e suas respectivas relações podem ser ligados à saída do pesquisador para melhorar o processo de descoberta científica e a eficiência do financiamento da pesquisa e colaboração dentro da comunidade de pesquisa.¹⁹

Tem como missão fornecer um registro de identificadores para os investigadores, garantir a adoção destes identificadores, trabalhando com a comunidade de pesquisa para incorporá-los em fluxos de trabalho de pesquisa tais como pedidos de concessão e manuscrito. E, como objetivo, a adoção internacional dos identificadores ORCID pela comunidade de pesquisadores - pesquisadores e organizações envolvidas no apoio à pesquisa – com o propósito de conectar pesquisa e pesquisadores, questões de ambiguidade com o sobrenome e transformar a pesquisa e o ecossistema acadêmico (ORCID, 2013).

ORCID é um esforço para criar e manter um registro de identificação exclusivo dos pesquisados e um método transparente das atividades de pesquisas e saídas para esses identificadores. O ORCID é único em sua capacidade de alcance e através das disciplinas, setores de pesquisa e das fronteiras nacionais se configura no eixo que liga os pesquisadores e a pesquisa através da incorporação de

¹⁸ Disponível em: < <http://reconnect.casrai.org/15/#schedule>>

¹⁹ Disponível em: < <http://orcid.org/about>>

identificadores do ORCID em fluxos de trabalho essenciais, como a manutenção de perfil de pesquisa, submissões de artigos e pedidos de patente.²⁰

O ORCID apresenta duas funções principais:

- ✓ Um registro para obter um identificador único e gerenciar um registro das atividades e,
- ✓ APIs (Interface de Programação de Aplicação) que suportam comunicação sistema-a-sistema e autenticação.

Seu código fica disponível sob uma licença de código aberto, e publica um arquivo de dados públicos, anualmente sob a licença Creative Commons para download gratuito.²⁰

Os registros do ORCID armazenam informações não-confidenciais, como nome, email, organização e atividades de pesquisas. ORCID entende como fundamental a necessidade que os indivíduos possuem de controlar a forma como os seus dados são compartilhados e, por isso, fornece ferramentas de gerenciamento de privacidade dos dados.²⁰

A subseção a seguir tratará dos softwares disponíveis para implantação de sistemas CRIS. Existem os softwares comerciais, os softwares livres e o desenvolvimento *in-house* que serão apresentados abaixo.

6.6 SOFTWARES DISPONÍVEIS: EXEMPLOS PRINCIPAIS E CARACTERÍSTICAS

A fim de serem criados os CRIS das mais diversas naturezas foram elaborados softwares que apoiam a idealização desses projetos. Esses softwares podem ser comerciais, baseados no Dspace ou de desenvolvimento próprio (*in-house*).

²⁰ Disponível em: < <http://orcid.org/about>>

6.6.1 PACOTES COMERCIAIS

PURE

O software Pure facilita a abordagem baseada na comprovação para pesquisa da instituição e estratégias de colaboração, exercícios de avaliação e decisões da instituição do dia a dia. Agrega informações sobre pesquisa da instituição a partir de fontes internas e externas e certifica a confiabilidade, a abrangência e acessibilidade dos dados em tempo real. Este software ainda permite a elaboração de relatórios, avaliações de desempenho, gerenciamento de perfis dos pesquisadores, habilita a rede de pesquisa. Todas essas tarefas podem ser realizadas ao mesmo tempo o que reduz os gastos administrativos²¹.

O Pure faz parte das Soluções de Inteligência de Pesquisa Elsevier. Esse sistema de gerenciamento de informações de pesquisa é administrado pelo Elsevier²¹. O Elsevier organizou em 2015 a sua primeira Conferência Internacional em Amsterdã. É centralizado e muito versátil e possui características como²²:

- ✓ Captura dados em todas as pesquisas do mundo;
- ✓ Diminui os encargos administrativos;
- ✓ Atesta que as informações sobre pesquisas são válidas e confiáveis;
- ✓ Utiliza o serviço de Refinamento de Perfis, prestado pelo software, cria perfis super detalhados dos pesquisadores;
- ✓ Identifica especialistas no assunto;
- ✓ Produz e compartilha relatórios;
- ✓ Apresenta declarações de avaliações nacionais.

Ao todo existem pelo mundo 200 sistemas que utilizam o software PURE, incluindo os que estão em desenvolvimento. Esses portais incluem perfis de pesquisadores que, disponíveis na web, possibilitam aos pesquisadores e suas publicações que sejam “encontrados” fora de suas redes pessoais, ou seja, cria oportunidades para os próprios pesquisadores e para as instituições das quais fazem parte. São 160.000 perfis de pesquisadores dispostos em 29 países²¹.

²¹ Disponível em: < <https://www.elsevier.com/solutions/pure>>

²² Disponível < <http://www.eurocris.org/news/elsevier-pure-conference-amsterdam>>

Seguem alguns exemplos de CRIS que utilizaram o software Pure e seus respectivos países²³:

- ✓ University of Western Australia – Austrália
- ✓ Korea University – Coreia
- ✓ University College Ghent - Bélgica
- ✓ Instituto Politécnico Nacional – México
- ✓ University of Amsterdam - Holanda
- ✓ Universidad de Santiago de Chile- Chile
- ✓ Universidade Nova De Lisboa – Portugal
- ✓ Ural Federal University - Rússia
- ✓ National Taiwan Normal University- Taiwan
- ✓ New York University - Estados Unidos
- ✓ Kumamoto University – Japão

O Pure fornece muitas opções de configurações para atender as necessidades das instituições. A partir de um mecanismo de gerência de informações de pesquisas, o mesmo sistema é apto a beneficiar diferentes usuários de maneiras distintas de acordo com sua função. Ao bibliotecário, garante a boa qualidade dos dados dentro do sistema através de gestão de fluxo de trabalho, agrega a visibilidade do conteúdo mediante integração com o sistema de bibliotecas, amplia a consciência da urgência em publicações de acesso aberto; para os gestores de pesquisas, cria e publica relatórios que avaliam o desempenho dos pesquisadores, equipes, departamento, ou grupos pré-selecionados, estabelece metas utilizando indicadores de desempenho e acompanham o respectivo progresso, apoia exercícios de avaliação de pesquisa nacional; aos pesquisadores, preenche automaticamente o repositório institucional da instituição com a informação que fornece nos seus perfis, fornece informações para relatórios que certificam que a comunicação é baseada em dados atuais e completos, gerencia os perfis através de recuperação de publicação automática.²³

²³ Disponível em: < <https://www.elsevier.com/solutions/pure>>

CONVERIS

O software é extremamente integrativo, fornecendo conectores para mais de 50 sistemas diferenciados, como sistemas de bibliotecas, repositórios institucionais, recursos humanos, banco de dados dentre outros. Ao mesmo tempo, o Converis funciona como fonte de informação para serviços como OAI – Havester. O Converis sustenta completamente o ciclo de vida da atividade de pesquisa, desde a submissão de projetos até os resultados das pesquisas e as inovações por estas permitidas. Conseqüentemente, isso proporciona benefícios importantes para os principais interessados ao longo do ciclo de vida da pesquisa²⁴.

Esse software foi fabricado e é gerenciado por Thomson Reuters²⁴, uma agência de notícias multinacional. Pode ser caracterizada como a maior fonte mundial de informação inteligente para empresas e profissionais.

As tarefas são realizadas de uma forma sistemática, através de fluxos de trabalho estruturados no Converis, quer seja através de participação direta ou através de integração com sistemas internos e externos, reutilizando informações no intuito de economizar custos e evitar que os pesquisadores consultem várias vezes a mesma informação. Do mesmo modo que o software Pure, o Converis dispõe de fluxos de trabalhos que são personalizáveis através do módulo de configuração para corresponder ao contexto da instituição e as suas necessidades específicas. Esse módulo de configuração permite a organização de uma interface de usuário, onde todos os elementos-chave do sistema podem ser personalizados para as necessidades individuais. As alterações nos módulos podem ser efetuadas a qualquer momento, o que possibilita um baixo custo e o atendimento às solicitações urgentes. Desse modo, é relevante mencionar a necessidade desse papel pró-ativo em iniciativas como CERIF para garantir que Converis esteja à frente das mais recentes iniciativas e padrões.²⁵

Converis torna possível a criação de vários relatórios que reúnam dados e requisitos através da combinação de integração de dados e análise

²⁴ Disponível em: < <http://thomsonreuters.com/en.html>>

²⁵ Disponível em : <<http://converis.thomsonreuters.com/>>

de negócios. Fornece o estado da arte da interface do usuário projetado especificamente para o relatório e a apresentação de tabelas, gráficos e painéis e, exporta dados em HTML, Excel, CSV, PDF e RTF.²⁶

6.6.2 SOFTWARE LIVRE

DSPACE

A Universidade de Hong Kong, em conjunto com o consórcio italiano CINECA, incorporou entidades CERIF ao DSpace o que possibilitou maior amplitude de representação. Esse trabalho colaborativo tornou-se em CRIS de fonte aberta que vem sendo utilizado por várias instituições de pesquisas (CASTRO, 2014).

O DSpace-CRIS estendeu o modelo de dados do DSpace dada a capacidade de gerenciar, coletar e exibir dados sobre todos os aspectos da pesquisa. Esse software ainda é capaz de gerir as entidades do CRIS, integrar entidades Dspace – CRIS com publicações, fornecer estatísticas de utilização, métricas no nível do artigo e garante de interoperabilidade.²⁷

6.6.3 DESENVOLVIMENTO IN-HOUSE

Os softwares que são de desenvolvimento próprio, ou seja, in-house, são softwares desenvolvidos somente para/naquela instituição onde há foco na necessidade e na missão de maneira específica para aquela organização.

Esse tipo de software pretende cumprir requisitos de excelência em pesquisa para uma instituição em específico. Geralmente, destina-se a reunir informações de várias bases de dados.

Nesse estudo, pode-se exemplificar que desenvolvimentos de softwares in-house podem ocorrer tanto em repositórios quanto em CRIS, por exemplo:

²⁶ Disponível em : <<http://converis.thomsonreuters.com/>>

²⁷ Disponível em : <<http://cineca.github.io/dspace-cris/index.html>>

- ✓ UCARIS - University for the Creative Arts – CRIS

6.7 PRINCIPAIS EXPERIÊNCIAS

EUROCRIS

Informalmente, o EuroCRIS existe desde a primeira Conferência Internacional CRIS que foi realizada em 1991. Tal diretório surgiu nos anos noventa, como uma plataforma, com o propósito de promover uma melhor prática do CRIS. Em 2002, o EuroCRIS foi refundado e, em 2004, registrado como uma organização sem fins lucrativos (ASSERSON; JEFFERY, 2009).

Até o ano de 2009, o EuroCRIS possuía mais de 100 membros. Ainda hoje, oferece uma conferência bienal, organiza seminários estratégicos todos os anos, bem como promove reunião dos países membros a cada dois anos (ASSERSON; JEFFERY, 2009). Dedicar-se ao desenvolvimento de Sistemas de Informação de Pesquisa e sua interoperabilidade, gerencia o portfólio das atividades de pesquisa e desenvolvimento de uma agência de financiamento ou instituição de pesquisa. Possibilita pesquisas em currículos e sites, portfólios de projetos de pesquisas, pesquisa e unidades bibliográficas, pesquisas especializadas, oportunidades de inovação, dentre outros.

Por conseguinte, oferece serviços que compartilham experiências e consultorias entre os membros em geral, mantém o CERIF e o DRIS - Diretório do CRIS - proporciona atividades conjuntas de pesquisa e desenvolvimento, oferece a Conferência Bienal do CRIS, disponibiliza reuniões semestrais de membros, seminários estratégicos anuais, workshops e notícias frequentes, oferta estratégias e infraestruturas.

De acordo com Laliou (2010), as áreas de competências do EuroCRIS são:

- ✓ Bases de dados de pesquisa e temática;
- ✓ Dados relacionados ao CRIS: conjuntos de dados científicos, repositórios institucionais, sistemas de coleta de dados assistida e sistemas de fluxo de

trabalho baseados em processos;

- ✓ Sistemas de manutenção e difusão de CERIF para acomodar a implementação e a interoperabilidade dos CRIS;
- ✓ Acesso a dados e taxas de mecanismos, normas e diretrizes e melhores práticas para os CRIS.

Em consequência do acesso à informação das pesquisas correntes de toda a Europa, o EUROCRIS tem sido essencial para o sucesso da política de inovação da União Europeia. Esta política de inovação tem como principal objetivo melhorar e reforçar a geração e exploração dos novos e dos correntes projetos de pesquisa e desenvolvimento, bem como a transferência de tecnologia. Ideias inovadoras de pesquisa e desenvolvimento podem surgir graças a uma abordagem de partilha de informação e intercâmbio para outros pesquisadores que acessam esses dados. Um “conjunto de CRIS” pode cooperar com essa tarefa de compartilhamento, por ser um mecanismo de informação muito poderoso de disseminação da informação, mas é preciso que a interoperabilidade seja alcançada para seu completo aproveitamento. Por tal motivo, desponta a necessidade de uma ferramenta que interopere essas informações (ZIMMERMAN, 2002). Essa ferramenta de interoperação é chamada de CERIF e será tratada na próxima seção.

NORWEGIAN SOCIAL SCIENCE DATA SERVICES (NSD)

De acordo com as informações no Site do NSD, essa base de dados é um centro de recursos que auxilia os pesquisadores em relação à coleta de dados, análise de dados, questões de metodologia, de privacidade e de pesquisa. Tem como principal objetivo melhorar as possibilidades e as condições de trabalho de pesquisa empírica que dependem, principalmente, do acesso aos dados. Desta forma, procura trabalhar para reduzir as barreiras financeiras, técnicas, jurídicas e administrativas entre usuários e recursos de dados²⁸.

Na homepage do NSD, averiguamos que o sistema possui parcerias com o Research Council of Norway, Statistics Norway e a The Norwegian Data Inspectorate que tem sido de suma importância para as condições favoráveis de acesso aos

²⁸ Disponível em: < <http://www.nsd.uib.no/nsd/english/>>

dados disponibilizados. O esquema Oficial de Proteção de Dados sugere que a exigência para a obtenção de licenças da Inspeção de Dados para a maior parte de projetos de pesquisa, seja substituída por uma notificação, em que o NSD é a última instância para analisar os pedidos de certificados²⁹.

Na NSD, os dados são divididos em quatro categorias²⁹:

✓ Dados Individuais: Informação sobre pessoas, através de pesquisas ou registros. Todos os dados disponíveis nessa coleção são disponibilizados aos pesquisadores anonimamente e possuem informações fornecidas pelos próprios indivíduos em pesquisas ou recolhidos a partir de registros. Os temas variam entre condições de vida, saúde, trabalho e uso de meios de comunicação para as eleições, democracia e valores sociais.

Os arquivos da NSD contêm dados de vários registros públicos, consultados do Sistema Nacional de Segurança, do Regime Geral GP, e outros censos nacionais.

As coleções de dados são divididas e subdivididas em:

- NSD data – Pesquisa Norueguesa;
- Poll Archive;
- Estudos do Curso de Vida (LOGG);
- Registro de Dados (Welfare Data; Base de Dados Geração NSD's; Banco de dados Census NSD's);
- Dados Multinacionais (Portal CESSDA; Eurobarômetros; Pesquisa Social Europeia (ESS); Programa de Pesquisa Social Internacional (ISSP); Saúde das Crianças em Idade Escolar (HBSC); Valor do Estudo).

✓ Dados Regionais: São constituídos pela descrição de unidades geográficas e pelos dados agregados às pessoas dentro dessas determinadas áreas, incluindo estatísticas de demografia, emprego, desenvolvimento etc.

As coleções de dados regionais são divididas em:

- Guia de Dados “Macro”;
- Base de Dados de Eleições Europeias;
- Base de Dados Regional da NSD;
- Base de Dados Eclesiástica;

²⁹ Disponível em: < <http://www.nsd.uib.no/nsd/english/>>

- Dados Constituintes;
- Dados Nórdicos;
- Inovação Norueguesa.

✓ Dados do Sistema Político (PolSys): Contêm informações sobre instituições, pessoas e recursos na área de serviço político e civil norueguês.

As coleções de dados do Sistema Político são seccionadas da seguinte forma:

- Parlamento;
- Governo;
- Partidos;
- Função Pública – Base e Dados da Administração Estatal Norueguesa;
- Outros Arquivos.

✓ Dados das Instituições: Descrevem variados tipos de organizações. Dividem-se nas seguintes coleções:

- Banco de Dados para Estatísticas do Nível Superior;
- Banco de Dados Eclesiástico;
- Inovação Norueguesa.

Os Softwares desenvolvidos pela NSD oferecem aos pesquisadores ferramentas que auxiliam o acesso e facilitam a análise. Os softwares são³⁰:

✓ NESSTAR - É um sistema para publicação de dados na web. Permite que o pesquisador analise e publique dados na web, criando um sistema de conhecimento para tomada de decisões, se baseando nas evidências. Lida com dados de pesquisa e tabelas multidimensionais;

✓ NSDstat – Sistema de entrada de dados que trata de dados discretos e regionais. Possui características como dados de representação gráfica, ponderação, recodificação, univariação, bivariação e multivariação;

³⁰ Disponível em: < <http://www.nsd.uib.no/nsd/english/>>

✓ CELIUS – Permite que o pesquisador estude as tendências políticas, analisando sistemas eleitorais e eleições.

NATIONAL ACADEMIC RESEARCH AND COLLABORATIONS INFORMATION SYSTEM (NARCIS)

O NARCIS é o principal site da Holanda para quem procura informações sobre pesquisas e seus respectivos autores. Promove acesso à informação científica, publicações dos repositórios de todas as universidades holandesas (incluindo livre acesso às teses eletrônicas de todas as universidades holandesas), instituições científicas, institutos de pesquisa, informações de arquivos de dados, descrições de projetos de pesquisa, pesquisadores e institutos de pesquisa³¹.

O NARCIS é utilizado por pesquisadores, estudantes, jornalistas e outros que trabalham em instituições educacionais, governamentais e setores empresariais³¹.

Segundo Ivanovic, Ivanovic e Surla (2012), o NARCIS fornece acesso a todas as informações de pesquisa científica da Holanda. E Joint (2008) garante que esse sistema foi criado com o intuito de aumentar a visibilidade e recuperação das pesquisas científicas holandesas.

Esse sistema de informações de pesquisa utiliza um identificador digital para autores a fim de identificar exclusivamente um pesquisador, e esse identificador, também, é utilizado pelas universidades holandesas em várias situações, mesmo quando não relacionadas com a produção científica e tecnológica. (PINTO; SIMÕES; AMARAL, 2014).

Os metadados das publicações e das bases de dados do NARCIS podem ser reutilizados em outros serviços. Estas informações podem ser recolhidas através do protocolo OAI-PMH. E é certificado que as publicações submetidas à NARCIS, com resumos e versões de texto completo são fornecidas para o Google Scholar. Também é possível encontrar todos os dados do NARCIS no Google.

Na página inicial do NARCIS, é possível encontrar a informação de que este sistema, ainda, não pode ser utilizado como meio para inclusão de sínteses

³¹ Disponível em: < <http://www.narcis.nl/search/coll/publication/genre/doctoralthesis/Language/en>>

completas das publicações dos pesquisadores. Entretanto, nacionalmente, é possível agregar os dados das publicações, partindo dos sistemas acadêmicos³².

A próxima seção abordará o papel significativo dos bibliotecários e das bibliotecas na gestão de informação para a pesquisa. Do mesmo modo, questionará o que esperar dos bibliotecários nessa constante evolução tecnológica e de que maneira eles podem ser necessários na gestão de informação para a pesquisa.

³² Disponível em: < <http://www.narcis.nl/search/coll/publication/genre/doctoralthesis/Language/en>>

7. O PAPEL DAS BIBLIOTECAS DE PESQUISA NA GESTÃO DE INFORMAÇÃO PARA A PESQUISA

Como as novas tecnologias digitais têm transformado continuamente a produção de conhecimento científico e a comunicação, a biblioteca não é mais o único fornecedor de acesso à informação. Hoje em dia, os professores, alunos e pesquisadores escolhem, normalmente, as ferramentas de busca da internet. Muitas vezes essas fontes não possuem a mesma qualidade e confiabilidade do material pesquisado na biblioteca, mas são muito mais baratas.

Quando a tecnologia evolui despontam novos mecanismos para auxiliar na disseminação e no acesso à informação científica. Esses novos mecanismos podem criar serviços de informação nunca antes imaginados pelos bibliotecários ou tendem a ocupar algumas funções que outrora eram realizadas pela biblioteca tradicional. Ao mesmo tempo, a comunidade científica transforma-se em um responsável pela criatividade e inovação, quando propõe novas ideias de serviços que auxiliem os pesquisadores, possibilitando o surgimento de novos padrões para a área. Como afirmado por Sales e Sayão (2015b), pode-se certificar que a área de informação para a pesquisa é aprimorada mediante influência de seus próprios pesquisadores, como, por exemplo, o padrão OAI-PMH que surgiu da urgência de disponibilização de forma eletrônica de *preprints* e foi proposta e desenvolvida por pesquisadores.

Alguns materiais que não são comumente fornecidos pelas bibliotecas de pesquisas, como, por exemplo, materiais baseados em novos tipos de mídias são muito importantes na publicação de resultados da pesquisa, pois geram maior acesso às publicações. Contudo, esses materiais são uma ampliação dos papéis da biblioteca de pesquisa. Tais conteúdos multimídias dominam a publicação de resultados de pesquisa e as relações de compartilhamento de informações e de pesquisas colaborativas que se realizam no ambiente definido pelas ciberinfraestruturas voltadas para a e-Pesquisa (DEFF, 2009).

Os pesquisadores sofrem com a constante intimidação por parte das instituições de pesquisa e agências de fomento para disponibilizar os seus resultados para um volume maior de “público estratégico”, devido à valorização do fazer científico. Isso influencia no padrão de desenvolvimento das bibliotecas de pesquisa o que gera o desafio de adaptação dos conteúdos científicos para novas audiências.

Os cientistas precisam acessar informação em ambientes virtuais adequados, tanto para grupos de pesquisa, como para pesquisas individuais. O modelo ideal de gestão e disseminação de informação para cada pesquisa na ciberinfraestrutura será definido de acordo com o desenvolvimento das atividades específicas.

A infraestrutura de pesquisa em que a biblioteca está inserida será expansiva, ubíqua e capaz de suportar uma multiplicidade de domínios acadêmicos, dada a natureza interdisciplinar da ePesquisa [pesquisa eletrônica]. A próxima geração de bibliotecas deve estar firmemente inserida nesta infraestrutura, porque esta será a plataforma na qual os acadêmicos irão ganhar acesso aos recursos informacionais de forma customizada (SALES; SAYÃO, 2015b, p.41)

Na verdade, grande parte da discussão sobre os papéis dos bibliotecários nos serviços de dados se deve à evolução da ciberinfraestrutura e da *e-Science* (GOLD, 2007). Da mesma forma, realizar gestão e curadoria de um grande volume de dados se assemelha com as funções habituais da biblioteca de pesquisa, pois a ciência tem utilizado cada vez mais os dados em suas hipóteses. Ao gerir e curar dados, a biblioteca científica obtém a chance de expandir seus serviços.

Assim, no futuro, todos os bibliotecários podem ter um papel ampliado nos serviços de informação antes do ciclo da pesquisa, auxiliando com o processo de documentação e certificando-se de que os dados serão preservados, utilizáveis e reutilizáveis para o longo prazo. Entretanto, uma vez que as bibliotecas e os bibliotecários compreenderem quão importante é a oportunidade de integração de dados em biblioteconomia, irão precisar investir tempo no desenvolvimento de novas habilidades e no cruzamento de fronteiras culturais. (GOLD, 2007).

A próxima seção dissertará sobre os principais padrões utilizados nos sistemas CRIS.

8 GUIA DE SISTEMATIZAÇÃO DA LITERATURA DO MODELO CRIS: CONCEITOS, TECNOLOGIAS, PADRÕES E APLICAÇÕES

Essa seção pretende demonstrar os resultados alcançados mediante a sistematização da literatura analisada na dissertação.

Antigamente os dados eram artefatos sem importância para as pesquisas, pois estavam sempre em segundo plano. Entretanto, atualmente, os dados são dignos e tão substanciais quanto a pesquisa em si; por isso, precisam ser armazenados em sistemas que forneçam o aporte necessário, tanto para sua recuperação quanto para possível reutilização pelos cientistas. Pois, publicar esses dados já não é mais suficiente. Atualmente, existe a urgência de esses dados serem armazenados, tratados, descritos e publicados, o que quer dizer; é necessário um sistema que interligue todos esses processos em uma só plataforma. Sem esquecer a importância do profissional da informação que frente à e-Science coopera com o desenvolvimento de ferramentas e técnicas de representação da informação e se responsabiliza pela gestão de dados produzidos pela e-Science (SALES, 2014).

A fim de que isso seja possível existe o processo da curadoria digital que é responsável pela gestão desses dados o que possibilita a disponibilização e reutilização deles. Além disso, através da interoperabilidade, tais dados podem ser transportados de um sistema para o outro, método esse que facilita a análise feita pelo pesquisador que não precisará acessar vários sistemas para conseguir as informações que lhe são necessárias, ou seja, terá a facilidade de adquirir todas as informações em um mesmo sistema e economizará tempo para produzir mais questões relevantes.

À vista de tudo que foi investigado durante a pesquisa, de toda a capacidade de suporte e de gestão do CRIS juntamente com todo o poder de interoperabilidade do CERIF, permite-nos certificar que esse sistema e seu padrão de interoperabilidade são, nos dias de hoje, o conjunto de excelência e eficácia no que tange à perspectiva da pesquisa científica.

Conforme mencionado no histórico desta dissertação, embora sejam raríssimas as iniciativas de criação de um CRIS para o Brasil, existem dois sistemas que poderiam interoperar com o CRIS caso houvesse um projeto de desenvolvimento deste: a Plataforma Lattes e o Sucupira. Em contrapartida, alguns pesquisadores tem se inquietado tanto com essa carência de sistemas gestores da

informação que sejam capazes de integrar as publicações científicas e seus respectivos dados de pesquisas, quanto com uma ciberinfraestrutura que produza uma composição entre os repositórios digitais e os CRIS. Essas iniciativas são relevantes para todo o Brasil no que tange à evolução da pesquisa científica.

Ao realizar um panorama mundial dos sistemas de informação para pesquisas percebeu-se que a necessidade de um CRIS é em razão da importância desse sistema na tomada de decisões da gestão das atividades de pesquisas, na divulgação dos resultados destas e no suporte que fornece nos processos de criação e gestão do conhecimento. Por isso, os mecanismos de busca e pesquisa da Web, por mais potentes que sejam, não substituem o CRIS. O CRIS ocupa um nível de completude relacionada às informações sobre pesquisa nunca antes alcançado por nenhum outro sistema que fornecesse informações sobre pesquisas. O CRIS é capaz de suprir necessidades de pesquisadores que antes deveriam recorrer a vários sistemas diferentes que não eram interoperáveis; com o CRIS, o pesquisador economiza tempo e dinheiro ao realizar sua pesquisa em um único lugar.

Com isso, universidades, instituições e organizações de todo o mundo estão recorrendo a novos softwares (PURE, CONVERIS, DSpace-CRIS) que lhes proporcione um sistema desse porte que facilite tanto a vida dos funcionários quanto dos usuários do sistema. Portanto, os CRIS estão surgindo sobre várias naturezas; agora existem não só os CRIS institucionais, mas também os temáticos que abrangem um determinado tema de inúmeras instituições, o que acaba permitindo que uma instituição que ainda não tenha seu próprio CRIS não seja excluída desse avanço tecnológico, ao menos algumas de suas disciplinas.

Como produto final, foram desenvolvidos alguns quadros com as principais definições, os tipos de sistemas CRIS, os principais usuários, os principais sistemas nacionais, regionais e internacionais, os principais softwares utilizados, os principais padrões utilizados pelos sistemas CRIS e os principais autores nacionais e internacionais. Esses quadros seguem abaixo com suas respectivas explicações:

1 . DEFINIÇÕES

DEFINIÇÃO	FONTE
<p>“CRIS é um modelo de dados que descreve um conjunto de objetos de interesse para as atividades de pesquisa e uma série de ferramentas que possibilitam ao usuário (pesquisador, gestor etc.) a gestão de seus dados de pesquisa em todos os processos, incluindo alocação de recursos, avaliação de projetos, identificação de novos mercados para produtos de pesquisa, análise de tendências entre outros serviços.”</p>	<p>SAYÃO; SALES / Instituição: CNEN</p>
<p>O CRIS é um sistema que visa proporcionar uma única fonte de informações de pesquisa que possibilita o planejamento, monitoramento, relatórios, comunicação e modelagem futura de atividades de pesquisas.</p>	<p>BEVAN; HARRINGTON / Instituição: Cranfield University</p>
<p>Um sistema de gestão de dados que possui a capacidade de integrar dados de vários outros sistemas.</p>	<p>QUIX; JARKE / Instituição: Fraunhofer-Institute for Applied Information Technology - FIT</p>
<p>Sistema de informação que gerencia todas as informações relevantes de uma pesquisa, desde as oportunidades de financiamento, da escrita e apresentação das etapas de propostas, até a conclusão do processo de pesquisa.</p>	<p>JOINT/ Instituição: University of Strathclyde</p>

Essas definições foram as que melhor representaram a completude de um CRIS ou foram as definições mais citadas, como, por exemplo, as definições de Joint e de Bevan e Harrington.

Resumidamente, o CRIS pode ser definido como uma base de dados que é responsável pelo armazenamento e pela gestão de dados sobre pesquisa de uma instituição, de um país, de uma região ou de uma determinada disciplina ou, como qualquer ferramenta de informação, dedicada a fornecer o acesso e a disseminar

informações sobre pesquisa, tais como pessoas, projetos, organizações, resultados (publicações, patentes e produtos), instalações e equipamentos.

2. TIPOS DE SISTEMAS CRIS

TIPOS	EXEMPLOS
Institucional	HKU Scholars Hub – Disponível em: < http://hub.hku.hk/ >
Nacional	SICRIS - Disponível em: < http://www.sicris.si/public/jqm/cris.aspx?lang=slv&opdescr=home&opt=1 > NARCIS - Disponível em: < http://www.narcis.nl/ >
Regional e Internacional	Socionet – Disponível em: < https://socionet.ru/ > IST World - Disponível em: < http://www.ist-world.org/ >
Por área de conhecimento ou por disciplina	Human Nutrition Research Information Management – Nutrição - Disponível em: < http://www.niddk.nih.gov/research-funding/Pages/default.aspx > CRIS of National Institute of Food and Agriculture (NIFA) – Agricultura - Disponível em: < http://cris.nifa.usda.gov/ >

Conforme demonstrado no corpo da dissertação, essa é a classificação mais utilizada nos textos que foram selecionados. Organizam o CRIS em: Institucional, Nacional, Regional e Internacional e Por Área de Conhecimento ou por Disciplina.

3. PRINCIPAIS USUÁRIOS

USUÁRIOS	BENEFÍCIOS
Gestores de Pesquisa	Permite o acesso a indicadores e análise detalhada de atividades de pesquisa, ao currículo dos pesquisadores, a informações comparativas sobre desempenho e instituições

	científicas e de áreas de pesquisa; <i>input</i> para desenvolvimento de estratégias.
Órgãos de Divulgação	Garante o acesso a informações que podem ser reformuladas em matérias que popularizem a pesquisa científica ou que sugiram questões de caráter social, ético, político ou econômico e pertençam a níveis da atividade de pesquisa que são de mais óbvio interesse público.
Pesquisadores	Facilitam ao pesquisador em suas anotações, nas apresentações de relatórios e nas tomadas de decisões relativas à pesquisa a ser realizada. É fundamental na avaliação das oportunidades de financiamento da pesquisa, reprime a duplicação de pesquisa, analisa as tendências, compila referências e links para um texto completo ou publicações acadêmicas multimídias, identifica novos contatos e, novos mercados para os produtos da pesquisa. Fornece acesso aos resultados de pesquisa e informações sobre oportunidades de parceria, currículos, empregos prêmios, bolsas de estudo e financiamento de pesquisa, instalações, equipamentos, laboratórios, projetos, eventos e pacotes de softwares; oferece informações relevantes, autênticas e confiáveis sobre particularidades do mundo da pesquisa científica.
Professores e Estudantes	Permite o acesso a materiais que apoiam o ensino de ciências.
Empresários	Possibilita o acesso a informações que levem a novas ideias que serão trabalhadas para a criação de riqueza ou melhoria da qualidade de vida; informações sobre contratos e patentes.
Instituição	Funciona como um instrumento na formulação de políticas, produz uma lista de pesquisas em andamento e mensura as pesquisas através do documento que registra atividades com base nas produções dos trabalhos.

Formuladores de política de Ciência, Tecnologia e Informação	Arrola instrumentos que correlacionam os produtos produzidos por diferentes continentes, países, instituições e grupos de pesquisa no intuito de auxiliar a concepção de marcos políticos e legais que dão suporte às atividades de pesquisa.
Cidadão comum	Incentiva o acesso a interfaces de sistemas voltados para o registro, processamento e exploração de resultados atualizados de atividades de pesquisa que refletem na vida social e na qualidade de vida.
Agências Financiadoras	Disponibiliza uma ferramenta importante para a melhoria e otimização do processo de avaliação de projetos; proporciona acesso a estudos que realizam comparação em relação a outras agências e o acompanhamento dos projetos em desenvolvimento.

O CRIS proporciona atendimento a um grande volume de usuários, sejam estes da comunidade científica ou não. Isso traz muitos benefícios para a sociedade, pois o acesso a esse sistema reflete na vida social e na qualidade de vida do indivíduo que o acessa.

Em suma, o CRIS propicia aos pesquisadores: acesso fácil à informação relevante, aos gestores de pesquisa e administradores: fácil medição e análise da atividade de pesquisa; aos conselhos de pesquisa: a otimização do processo de financiamento; aos empreendedores e instituições de transferência de tecnologia: uma fácil recuperação de ideias inovadoras e tecnologia em um ambiente assistida por conhecimento e facilita a identificação dos concorrentes e, anteriormente, feita uma pesquisa semelhante; à mídia e ao público: acesso fácil à informação, software e poder do computador para permitir a apresentação facilmente assimilada dos resultados da investigação em contextos apropriados.

4. PRINCIPAIS SISTEMAS NACIONAIS, REGIONAIS E INTERNACIONAIS

SISTEMA	TIPO	URL
EuroCRIS	Diretório Regional	http://www.eurocris.org/
NSD	Área de conhecimento	http://www.nsd.uib.no/nsd/english/
NARCIS	Nacional	http://www.narcis.nl/search/coll/publication/genre/doctoralthesis/Language/en

Esses sistemas foram selecionados para os resultados finais por estarem sempre presentes nas literaturas escolhidas para a referência bibliográfica desta dissertação, se tornando assim, representativos em suas categorias.

5. PRINCIPAIS FERRAMENTAS DE SOFTWARES UTILIZADOS

SOFTWARE	FORNECEDOR/URL
LIVRES DSPACE	http://cineca.github.io/dspace-cris/index.html
COMERCIAIS PURE	https://www.elsevier.com/solutions/pure
CONVERIS	http://converis.thomsonreuters.com http://thomsonreuters.com/en.html

É importante acentuar que o CRIS é um sistema que pode ser desenvolvido de várias maneiras. E, no decorrer da pesquisa, se pode perceber que além da criação desse sistema com características *in-house*, é possível que softwares sejam utilizados para auxiliar em sua origem.

Além de softwares comerciais, como o PURE e o CONVERIS, ao longo da pesquisa se percebeu que existe um software livre para criação de CRIS. Esse software é muito conhecido na esfera biblioteconômica por causa dos Repositórios

Institucionais; contudo, ele foi adaptado para o ambiente CRIS. O DSpace-CRIS estendeu o modelo de dados do DSpace e é capaz de gerir as entidades do CRIS.

6. PRINCIPAIS PADRÕES UTILIZADOS PELOS SISTEMAS CRIS

NORMA/PADRÃO/PROTOCOLO	URL
Dublin Core	http://dublincore.org/
OAI-PMH	https://www.openarchives.org/pmh/
OAI-ORE	https://www.openarchives.org/ore/
CERIF	http://eurocris.org/cerif/main-features-cerif
NL CERIF	https://www.surf.nl/en
CASRAI	http://reconnect.casrai.org/15/#schedule
ORCID	http://orcid.org/about

Existem alguns padrões que são adotados para a troca de informações entre sistemas de informações na internet. Os principais são Dublin Core, OAI-PMH e OAI-ORE. Essas normas são relevantes tanto para repositórios institucionais quanto para CRIS.

Contudo, há algumas normas específicas para o CRIS que são classificadas como padrões de modelos de dados, padrões de identificadores únicos e padrões de dicionários de dados comuns. O CERIF e o NL CERIF são padrões de modelos de dados que garantem a interoperabilidade entre os sistemas. O CASRAI é um padrão de dicionários de dados comuns e o ORCID é um padrão de identificador único.

Vale ressaltar que o principal padrão é o CERIF. Este é o mais citado em artigos por autores nacionais, como Sales e Sayão, e internacionais, como Zimmerman e Bevan e Harrington. O CERIF é o melhor padrão de compartilhamento de dados de pesquisas que proporciona a interoperabilidade entre os CRIS e sistemas associados.

7. PRINCIPAIS AUTORES NACIONAIS E INTERNACIONAIS

AUTOR		REFERÊNCIA
NACIONAIS	SAYÃO, Luis Fernando	SAYÃO, Luís Fernando; SALES, Luana Farias. Ciberinfraestrutura de informação para a pesquisa : uma proposta de arquitetura para integração de repositórios e sistemas CRIS. 2015.
	SALES, Luana	SAYÃO, Luís Fernando; SALES, Luana Farias. Dados abertos de pesquisa: ampliando o conceito de acesso livre. RECIIS – Rev. Eletron. de Comun. Inf. Inov. Saúde , [s.l.], v. 8, n.2, p.76-92, jun. 2014.
	PEREIRA, Maria Nazaré Freitas Pereira	PEREIRA, Maria Nazaré Freitas Pereira. Mapa da Competência, Sistemas de Informação de Pesquisa em Processo . Nov. 2014
	PACHECO, Roberto Carlos dos Santos	PACHECO, Roberto C. S.; KERN, Vinícius Medina ; SALM JR, José Francisco ; PACKER, Abel Laerte ; MURASAKI, Renato Toshiyuki ; AMARAL, Luis ; SANTOS, Leonel Duarte ; CABEZAS, Alberto Raul Bullemore . Toward CERIF-ScienTI cooperation and interoperability. In: Anne G.S. Asserson and Eduard J. Simons (eds.) International Conference on Current Research Information Systems (CRIS) , 8., 2006, Bergen, Norway. Enabling interaction and quality: beyond the Hanseatic League. Leuven (Belgium): Leuven University Press, 2006. p. 179-188.
INTERNACIONAIS	CASTRO, Pablo	CASTRO, Pablo; SHEARER , Kathleen; SUMMANN, Friedrich. The gradual merging of repository and CRIS solutions to meet institutional research information management requirements. In: INTERNATIONAL CONFERENCE ON CURRENT RESEARCH INFORMATION SYSTEMS, 12, 2014.

		Proceedings... Rome: euroCRIS, 2014, p. 39-46. Disponível em: < http://dx.doi.org/10.1016/j.procs.2014.06.007 >. Acesso em: 29 abr. 2015.
	ZIMMERMAN, Eric H.	ZIMMERMAN, Eric H; JEFFERY, Keith G. The Need for a Standardized Current Research Information System (CRIS): A Call to Arms. In: International Conference on Current Research Information Systems, 7., 2004, Antuérpia. Proceedings... Antuérpia: euroCRIS, 2004, p.153-160.
	JOINT, Nicholas	JOINT, Nicholas. Current research information systems, open access repositories and libraries: ANTAEUS. Library Review , [s.l.], v. 57 n. 8, p.570- 575, 2008.

No decorrer da pesquisa pode-se notar que existe uma grande variedade de artigos científicos sobre o CRIS na língua inglesa. Porém, no Brasil esse número cai drasticamente, chegando a ter quatro autores de relevância que dissertam sobre o tema, enquanto um deles só publica em inglês. E, mesmo assim, esses autores nacionais esforçam-se para manter a literatura nacional atualizada.

Todas essas tabelas construídas, os pontos mencionados e, principalmente o interesse dos grandes polos de pesquisa científicas mundiais por esses sistemas, veem corroborar a importância desses sistemas para a sociedade como um todo, inovando nas pesquisas científicas e possibilitando maior evolução para as comunidades nas quais estão inseridos.

9 CONCLUSÃO

Os CRIS são sistemas que estão em contínuo crescimento no mundo devido a sua importância na área da pesquisa científica e, conseqüentemente, para a comunicação científica. Eles possibilitam maior visibilidade à produção intelectual da instituição, da região ou do tema em questão. Tais sistemas estão em grande fase de expansão por todo o mundo, o que resulta no surgimento de diferentes pacotes de software para implantação de sistemas CRIS, tais como softwares comerciais, softwares livres e até mesmo desenvolvimento *in-house*. Por este motivo, são necessários padrões que permitam a interoperação desses sistemas; caso contrário, eles não se comunicariam entre si o que geraria um grande transtorno para a comunidade acadêmica como um todo, pois os pesquisadores retornariam ao marco zero, onde seria necessário acessar várias plataformas diferentes para conseguir determinada informação de que precisam.

A Biblioteconomia e a Ciência da Informação têm grande importância na concepção, desenvolvimento e operação dos sistemas CRIS, pois conduzem, em suas estruturas, princípios da organização e da gestão de informação em qualquer que seja o suporte. A essência da biblioteconomia permite o melhor direcionamento e administração dos CRIS através de suas subáreas, como a classificação, representação do conhecimento, intercâmbio e recuperação de informações, formação de coleção e descrição de metadados. Dessa forma, os pressupostos fundamentais da Biblioteconomia têm uma contribuição significativa no desenvolvimento dos sistemas de informação para a pesquisa e para a integração desses sistemas nas infraestruturas de pesquisa e informacionais das instituições acadêmicas.

Da mesma maneira, o CRIS contribui com a Biblioteconomia ao expandir suas fronteiras de pesquisa, permitindo que a Biblioteconomia ganhe espaço no ambiente virtual onde seus conceitos são rerepresentados, expandidos, valorizados e enriquecidos, mediante as inúmeras oportunidades advindas do surgimento da internet e das novas tecnologias que interagem diretamente com a pesquisa científica de modo geral. Ao ampliar o seu interesse para essas novas concepções de sistemas de informação, a Biblioteconomia e a Ciência da Informação definem novos espaços de investigação, ampliando a sua agenda de pesquisa com visões mais contemporâneas sobre os problemas de informação.

Do ponto de vista de afirmação profissional nesse novo universo, os sistemas de informações sobre pesquisas auxiliam aos profissionais da informação em seu trabalho de recuperação da informação. E, em contraponto, os profissionais de Biblioteconomia e da Ciência da Informação gerenciam esses sistemas, organizando as informações, disponibilizando as publicações completas, seus dados e metadados de maneira organizada e recuperável. Pois, atualmente, a questão que tem sido adversa não é a falta de informação e sim, o grande volume de informação e de dados reproduzidos diariamente o que gera a necessidade que profissionais da informação, devidamente preparados, estejam liderando todo o processo de tratamento e gestão dessas informações. A falta de gestão de tais informações pode gerar a não recuperação dessas informações, a não reutilização de dados, a duplicação de informações já desenvolvidas e de dados já tabulados o que, conseqüentemente, pode resultar no gasto desnecessário de fundos.

Muitas vezes esse profissional da informação é fundamental na concepção de um CRIS, pois trabalha em conjunto com o olhar crítico do pesquisador. Por este motivo seria de suma importância que os alunos de biblioteconomia, os futuros gestores da informação, soubessem da relevância desses sistemas desde o início de seus estudos na faculdade, onde estudariam como manipular a informação nos CRIS, como melhorar a usabilidade desses sistemas e, por que não avaliá-los? Tal medida facilitaria sua vida profissional e, principalmente, seriam capacitados a ajudar os pesquisadores que utilizam esses sistemas.

Por ser uma questão tão importante para a Biblioteconomia e para a ciência, os CRIS deveriam ser acompanhados, de forma cuidadosa, pelas escolas de biblioteconomia, tanto na graduação quanto na pós-graduação. É urgente a incorporação do estudo do histórico e do estado da arte dos CRIS nos currículos e nas agendas de pesquisa das escolas de Biblioteconomia, no intuito de que haja a observação e a assistência da evolução dos sistemas de informação sobre pesquisa pelos futuros profissionais da informação. Haveria o enriquecimento e a troca de conhecimentos preciosos, tanto quando os alunos estudassem sobre os CRIS, quanto quando os alunos contribuíssem para o desenvolvimento dos CRIS, através de suas questões e inquietações científicas.

Os profissionais de informação possuem a oportunidade de aumentar, de maneira virtual, os acervos da biblioteca e expandir a abrangência os seus serviços devido à possibilidade de obter informação livre e de qualidade na internet. Porém,

para tanto, a biblioteca deve dar coerência às fontes de informação que estão subdivididas na internet, integrando-as ao fluxo local onde serão configuradas de maneira unificada e particularizada. Ou seja, a Biblioteconomia encontra-se em uma encruzilhada onde deve associar os novos serviços para tratar das novas estratégias intelectuais de pesquisa, integrando os novos sistemas de gestão de pesquisa na sua função na pesquisa científica e, ao mesmo tempo, deve se conservar na fronteira de sua área de competência.

Com os pensamentos anteriores, é simples afirmar que os serviços tradicionais da biblioteca talvez percam o valor, porém os serviços oferecidos *on-line* tem se tornado a cada dia mais relevantes e requisitados para os pesquisadores. Isso sugere que por mais que os serviços tradicionais sejam ocultados pelo tempo, quanto mais a tecnologia evoluir mais serão necessários os serviços dos profissionais da informação para gerir a informação virtual e integrar os usuários aos ambientes virtuais.

Durante a produção do trabalho, foi possível obter um alto nível de conhecimento sobre o CRIS que embasou todo o mapeamento das definições, dos tipos de sistemas CRIS, dos principais usuários, dos principais sistemas nacionais, regionais e internacionais, das ferramentas de softwares predominantes que são utilizadas para o desenvolvimento de um CRIS, dos principais padrões utilizados pelos sistemas CRIS, dos autores indispensáveis à leitura nos âmbitos nacionais e internacionais realizado pela autora. Desta maneira, acredita-se que o levantamento elaborado foi de grande relevância para a comunidade científica da Biblioteconomia e da Ciência da Informação e conclui-se que o propósito inicial desta dissertação de compilar padrões, tecnologias, aplicações e conceitos do CRIS foi alcançado, visto que a autora deixa como contribuição tal compilação de informações que está inserida nos quadros do guia de sistematização da literatura, na análise realizada com base nos textos e na crítica da carência de produção científica sobre o tema em questão.

Acrescenta-se que foi importante todo o conhecimento adquirido com todo o referencial teórico utilizado. Cada autor contribuiu de maneira única para a seleção de informações realizada nessa dissertação.

Com o propósito de seguir a linha de raciocínio trabalhada no decorrer da dissertação, almeja-se dar continuidade a este projeto através do desenvolvimento

de uma pesquisa baseada na infraestrutura da integração do Sistema Maxwell com o CRIS.

9.1 TRABALHOS FUTUROS

O sistema Maxwell foi desenvolvido no LAMBDA - Laboratório de Automação de Museus, Bibliotecas Digitais e Arquivos. O foco do laboratório é aplicar tecnologia da informação na gestão da informação. Suas atividades são subdivididas em: desenvolvimento de sistemas e desenvolvimento de arquivos multimídias para alimentar o banco de dados.

É possível que portadores de deficiência visual e cegos acessem o repositório, que já está na quarta versão. Para isso são utilizadas ferramentas de manipulação das páginas e uma programação que pode ser interpretada pelo TTS (Text-to-Speech). Por ser um repositório de uma instituição privada, o acesso público é restringido e o proprietário original que gerencia o conteúdo passa o controle direto do material a mãos de terceiros.

No entanto, após o averiguado durante a análise do presente trabalho, se verificou que os repositórios institucionais, embora completos no que se propõem, são eficazes e eficientes para outro propósito no auxílio da pesquisa científica. Assim, foi dado o início ao projeto de interoperação do Sistema Maxwell com o CRIS.

O projeto visa à integração desses dois sistemas em que seria permitido que aos usuários acessarem a CRIS e o Sistema Maxwell ao mesmo tempo, sem precisar iniciar o *login* em dois sistemas diferentes ao mesmo tempo. Tal iniciativa facilitaria as análises dos pesquisadores da instituição e dos profissionais de informação que gerenciam o sistema.

Contudo, com a pesquisa já iniciada percebeu-se que é necessário a interoperação com outros sistemas da PUC-Rio, tanto de sistemas administrativos que tratam de históricos, matrículas, currículos e outros, como o sistema da biblioteca física da PUC-Rio, pois isso facilitaria ao usuário quando precisasse de alguma publicação que fosse de domínio exclusivo da biblioteca ou, quando necessitasse de um número de chamada. A ideia é de interoperar os sistemas da PUC-Rio tanto educacionais quanto administrativos.

O Sistema Maxwell já utiliza padrões internacionais como o Dublin Core e o *Open Archives* o que facilita a interoperação. Atualmente, publica materiais como dados de pesquisa, vídeos que explicam experimentos e vídeos-aula o que tem ampliado a variedade e a qualidade dos materiais que disponibiliza.

Com essa visão, a coordenadora do projeto Maxwell iniciou o planejamento de integração do Sistema Maxwell com o CRIS. Nesse intuito, reuniu sua equipe, da qual a autora faz parte, para dar início ao projeto que já está em evolução. Tal plano, seu desenvolvimento e seus respectivos resultados serão alvo de estudo de futuras pesquisas.

REFERÊNCIAS

ABBOTT, Daisy. **What is digital curation?** Edinburgh, UK : Digital Curation Centre, 2008. Disponível em: <<http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/what-digital-curation>>. Acesso em: 20 jun. 2014.

AMORIM, Fabiana Borelli; TOMAÉL, Maria Inês. Gestão da informação e gestão do conhecimento na prática organizacional: análise de estudos de casos. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v8, n. 2, p. 01-22, jan./jun. 2011.

ASSERSON, Anne; JEFFERY, Keith G. Current Research Information Systems (CRIS): Past, Present and Future. **Wissenschaftsmanagement**, v. 15, n. 1, p. 41-44, 2009. Disponível em: <http://www.wissenschaftsmanagement.de/dateien/dateien/schwerpunkt/downloaddateien/wiw_2009_01_anne_asserson_keith_g_jeffery_current_research_information_systems_chris_papa_present_future.pdf>. Acesso em: 12 nov. 2014.

ATKINS, D.E., DROEGEMEIER, K.K., FELDMAN, S.I., GARCIA-MOLINA, H., Klein, M.L., MESSERSCHMITT, D.G., MESSINA, P., OSTRIKER, J.P.; WRIGHT, M.H. **Revolutionizing Science and Engineering through Cyberinfrastructure**: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. 2003. Disponível em:< www.nsf.gov/od/oci/reports/CH1.pdf>. Acesso em 20 out. 2015.

BARRETO, Auta Rojas. A informação para a agro-industria. **Revista AIBDA**, v. 2, n. 2, julho – diciembre 1981, p. 101- 121.

BENTANCOURT, Silvia Maria Puentes; ROCHA, Rafael Port da. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação. v. 17, n. Esp. 2, 2012.

BETTENCOURT, Marcia Pires da Luz; CIANCONI, Regina de Barros. Gestão do conhecimento: um olhar sob a perspectiva da ciência da informação. **Tendências da Pesquisa Brasileira em Ciência da Informação**, Vol. 5, No 1 (2012).

BEVAN, Simon J.; HARRINGTON; John. Managing research publications: lessons learned from the implementation of a Current Research Information System. **Serials: The Journal for the Serials Community**, [s.l], v. 24, n. 1, p.26-30, mar. 2011. Disponível em: < <http://uksg.metapress.com/content/y1q0ut5465708t16/fulltext.pdf>>. Acesso em: 5 jun. 2014.

BORBA, Danice Kinner; SANTOS, Luiz Felipe Almeida; KAWAMOTO JR., Luiz Teruo. Big Data: percepção dos usuários sobre vantagens e invasão de privacidade. VIII WORKSHOP DE PÓS-GRADUAÇÃO E PESQUISA DO CENTRO PAULA SOUZA, 8., 2013, p. 695-704. Disponível em:< <http://www.centropaulasouza.sp.gov.br/pos-graduacao/workshop-de-pos-graduacao-e-pesquisa/008-workshop->

2013/trabalhos/gestao_estrategica_de_tecnologia_da_informacao/121467_695_704_final.pdf> . Acesso em: 2 set. 2015

BORGMAN, Cristine. Research data: who will share what, with whom, when, and why? In: CHINA-NORTH AMERICAN LIBRARY CONFERENCE, 5.Beijing, 17 aug. 2010. Disponível em: <<http://works.bepress.com/borgman/238/>>. Acesso em: 16 jun. 2014.

BRASE, Jan; FARQUHAR, Adam. Access to research data. **D-Lib Magazine**, v. 17, n. 1/2, Jan. / Feb. 2011. Disponível em: <<http://www.dlib.org/dlib/january11/brase/01brase.html>>. Acesso em: 16 set. 2014.

CARMO, Juliana Rabelo do; PECEGUEIRO, Cláudia Maria Pinho de Abreu. University scientific knowledge organization: a case study. **BJIS**, São Paulo, v.5, n.2, p.88-98, jul./dec. 2011. Disponível em <<http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/1330/2173>> Acesso em: 10 nov. 2013.

CASTRO, Pablo; SHEARER , Kathleen; SUMMANN, Friedrich. The gradual merging of repository and CRIS solutions to meet institutional research information management requirements. In: INTERNATIONAL CONFERENCE ON CURRENT RESEARCH INFORMATION SYSTEMS, 12, 2014. **Proceedings...** Rome: euroCRIS, 2014, p. 39-46. Disponível em: <<http://dx.doi.org/10.1016/j.procs.2014.06.007>>. Acesso em: 29 abr. 2015.

CARVALHO, Maria da Conceição R. de. **O reuso da informação técnico-científica a partir de um repositório institucional (RJ): um estudo exploratório**. 2011. 101f. Dissertação (Mestrado em Ciência da Informação) – Instituto de Arte e Comunicação Social, Universidade Federal Fluminense, Niterói, 2011.

CORDEIRO, Daniel; et al. Da ciência à e-ciência: paradigmas da descoberta do conhecimento. **Revista USP**, São Paulo, n. 97, p. 71-80^a, março/abril/maio 2013.

COSTA, Maira Murrieta; CUNHA, Murilo Bastos da. O bibliotecário no tratamento de dados oriundos da e-science: considerações iniciais. **Perspectivas em Ciência da Informação**, v.19, n.3, p.189-206, jul./set. 2014.

CUNHA, Murilo Bastos da; CAVALCANTI, Cordélia Robalinho de Oliveira. **Dicionário de Biblioteconomia e Arquivologia**. Brasília, DF: Brique de Lemos/Livros, 2008.

CUNHA, Murilo Bastos da. Metodologia para estudos dos usuários de informação científica e tecnológica. **Revista de Biblioteconomia de Brasília**, Brasília, v. 10, n. 2, p. 5-19, jul./dez. 1982. Disponível em: <http://bogliolo.eci.ufmg.br/downloads/CUNHA_1982.pdf>. Acesso em: 15 nov. 2013.

DEFF. **The future of research and research library**. Copenhagen: DEFF, 2009. Disponível em: <http://www.deff.dk/fileadmin/user_upload/DEFF/publikationer/Oevrige/The_Fut

ure_of_Research_and_the_Research_Library.pdf>. Acesso em: 22 jan 2016.

ERBACH, Gregor. Data-centric view in e-science information systems. **Data Science Journal**, [s.l.], v. 5, n. 19, p. 219-222, out. 2006.

GOLD, Anna. Cyberinfrastructure, data, and libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries. **D-Lib Magazine**, v. 13, n. 9/10, 2007. Disponível em: <<http://www.dlib.org/dlib/september07/gold/09gold-pt2.html>>. Acesso em: 22 jan. 2016.

GOMES, Cristina Marques. **Comunicação científica**: alicerces, transformações e tendências. [s.l.]: LABCOM, 2013.

GOMÉZ, Maria Néida González de. As relações entre ciência, Estado e sociedade: um domínio de visibilidade para as questões da informação. **Ci. Inf.**, Brasília, v. 32, n. 1, p. 60-76, jan./abr. 2003

HEY, T.; HEY, J. E-science and its implications for the library community. **Library Hi Tech**, [s.l.], v. 24, n. 4, p. 515-528, 2006.

HORNBOSTEL, Sthefan. **From CRIS to CRIS**: Integration and Interoperability. 2006

IVANOVIC, Dragan; IVANOVIC, Lidija; SURLA, Dusan. Notes on operations: integration of a research management system and an OAI-PMH compatible ETDs repository of the university of Novi Sad, Republic of Serbia. **Library Resources & Technical Services**, [s.l.], v. 56, n. 2, p. 104-112, apr. 2012.

JACKSON, S.J., EDWARDS, P.N., BOWKER, G.C., KNOBEL, C., Understanding infrastructure: history, heuristics, and cyberinfrastructure policy, 12v, 6. Disponível em: <http://firstmonday.org/issues/issue12_6/index.html>. Acesso em: 10 nov. 2014

JOINT, Nicholas. Current research information systems, open access repositories and libraries: ANTAEUS. **Library Review**, [s.l.], v. 57 n. 8, p.570- 575, 2008.

JONES, Richard. Giving birth to next generation repositories. **International Journal of Information Management**, [s.l.], v. 27, n. 3, p. 154-158, 2007.

JÖRG, Brigitte; et al. **CERIF 2008: 1.0 Full Data Model (FDM)**: Introduction and Specification, 2009. Disponível em: <http://www.eurocris.org/fileadmin/cerif-2008/CERIF2008_1.0_FDM.pdf>. Acesso em: 13 nov. 2013.

KOVACEVIC, Aleksandar; et al. Automatic extraction of metadata from scientific publications for CRIS systems. **Program: electronic library and information systems**, [s.l.], v. 45 n. 4, p. 376-396, 2011. Disponível em: <<http://www.emeraldinsight.com/doi/pdfplus/10.1108/00330331111182094>>. Acesso em: 10 dez. 2013.

KVASNY, Lynette; RICHARDSON, Helen. Critical research in information systems: looking forward, looking back. **Information Technology & People**, [s.l.], v. 19, n. 3, p. 196-202, 2006. Disponível em:

<<http://www.emeraldinsight.com/doi/pdfplus/10.1108/09593840610689813>>. Acesso em: 10 out. 2013.

LALIEU, Harrie. **Introduction to euroCRIS: the organisation, its objectives and way of working.pptx**, 2010. Disponível em: <<http://www.eurocris.org/Index.php?page=structure&t=1>>. Acesso em 01 nov. 2013.

LEITE, Fernando César Lima; COSTA, Sely Maria de Souza. Gestão do conhecimento científico: proposta de um modelo conceitual com base em processos de comunicação científica. **Ci. Inf.** [online]. 2007, vol.36, n.1, pp. 92-107. Disponível em: <http://www.scielo.br/readcube/epdf.php?doi=10.1590/S0100-19652007000100007&pid=S0100-19652007000100007&pdf_path=ci/v36n1/a07v36n1.pdf&lang=pt>. Acesso em: 19 ago 2015.

LIMA, J. L. O.; ALVARES, L. Organização e representação da informação e do conhecimento. In: ALVARES, L. (Org.). **Organização da informação e do conhecimento: conceitos, subsídios, interdisciplinares e aplicações**. São Paulo: B4 Editores, 2012. p. 21-48.

LOPES, Elaine Cristina; VALENTIM, Marta Lúcia Pomim. Processos de gestão da informação: tratamento, recuperação e uso da informação no mercado de capitais. **Perspectivas em Gestão & Conhecimento**, João Pessoa, v. 3, n. 1, p. 157-174, jan./jun. 2013.

MARCHIORI, Patricia Zeni. Bibliotecas Digitais e Repositórios de Objetos de Aprendizagem. **Inf. & Soc.:Est.**, João Pessoa, v.22, n.2, p. 13-21, maio/ago. 2012. Disponível em:< <http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/12207/7754>>. Acesso em: 21 nov. 2013.

MARCONDES, **Carlos Henrique**. Organização e representação do conhecimento científico em ambiente web: do formato textual linear aos artigos semânticos. **Ponto de Acesso**, Salvador, V.7, n.1 ,p. 7-41, abr 2013. Disponível em: <<http://www.portalseer.ufba.br/index.php/revistaici/article/view/8134>>. Acesso em: 8 set. 2015.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Fundamentos de metodologia científica**. 5. ed. São Paulo: Atlas, 2003. 311 p.

MARTINS, Dalton Lopes; FERREIRA, Sueli Mara Soares Pinto. Proposta de metodologia de mapeamento e avaliação da produção científica da Universidade de São Paulo com foco na estrutura e dinâmica de suas redes de colaboração científica: em busca de novos modelos causais. **Liinc em Revista**, Rio de Janeiro, v. 9, n. 1, p. 181-195, maio 2013. Disponível em:<<http://revista.ibict.br/liinc/index.php/liinc/article/viewFile/513/394>>. Acesso em: 20 nov. 2013.

MEDEIROS, Jackson da Silva; CAREGNATO, Sônia Elisa. Compartilhamento de dados e e-Science: explorando um novo conceito para a comunicação científica. **Liinc em Revista**, Rio de Janeiro, v.8, n.2, p. 311-322, set. 2012.

MONS, Barend; et. al. International Research Information System: Support to Science Management. In: INTERNATIONAL CONFERENCE ON CURRENT RESEARCH INFORMATION SYSTEMS, 6., 2002, Kassel. **Proceedings...** Kassel: euroCRIS; Kassel University Press, 2002, p. 121-229. Disponível em: <http://dspacecris.eurocris.org/bitstream/11366/141/1/CRIS2002_Mons_International_paper.pdp>. Acesso em: 29 abr. 2015

MORENO, Fernanda Passini; LEITE, Fernando César Lima; ARELLANO, Miguel Ángel Márdero. 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.

MUELLER, S. P. M. A comunicação científica e o movimento do acesso livre ao conhecimento. **Ciência da Informação**, Brasília, v. 35, n. 2, p. 27-38, mai./ago. 2006. Disponível em; <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/826/667>>. Acesso em: 09 jul. 2015.

NESELLO, Priscila; FACHINELLI, Ana Cristina. Big Data: o novo desafio para a gestão. Revista Inteligência Competitiva, São Paulo, v.4, n.1, p. 18-38, jan/mr.2014. Disponível em: <<http://www.inteligenciacompetitivarev.com.br/ojs/index.php/rev/article/view/76/101>>. Acesso em: 19 ago. 2015

ORCID Annual Report – 2012. 2013. Disponível em: http://orcid.org/sites/default/files/ORCID_2012%20Annual%20Report-FINAL.pdf. Acesso em: 20 dez. 2015.

PEREIRA, Maria Nazaré Freitas Pereira. **Mapa da Competência, Sistemas de Informação de Pesquisa em Processo**. Nov. 2014

PIMENTA, Ricardo M. Big data e controle da informação na era digital: tecnogênese de uma memória a serviço do mercado e do estado. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 6, n. 2, jul./dez. 2013. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/article/view/117/159>>. Acesso em: 1º set. 2015.

PINHEIRO, Lena Vania R. Comunidades científicas e infraestrutura tecnológica no Brasil para uso de recursos eletrônicos de comunicação e informação na pesquisa. **Ciência da Informação**, v. 32, n. 3, p. 62-73, 2003.

PINTO, Carlos Souza; SIMÕES, Cláudia; AMARAL, Luís. CERIF – Is the standard helping to improve CRIS?. In: INTERNATIONAL CONFERENCE ON CURRENT RESEARCH INFORMATION SYSTEMS, 12, 2014. **Proceedings...** Rome: euroCRIS, 2014, p. 80-85. Disponível em: <http://dspacecris.eurocris.org/bitstream/11366/216/3/26_Pinto%20Sousa%2c%20Sim%2c%20Amaral_fin.pdf>. Acesso em: 16 dez. 2014.

PORÉM, Eugênia; GUARALDO, Tamara de Souza Brandão. Informação, conhecimento e comunicação em organizações do conhecimento. **DataGramZero**. v.13 n.1 fev/12. Disponível em:< http://www.dgz.org.br/fev12/Art_06.htm> . Acesso em: 9 set. 2015.

PRADO, Alysson Bolognesi; BARANAUSKAS, Maria Cecília Calani. Panorama do suporte computacional às atividades de pesquisa em universidades brasileiras: um estudo de caso. **Liinc em Revista**, Rio de Janeiro, v. 9, n. 1, p. 307-325, maio 2013.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico**: métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Novo Hamburgo: Feevale, 2013. 276 p.

QUIX, Christoph; JARKE, Matthias. Information Integration in Research Information Systems . **Procedia Computer Science**, Roma, v. 33, p. 18-24, 2014. Disponível em: http://dspacecris.eurocris.org/bitstream/11366/199/1/40_Quix_Jarke_CRIS2014_Rome.pdf . Acesso em: 10 dez. 2014.

RODRIGUES, Alessandra Pereira et al. A interoperação com repositórios digitais: protocolos e exemplos. **Ci. Inf.**, Brasília, DF, v. 40 n. 3, p.349-363, set./dez., 2011

RODRIGUES, Eloi et al. **Os repositórios de dados científicos**: estado da arte. 2010. (Relatório D-24: RCAAP). Disponível em: http://projeto.rcaap.pt/index.php?option=com_remository&Itemid=2&func=startdown&id=271&lang=pt. Acesso em: 10 maio. 2015.

ROWLEY, Jennifer. **A biblioteca eletrônica**. Brasília: Briquet de Lemos, 2002.

SALES, Luana Farias. **Integração semântica de publicações científicas e dados de pesquisa**: proposta de modelo de publicação ampliada para a área de ciências nucleares. Rio de Janeiro, 2014. Tese (Doutorado em Ciência da Informação) -- Escola de Comunicação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014.

SAYÃO, Luís Fernando. Afinal, o que é biblioteca digital? **Revista USP**, São Paulo, n.80, p. 6-17, 2008-2009. Disponível em: <http://eprints.rclis.org/14675/1/biblioteca-digital.pdf> >. Acesso em: 3 dez. 2015.

SAYÃO, Luís Fernando; SALES, Luana Farias. **Ciberinfraestrutura de informação para a pesquisa**: uma proposta de arquitetura para integração de repositórios e sistemas CRIS. 2015a.

SAYÃO, Luís Fernando; SALES, Luana Farias. Curadoria Digital: um novo patamar para preservação de dados digitais de pesquisa. **Inf. & Soc.:Est.**, João Pessoa, v.22, n.3, p. 179-191, set./dez. 2012. Disponível em: <http://www.ies.ufpb.br/ojs/index.php/ies/article/view/12224/8586> . Acesso em 20 dez. 2013.

SAYÃO, Luís Fernando; SALES, Luana Farias. Dados abertos de pesquisa: ampliando o conceito de acesso livre. **RECIIS – Rev. Eletron. de Comun. Inf. Inov. Saúde**, [s.l.], v. 8, n.2, p.76-92, jun. 2014.

SAYÃO, Luís Fernando; SALES, Luana Farias. Dados de Pesquisa: contribuição para o estabelecimento de um modelo de curadoria digital para o país. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 6, n. 1, 2013. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/article/view/102/146>>. Acesso em: 18 out. 2014

SAYÃO, Luís Fernando; SALES, Luana Farias. Há Futuro para as Bibliotecas de Pesquisa no Ambiente de e-Science?. **Informação & Tecnologia (ITEC)**: Marília/João Pessoa, 2(1): 30-52, jan./jul., 2015b.

SELJAK, Marta; SELJAK, Tomazĭ. COBISS: National Union Catalogue, online bibliography and gateway to other resources. **New Library World**, [s.l.], v.101, n. 1153, p. 12-20, 2000.

SOUZA, Beatriz Alves de. Proposta de Criação de um Repositório Institucional para o Instituto Federal de Educação, Ciência e Tecnologia da Paraíba – IFPB. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v.8, n.1, p. 66-84, jan./jun. 2012. Disponível em: <<http://rbbd.febab.org.br/rbbd/article/view/196/228>>. Acesso em: 21 nov. 2013.

SOUZA, Edivanio Duarte de; DIAS, Eduardo José Wense; NASSIF, Mônica Erichsen. A gestão da informação e do conhecimento na ciência da informação: perspectivas teóricas e práticas organizacionais. **Inf. & Soc.:Est.**, João Pessoa, v.21, n.1, p. 55-70, jan./abr. 2011. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/4039/5598>>. Acesso em: 10 set. 2015.

SOUZA, Nicolau Frederico de. O BRACARIS: a experiência brasileira de um sistema de informação sobre pesquisa corrente em agricultura. **R. Bibliotecon.** Brasília, 11 (1):85-104jan./jun. 1983.

TARGINO, M. das G. Comunicação científica: uma revisão de seus elementos básicos. **Informação & Sociedade**, [s.l.], v. 10, n. 2, p.1-27, 2000. Disponível em: <<http://www.ies.ufpb.br/ojs/index.php/ies/article/view/326>>. Acesso em 10 ago. 2015

TAURION, Cezar. **O caos conceitual e os 5 Vs do Big Data**. 2012. Disponível em <<http://cio.com.br/opiniao/2012/05/11/o-caos-conceitual-e-os-5-vs-do-big-data/>>. Acesso em: 10 ago. 2015

VALENTIM, Marta Lígia Pomim. Gestão da Informação e Gestão do Conhecimento em ambientes organizacionais: conceitos e compreensões. **Ciência da Informação**, v. 1, n. 1, 2008. Disponível em : <<http://inseer.ibict.br/ancib/index.php/tpbci/article/view/110/151>>. Acesso em: 10 set. 2015.

XEXÉO, Geraldo. Big data: computação para uma sociedade conectada e digitalizada. **Ciência Hoje**, 306., p.18-23, ago. 2013. Disponível em: <cienciahoje.uol.com.br/revista---ch/2013/306/pdf.../bigdata306.pdf/.../file>. Acesso em: 15 AGO.2015.

YAKEL, Elizabeth. Archives and manuscripts: Digital curation. **OCLC Systems & Services: International digital library perspectives**, [s.l.], v. 23, n. 4, p. 335-340, 2007. Disponível em: <<http://50.17.193.184/omeka/files/original/8f5d346c12e3f7e9f09bce5a64cf7728.pdf>>. Acesso em: 9 nov. 2013.

ZIMMERMAN, Eric H. CRIS-Cross: Current Research Information Systems at a Crossroads. In: INTERNATIONAL CONFERENCE ON CURRENT RESEARCH INFORMATION SYSTEMS, 6., 2002, Kassel. **Proceedings...** Kassel: euroCRIS; Kassel University Press, 2002, p. 11-20. Disponível em: <http://dspacecris.eurocris.org/bitstream/11366/129/1/CRIS2002_Zimmerman_CRIS-Cross_paper.pdf>. Acesso em 29 abr. 2015.

ZIMMERMAN, Eric H; JEFFERY, Keith G. The Need for a Standardized Current Research Information System (CRIS): A Call to Arms. In: International Conference on Current Research Information Systems, 7., 2004, Antuérpia. **Proceedings...** Antuérpia: euroCRIS, 2004, p.153-160.

ZUPAN, Vesna. Focusing on the human resources in academic librarianship: an outlook from Serbia. **Library Management**, [s.l.], v. 33, n. 3, p. 168-173, 2012.