

CIBERINFRAESTRUTURA DE INFORMAÇÃO PARA A PESQUISA: uma proposta de arquitetura para integração de repositórios e sistemas CRIS

Luana Farias Sales*
Luís Fernando Sayão**

RESUMO

Apesar da importância dos repositórios institucionais (RI) e dos repositórios de dados (e-repositórios) como parte essencial da atual infraestrutura mundial de informações para a pesquisa, os modelos de dados desses sistemas não comportam a complexidade e a diversidade das informações necessárias para a gestão integral de ambientes de pesquisa. Por outro lado, governos, instituições de pesquisa e órgãos de fomento estão implantando sistemas mais sofisticados conhecidos pela sigla em inglês CRIS (Current Research Information System) voltados para gerenciar a grande quantidade de metadados sobre entidades envolvidas nas atividades de pesquisa - como projetos, pesquisadores, financiamentos, publicações, patentes, laboratórios, equipamentos, currículos etc. -, bem como os relacionamentos e os fluxos de informações que se estabelecem entre elas. Nesse cenário, um desafio importante se coloca que é a integração e a sincronização dos sistemas CRIS com os RI's e os e-repositórios, de forma que se possa aumentar a amplitude de aplicações e de serviços de ambos os sistemas, tanto em âmbito administrativo como no âmbito acadêmico. Nessa direção, este trabalho propõe uma arquitetura integrada que sincronize os fluxos de informação de sistemas CRIS baseados na norma europeia CERIF (Common European Research Information Format), de RI's baseado no protocolo OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting), de e-Repósitos e de outros sistemas acadêmicos e administrativos internos e externos a uma instituição de pesquisa. Como metodologia foram examinadas as normas, padrões e as tecnologias que tornam possível a interoperabilidade desses sistemas no contexto de uma ciberinfraestrutura voltada para pesquisa.

Palavras-chave: CRIS. CERIF. Repositório institucional. Repositório de dados. Ciberinfraestrutura de pesquisa.

* Doutora em Ciência da Informação pela Universidade Federal do Rio de Janeiro, Brasil. Analista em Ciência e Tecnologia na Comissão Nacional de Energia Nuclear, Brasil.
E-mail: lsales@ien.gov.br.

** Doutor em Ciência da Informação pela Universidade Federal do Rio de Janeiro, Brasil. Analista em Ciência e Tecnologia na Comissão Nacional de Energia Nuclear, Brasil.
E-mail: lsayao@cnen.gov.br.

1 INTRODUÇÃO

A ideia predominante do que é repositório institucional (RI) está circunscrita por conjunto de ações que coletivamente visam coletar, organizar, arquivar e disseminar, na forma de base de dados, publicações acadêmicas, metadados e, em alguns casos, dados e conjunto de dados de pesquisa. Esta dinâmica tem como perspectiva a

formação e preservação da memória digital das organizações de pesquisa, o acesso livre aos seus ativos informacionais e o aumento do impacto desses ativos, traduzido em citações.

Entretanto é necessário considerar que a complexidade do ambiente da pesquisa contemporânea envolve inúmeros elementos que se relacionam entre si - como pesquisadores, projetos, infraestruturas, aportes financeiros,

instalações, para citar alguns – que para serem descritos e controlados precisam de sistemas que estendam seus modelos de dados além dos necessários para descrição de publicações acadêmicas, e demandam, por conseguinte, sistemas estruturalmente mais sofisticados do que os repositórios institucionais.

Nessa direção, indo além das funções tradicionais dos RI's, os sistemas de gestão de informações sobre pesquisa conhecidos pela sigla em inglês CRIS (Current Research Information System) coletam um espectro amplo de metadados que descrevem, idealmente, todas as entidades e relações envolvidas nos fluxos de trabalho que caracterizam as atividades de uma instituição científica, incluindo as interações com os seus *stakeholders* mais próximos, como, por exemplo, os órgãos de fomento à pesquisa. Dessa forma, o CRIS gerencia informações que permitem aos seus usuários registrar, reportar e tomar decisões em relação aos processos envolvidos no ciclo de gestão de pesquisa e da inovação, como estudos comparativos, produtividade, métricas e avaliações, acompanhamento de projetos, resultados de pesquisa, oportunidade de empregos, patentes e muitos outros. Um resultado importante da plataforma CRIS está expresso por maiores níveis de visibilidade sobre as atividades institucionais de pesquisa para os vários segmentos interessados, incluindo o cidadão comum.

Para cumprir seus objetivos, um sistema CRIS integra dados de uma grande variedade de sistemas, internos e externos à instituição. A partir desse ponto surgem indagações importantes sobre como os RI's podem sincronizar suas funções com as desses sistemas de gestão de informação sobre pesquisa de forma que as sobreposições e diferenças possam ser exploradas para a formação de infraestruturas informacionais de maior alcance. Para intercâmbio de informações os sistemas CRIS adotam preferencialmente a norma elaborada pela EuroCRIS¹ e recomendada pela União Europeia conhecida pela sigla CERIF (Common European Research Information System)². Esta norma estabelece um modelo de dados que

descreve, no nível conceitual e lógico, o domínio da pesquisa, relação entre as entidades desse domínio e como essas relações variam no tempo; e no nível físico, uma formalização do domínio na forma de *scripts* para sistemas de banco de dados (RUSSELL, 2010).

Mesmo considerando que os RI's e os sistemas CRIS têm objetivos que os diferenciam conceitualmente e funcionalmente, e que, além do mais, possuem trajetórias evolutivas que são independentes (CASTRO, 2014), a possibilidade de integração dos sistemas RI e dos sistemas CRIS, na direção de uma única infraestrutura institucional de gestão de informações sobre pesquisa, configura um cenário rico de novas demandas e uso para os repositórios institucionais, redimensionando o papel desses sistemas. Essas novas possibilidades de exploração dos RI's – que vão além do seu papel de base de dados de memória institucional – reforçam o seu potencial como ferramenta de apoio à governança das atividades de gestão de pesquisa e de inovação e os colocam como parte dos ambientes tecnologicamente avançados e altamente integrados voltados para e-pesquisa, conhecidos como ciberinfraestrutura. Nessa direção, o presente estudo examina as possibilidades de integração de RI's aderentes ao protocolo de interoperabilidade OAI-PMH³, de repositórios de dados e de sistemas CRIS que atendam a camada semântica estabelecida pela norma CERIF, oferecendo como resultado uma arquitetura que sincronize os fluxos de informação desses sistemas operando no contexto de uma ciberinfraestrutura voltada para a pesquisa.

A discussão sobre a integração e interoperabilidade CRIS/RI tem sido colocada em pauta nos principais fóruns sobre gestão de pesquisa, constituindo um polo de intensos debates que envolvem fortemente a área de Ciência da Informação.

À guisa de metodologia, o estudo examinou algumas experiências descritas pela literatura e analisou os principais conceitos necessários à contextualização do problema de integração CRIS/RI, como: sistemas de gestão da informação sobre pesquisa, o modelo de dados do CERIF, ciberinfraestrutura de

¹ EuroCRIS <<http://www.eurocris.org>> é uma organização sem fins lucrativos que atua no desenvolvimento de sistemas de informação sobre pesquisa e na interoperabilidade entre esses sistemas no âmbito da Comunidade Europeia.

² CERIF: <<http://www.eurocris.org/cerif/main-features-cerif/>>

³ OAI-PMH – sigla em inglês para Open Archives Initiative Protocol for Metadata Harvesting; protocolo voltado para interoperabilidade entre repositórios digitais. <<http://www.openarchives.org/pmh/>>

pesquisa e, de forma privilegiada, diferenças e semelhanças entre os dois sistemas. Como parte dos resultados - e também como ponto de partida para novas discussões -, é proposto uma possível arquitetura de integração CRIS/RI que explora complementaridade entre os dois conceitos de sistema de informação para a pesquisa.

2 A RELEVÂNCIA DO PROBLEMA: UMA PONTE ENTRE DOCUMENTOS E DADOS EM PROL DA GESTÃO DA INFORMAÇÃO SOBRE PESQUISA

O efeito direto que as atividades científicas têm sobre o progresso econômico, tecnológico e social e, como resultado mais perceptível, sobre a qualidade de vida do cidadão comum, implica na necessidade de intervenção e de suporte governamental para que estas atividades sejam sustentáveis (PINTO; SIMÕES; AMARAL, 2014). Uma parcela das ações do governo está centrada na criação de serviços de informação como suporte tecnológico e gerencial para a ciência, tecnologia e inovação (SAYÃO; SALES, 2013). No Brasil, o Portal de Periódicos CAPES⁴ e a Plataforma Lattes⁵ são exemplos claros dessas ações.

Em comum com outros empreendimentos estrategicamente importantes para a sociedade, a pesquisa científica está se tornando altamente dependente de sistemas de informação que apoiem a gestão do complexo fluxo de informação que ela própria gera nos seus processos de negócio e também dos fluxos externos que ela tem que consolidar e agregar. Sistemas que respondam com confiabilidade e qualidade o gerenciamento de informações sobre a pesquisa se tornam cada dia mais imprescindíveis para as instituições envolvidas com atividades científicas, seja na qualidade de geradores de conhecimento - como as universidades e institutos de pesquisa - ou na qualidade de gestores, financiadores ou formuladores de políticas para as áreas científicas.

A complexidade das relações entre os diversos atores e produtos de pesquisa que compõem os cenários de geração de conhecimento científico, aliada à necessidade

de respostas mais adequadas às indagações de pesquisadores, administradores de organizações de pesquisa, das empresas de tecnologia e de agências de financiamento, só para citar alguns atores, demandam sistemas de informação que sejam projetados para se adequarem aos novos modos de fazer ciência. Espera-se que essas plataformas se integrem aos ambientes de e-pesquisa, que são baseados fortemente em "sistemas computacionais, sistemas de armazenamento, instrumentos avançados e repositórios de dados, ambientes de visualização e pessoas, todos "linkados" por redes de alto desempenho que tornam possível inovação e descobertas, que de outra forma não seriam possíveis" (INDIANA UNIVERSITY, 2014).

Apesar do avanço dos repositórios institucionais como parte integrante da infraestrutura mundial de informação científica e de seu papel como memória digital das instituições de pesquisa, a orientação dessas bases de dados ainda se volta para a ideia de documento. Os repositórios digitais, em geral, entregam como resultado de uma consulta, ainda sob uma ótica tradicional, itens de informação individuais, que desconsideram a riqueza das relações entre os diversos produtos de pesquisa e a multiplicidade de atores envolvidos nos fluxos de geração de conhecimento científico. De uma forma geral os repositórios estão isolados e não suficientemente conectados aos outros sistemas da instituição (SHEPPARD, 2010), como por exemplo, os sistemas de gestão de informações sobre pesquisa, recursos humanos, currículos e outros sistemas corporativos, ou, ainda desconsideram sistemas externos que controlam ativos importantes, como a Biblioteca Digital de Teses e Dissertações (IBICT/BDTD)⁶, no caso do Brasil.

Entretanto, os RI's arquivam metadados e publicações que são fontes relevantes para os fluxos de trabalho de pesquisa. Esses estoques informacionais, se organicamente articulados e contextualizados, podem intensificar a sua utilidade nas esferas gerenciais e administrativas das instituições envolvidas no trabalho de pesquisa científica. O que se espera é que os repositórios possam trocar informações com outras plataformas ampliando o seu alcance como fonte de informação e como provedor

4 Portal de Periódicos CAPES: <<http://www.periodicos.capes.gov.br/>>

5 Plataforma Lattes: <<http://lattes.cnpq.br/>>

6 Disponível em: <<http://bdt.d.ibrict.br/>> Acesso em 20 mar. 2015.

de novos serviços. Nessa direção, os sistemas de gestão de informação sobre pesquisa, pela sua proximidade e complementariedade em relação aos RI's, se mostram adequados para a formulação de infraestruturas informacionais mais completas e contextualizadas.

Os sistemas voltados para gestão de pesquisa, em particular os que se dedicam à gestão de informação sobre pesquisa, vêm, historicamente, se ajustando às novas demandas impostas pela ciência contemporânea, na medida em que registram, processam e apresentam, na forma de dados abertos, métricas e cifras relacionadas ao ciclo de vida das atividades da pesquisa científica. Os indicadores e métricas que são registrados pelos sistemas oferecem informações de qualidade e contextualizadas sobre os produtos de pesquisa – por exemplo, publicações, patentes, dados de pesquisa e serviços – e sobre suas fontes – pesquisadores, instituições, projetos financiados, e-infraestrutura, entre outras. O uso dos sistemas de informação com essas características tem se expandido, adotando vários nomes, com sucesso em âmbito institucional, nacional e, em escala regional, principalmente na Europa, liderado pela Comunidade Europeia, constituindo-se uma ferramenta importante para a gestão, distribuição e exploração e contextualização de resultados de pesquisa

Atestada a importância dessas plataformas informacionais como ferramentas importantes para o mundo da pesquisa, fica o desafio para várias áreas, incluindo a Ciência da Informação, de como explorar as diferenças e sobreposições desses sistemas sincronizando os seus processos em configurações e arquiteturas que resultem em serviços mais avançados e mais próximos aos pressupostos de uma ciência mais aberta e colaborativa.

Apesar da alegada dicotomia entre os repositórios institucionais e os sistemas de gestão de informação sobre pesquisa, há um reconhecimento de que o parceiro ideal para os repositórios é, de fato, o CRIS. “Aparentemente ambos os sistemas gerenciam os mesmos dados, para propósitos análogos e para os mesmos usuários-finais” (SHEPPARD, 2010). Para tal, ambos devem se integrar apropriadamente aos fluxos das atividades de pesquisa entregando serviços específicos, como, por exemplo, arquivando versões em texto completo de

artigos científicos, para o caso dos repositórios, e documentando informações sobre financiamento da pesquisa que gerou o artigo, para o caso do CRIS. Poucas coisas impedem que esses serviços sejam enriquecidos mutuamente via troca de informações.

Como atestam Castro e seus colaboradores (2014, p.39), o panorama geral tende para um estado em que os sistemas de gestão de pesquisa integrem ambos os sistemas. “A interoperabilidade CRIS/RI é uma característica que está se generalizando e que vai permitir que ambas as plataformas intercambiem informações de forma mais eficiente reforçando as características do que cada um é e faz [...]”, mesmo considerando que as fronteiras que os delimitam se tornam cada vez mais tênues.

São várias as possibilidades de integração CRIS/RI, mas para entender melhor as possibilidades de interoperabilidade entre esses sistemas é desejável analisar com um grau a mais de profundidade as características dos sistemas CRIS, seu escopo de atuação e seus benefícios, e, ainda, o ambiente tecnológico onde eles se situam.

3 CRIS: SISTEMA DE GESTÃO DE INFORMAÇÃO SOBRE PESQUISA

Qualquer tentativa de gerenciar o ciclo completo do processo de pesquisa – conforme nos alerta Joint (2008) – é por definição uma experiência bastante ambiciosa, posto que uma parte essencial da complexidade desse empreendimento é vislumbrar o fluxo global de informações que permeia a totalidade das atividades do mundo científico de uma forma simples e coerente. Isto implica projetar sistemas de informação que sejam capazes de apoiar todos os aspectos informacionais importantes para os processos de gestão da pesquisa.

A amplitude das demandas impostas sobre os sistemas de gestão de informações sobre pesquisa pode ser qualitativamente dimensionada pelo levantamento realizado pelo Imperial College London⁷ e Elsevier (2010). Segundo essa fonte, as instituições e seus pesquisadores e gestores necessitam de informações que: apoiem a descoberta de

7 Disponível em: <<http://www.imperial.ac.uk>> Acesso em 20 mar. 2015

oportunidades de financiamento; facilitem o cálculo de custo de desenvolvimento de uma pesquisa; monitorem a aplicação dos investimentos e a taxa de sucesso de uma pesquisa; facilitem a agregação e a avaliação padronizada dos resultados e benefícios da pesquisa, incluindo publicações, patentes e licenças; a gestão de fundos já concedidos; promovam o aumento da visibilidade das atividades de pesquisa individuais e institucionais; assistam os pesquisadores na descoberta de oportunidades de colaboração, especialmente em áreas interdisciplinares, transversalmente a departamentos da instituição e com pesquisadores de outras instituições; ajudem as instituições a rastream e identifiquem oportunidades de colaboração com empresas, órgãos governamentais e outras instituições de pesquisa; facilitem o desenvolvimento de atividades empresariais; identifiquem talentos fora da instituição para potencial recrutamento para o mundo acadêmico; facilitem o planejamento de cenários futuros e identificação de mudanças significativas. Outros itens podem ser alinhados a essa lista.

Os sistemas de informação que têm como princípio de funcionamento a visão totalizante, que podem ser caracterizadas pelas demandas acima, são geralmente conhecidos por CRIS ou Research Information System⁸ (RIM). De uma forma abrangente um sistema CRIS é um banco de dados que armazena metadados sobre atividades de pesquisas de interesse corrente, como interpreta Lambert (2010), e gerenciam de maneira integrada as informações sobre o ciclo de vida das atividades de pesquisa, sobre as entidades que estão envolvidas e as relações que se estabelecem entre elas, por exemplo: pessoas (pesquisadores, administradores, etc.); organizações (universidades, centros de pesquisa, órgãos de fomento, etc.); instalações (laboratórios, equipamentos, etc.); produtos de pesquisa (dados, publicações, patentes, etc.); projetos e outras entidades que são protagonistas da dinâmica do universo da pesquisa científica (SALES, 2014).

O resultado mais evidente da implantação de sistemas CRIS é o grande aumento da visibilidade das atividades de pesquisa da instituição para os diversos segmentos de

interessados: de pesquisadores a gestores, passando por formadores de opinião e as mídias de divulgação científica voltadas para o cidadão comum e para estudantes e professores. Pode-se incluir também como motivação uma carteira de serviços que apoia a conformidade e proporciona o aumento da capacidade interna de reportar, analisar, avaliar e entregar informações de forma mais precisa, integrada e organizada.

Não obstante as atuais e intensas discussões em torno dos sistemas de gestão de informações sobre pesquisa, colocadas principalmente pelas condicionantes da ciência contemporânea e suas inúmeras articulações, o conceito subjacente à sigla CRIS, não é, como se poderia esperar, algo exatamente novo no mundo da gestão da Ciência, Tecnologia e Inovação (CT&I). A importância cada vez mais imperativa para as infraestruturas de gestão de pesquisa de gerir todo o processo de pesquisa como uma sequência de eventos de informação é uma ideia que tem uma história de 40 anos de desenvolvimento, como lembraram Jeffery e Asserson (2009, p.74). Ainda que os sistemas CRIS tenham sido identificados no passado – e mesmo no presente – por outros nomes, essencialmente, eles estão baseados nos mesmos princípios funcionais dos atuais sistemas. Isto reforça o fato de que conceitos tais como “ambientes virtuais de pesquisa” e “sistemas de gestão de pesquisa” também evocam a ideia de que gestão de informação sobre pesquisa, são, em alguma medida, sinônimos de CRIS ou incluem o conceito de CRIS (JOINT, 2008, p.571).

3.1 Usuários e benefícios

Um dos pressupostos importantes dos sistemas CRIS, conforme já visto, é a abrangência e capacidade de filtrar e agregar informações, traduzidos em serviços, que podem atingir um amplo espectro de usuários que não estão restritos somente ao universo científico, como visto a seguir:

- **Pesquisadores** – acesso a informações relevantes, autênticas e confiáveis sobre vários aspectos do mundo da pesquisa científica; acesso aos resultados de pesquisa (publicações e dados de pesquisa) e informações sobre oportunidades de parceria, currículos, empregos, prêmios,

8 Sistema de Gestão de Informação sobre Pesquisa (tradução nossa)

bolsas de estudo e financiamento de pesquisa, instalações, equipamentos, laboratórios, projetos, eventos e pacotes de *software*.

- **Gestores de pesquisa** - acesso a indicadores e análise detalhada de atividades de pesquisa, a informações bibliométricas e comparativas sobre desempenho de instituições científicas e de área de pesquisa; input para desenvolvimento de estratégias e para o planejamento; acesso ao currículo dos pesquisadores e para avaliação do nível individual de produtividade.
- **Agências de financiamento á pesquisa** - constitui uma ferramenta importante para melhoria e otimização do processo de avaliação de projetos e de instituições de pesquisa candidatas a financiamento; permite acesso a estudos comparativos em relação a outras agências e o acompanhamento dos projetos em desenvolvimento.
- **Formuladores de política de CT&I** - constitui instrumento para confrontar os produtos produzidos por diferentes continentes, países, instituições e grupos de pesquisa, para subsidiar a formulação de marcos políticos e legais que dão sustentação às atividades de pesquisa de um país ou região.
- **Empreendedores e inovadores** - acesso a informações que levem a novas ideias que podem ser exploradas para a criação de riqueza ou melhoria da qualidade de vida; identificação de competidores e parceiros que trabalham em projetos correlatos; informações sobre contratos e patentes.
- **Professores e estudantes** - acesso a materiais que possam apoiar o ensino de ciências, como imagens e materiais multimídias e simulações.
- **Órgãos de divulgação** - acesso a informações que podem ser reformatadas em matérias que popularizem a pesquisa científica ou que suscitem questões de caráter social, ético, político ou econômico concernentes a segmentos da atividade de pesquisa que são de mais óbvio interesse público.
- **Cidadão comum** - acesso a interfaces de sistemas voltados para registro,

processamento e exploração de resultados atualizados de atividades de pesquisa que têm efeitos sobre a prosperidade social e sobre a qualidade de vida em contextos mais sensíveis à sociedade como, por exemplo, saúde, meio-ambiente e educação.

Para atender a essas demandas e usos, os sistemas de gestão de informação sobre pesquisa precisam integrar dados provenientes de diversas fontes e sistemas, internos e externos à instituição. Tipicamente, as instituições de pesquisa gerenciam as informações sobre o ciclo de vida da pesquisa em seus domínios extraindo-os, de forma transversal, de seus próprios departamentos acadêmicos e administrativos. Entretanto, dados externos precisam ser integrados, notadamente os dados sobre oportunidades de financiamento e sobre publicações acadêmicas (DEMPSEI, 2014).

Essas exigências de integração das informações sobre publicações implicam necessidade de sincronização dos sistemas CRIS com os repositórios institucionais e repositórios de dados, tudo isso em conformidade com as políticas internas de acesso aberto. A visão integrada que esta sincronização oferece é precisamente o objeto do presente estudo.

3.2 Uma tentativa de classificação

Considerando a abrangência geográfica dos sistemas CRIS, Pinto e seus colaboradores (2014) estabelecem uma categorização que ajuda a compreender o escopo de ação desses sistemas no contexto dos diversos níveis de infraestruturas de pesquisa e a sua pluralidade de concepções. Nessa direção temos: CRIS institucional que inclui informações sobre pesquisas circunscritas a uma única instituição; CRIS nacional que gerencia informações científicas, tecnológicas, de inovação ou curriculares de várias instituições pertencentes a um país; CRIS regional e internacional que lida com informações sobre pesquisa de países, de uma região ou de um grupo de países reunidos por algum princípio econômico, cultural ou proximidade geográfica.

Num outro eixo pode-se considerar o escopo temático dos sistemas CRIS: há os sistemas que atuam sobre diversas áreas temáticas; em

contraste, há os que se dedicam a uma área de pesquisa específica, como agricultura ou saúde. Além do mais, há sistemas que focalizam sua atenção em segmentos específicos, como projetos ou currículos.

3.3 Ferramentas de software para implementação de sistemas CRIS

De acordo com DEMPSEY (2014), várias ferramentas de *software* surgiram no mercado para assistir a implantação de ambientes CRIS nos últimos anos. Isto deixa patente a demanda crescente, em escala mundial, por produtos comerciais voltadas para a gestão de informações sobre pesquisa. Esta demanda é inflada por uma política agressiva de expansão do mercado por parte dos tradicionais fornecedores de *software* para a área de informação, que reposicionam seus investimentos para um florescente segmento de mercado. O Pure da Atira-Elsevier⁹ e o Converis¹⁰ da Avedas-Thomson Reuters são os principais pacotes de *software* comerciais para CRIS que se enquadram nessa situação.

Porém, em paralelo, muitas instituições acadêmicas e de pesquisa estão optando por desenvolver soluções baseadas em pacotes de *software* originalmente destinados a gestão de repositórios digitais, como o Dspace¹¹ ou o EPrints¹². Esses programas têm os seus modelos de dados estendidos para oferecer algumas funcionalidades mais próximas aos dos sistemas CRIS. Um exemplo marcante dessa situação é a experiência realizada na Universidade de Honk Kong que, em parceria com o consórcio italiano CINECA¹³, adicionou entidades CERIF no modelo de dados do DSpace, permitindo uma maior amplitude de representação. Esse trabalho colaborativo “[...] resultou na liberação da nova plataforma DSpace-CRIS¹⁴, um CRIS de fonte aberta que está sendo gradualmente adotado por instituições de [pesquisa] em escala mundial” (CASTRO, 2014, p. 42).

Indo além, algumas instituições preferiram desenvolverem *in-house* seus próprios sistemas.

Esses desenvolvimentos domésticos têm como principal motivação a necessidade que essas instituições têm de ter o controle completo sobre requisitos específicos de gestão de seu ciclo de vida de pesquisa (CASTRO, 2014).

A natureza diversificada e a abrangência dos sistemas CRIS implicam uma inevitável heterogeneidade que vai demandar algum grau de padronização, que além de orientar o desenvolvimento e os processos desses sistemas, possam também contribuir para a troca de informações entre eles. A norma CERIF, que essencialmente, oferece um formato de intercâmbio de dados entre sistemas CRIS e um modelo de dados, tem um papel importante na consolidação das redes de sistemas de gestão de pesquisa. Pela sua importância para o presente estudo, ela será rapidamente analisada a seguir.

4 INTERCAMBIANDO INFORMAÇÕES SOBRE PESQUISA

A pesquisa científica é hoje um empreendimento necessariamente de amplitude internacional, isto acontece por vários motivos, mas especialmente porque uma pesquisa realizada dentro das fronteiras de um país, provavelmente, está baseada em pesquisas anteriores realizadas em diversos outros países e pela intensa interação, via as novas formas de interlocução entre os pesquisadores, além dos limites nacionais. A ciência contemporânea tem muitos exemplos bem conhecidos de pesquisas transnacionais, como os projetos de mapeamento do genoma humano e os estudos sobre as mudanças climáticas; existem muitos outros exemplos, especialmente quando infraestruturas caras e tecnologicamente sofisticadas são necessárias, como em física das partículas e em ciências espaciais. Por isso existe uma necessidade absoluta de compartilhamento de informações sobre pesquisa entre países, ou mesmo entre instituições de pesquisa e órgãos de fomento dentro do mesmo país (EUROCRIS, 2010)

Uma sinalização importante de que uma categoria de serviço está se tornando relevante para as áreas de CT&I é o movimento em torno da elaboração de normas e padrões que oriente o seu desenvolvimento e a sua capacidade de interoperar em escala nacional e internacional.

9 Pure: <<http://www.elsevier.com/online-tools/research-intelligence/products-and-services/pure>>

10 Converis: <<http://converis.thomsonreuters.com/>>

11 DSpace: <<http://www.dspace.org/>>

12 EPrints: <<http://www.eprints.org/uk/>>

13 CINECA: <<http://www.cineca.it/>>

14 DSpace-CRIS: <<http://cineca.github.io/dspace-cris/>>

Foi assim com as bibliotecas e repositórios digitais e esta sendo assim com os sistemas de gestão de informações sobre pesquisa.

A multiplicidade de tipos e de concepções de plataformas CRIS, conforme visto anteriormente, tem como consequência direta a dificuldade de implementação de mecanismos de interoperabilidade e de intercâmbio de dados e informações entre elas; somado a isso está a diversidade e a complexidade inerente aos fluxos informacionais que permeiam a gestão da pesquisa científica. Estes fatos impõem o desafio de se criar instrumentos de padronização de sistemas CRIS, principalmente para os sistemas que compõem as redes de maior envergadura, que permitam níveis de interoperabilidade e que regule o desenvolvimento desses sistemas na direção de concepções mais integradas entre si e entre outras infraestruturas de informação, como os RI's. Para tal, é necessário considerar que as normas criadas por entidades de padronização de um determinado país não cobrem as necessidades de outros países, dado o seu escopo limitado às especificidades locais; por outro lado, "as iniciativas internacionais e regionais (que incluem vários países) são mais complexas e a sua adoção é mais difícil, posto que nesse caso, as normas são transversais a governos, políticas e países" (PINTO; SIMÕES; AMARAL, 2014, p.3).

Neste cenário de modelos e de visões díspares, o CERIF desponta como a norma mais importante e conceitualmente mais bem elaborada, tendo como objeto principal o intercâmbio de informações entre as plataformas CRIS no âmbito da Comunidade Européia. Pela amplitude do domínio que trata - a pesquisa científica -, o CERIF é frequentemente descrita como uma norma de grande complexidade.

O CERIF apresenta um modelo conceitual que descreve o domínio da Pesquisa Científica, esta é sua essência. Formalmente ele é mantido como um Modelo Entidade Relacionamento (MER) a partir do qual *scripts* SQL¹⁵ para múltiplos sistemas de banco de dados (Oracle, mySQL, etc.) podem ser gerados. Funcionalmente, o CERIF é definido como uma norma voltada para o gerencialmente e intercâmbio de informações sobre o ciclo de vida da pesquisa científica, ou seja, informação

sobre pesquisadores, projetos, resultados de pesquisa, financiamentos, instalações, equipamentos e demais informações que têm origem nos processos subjacentes às atividades de pesquisa (RUSSEL, 2011). A norma CERIF foi desenvolvida como um projeto multi-institucional e multinacional gerenciado e mantido pelo EuroCRIS e, oficialmente, ela é uma recomendação da União Europeia para os seus países membros, tendo como perspectiva se consolidar como um quadro referencial para o desenvolvimento e interoperabilidade de sistemas CRIS no âmbito da região. Nessa direção, a norma tem sido "[...] aplicada em vários projetos como uma base referencial não somente para o desenvolvimento de sistemas de informação em ciência e tecnologia, mas também para torná-los interoperáveis" (PACHECO, et al, 2006, p.180).

O modelo formal de dados estabelecido pelo CERIF oferece uma ferramenta de representação voltada para descrever as entidades envolvidas no universo da pesquisa, os relacionamentos entre elas e como estes relacionamentos mudam com o tempo, ou seja, o modelo tem embutido um conceito relacional que é um constructo imprescindível para a representação do universo da pesquisa. O CERIF cria, dessa forma, uma abstração dos possíveis contextos do ambiente de pesquisa articulando resultados, produtos, atores, infraestruturas, entre outros. O modelo contempla também a multiplicidade de idiomas que caracteriza o multiculturalismo dos países europeus.

O conceito relacional do CERIF define quatro tipos ou níveis de entidades: entidades básicas; entidades resultados; entidades de 2º nível; e entidades *links*. As três primeiras classes de entidades formam um modelo de camadas concêntricas tendo como núcleo as entidades básicas e como foco principal a entidade Projeto.

As entidades básicas - que incluem as entidades Projeto, Organização e Pessoa - indicam e corporificam o propósito central do CERIF. Porém, a entidade Projeto é o ponto central da norma, cujo desenvolvimento tem como ponto de partida a troca de informações padronizadas sobre projetos entre os Estados membros da União Europeia. Nessa direção, ao registrar dados alinhados à norma CERIF sobre projetos, pessoas e organizações associadas ao ciclo de vida da pesquisa, as instituições

15 Sigla em inglês para Structured Query Language (Linguagem de Consulta Estruturada) que é a linguagem de consulta padrão para sistemas de banco de dados relacional

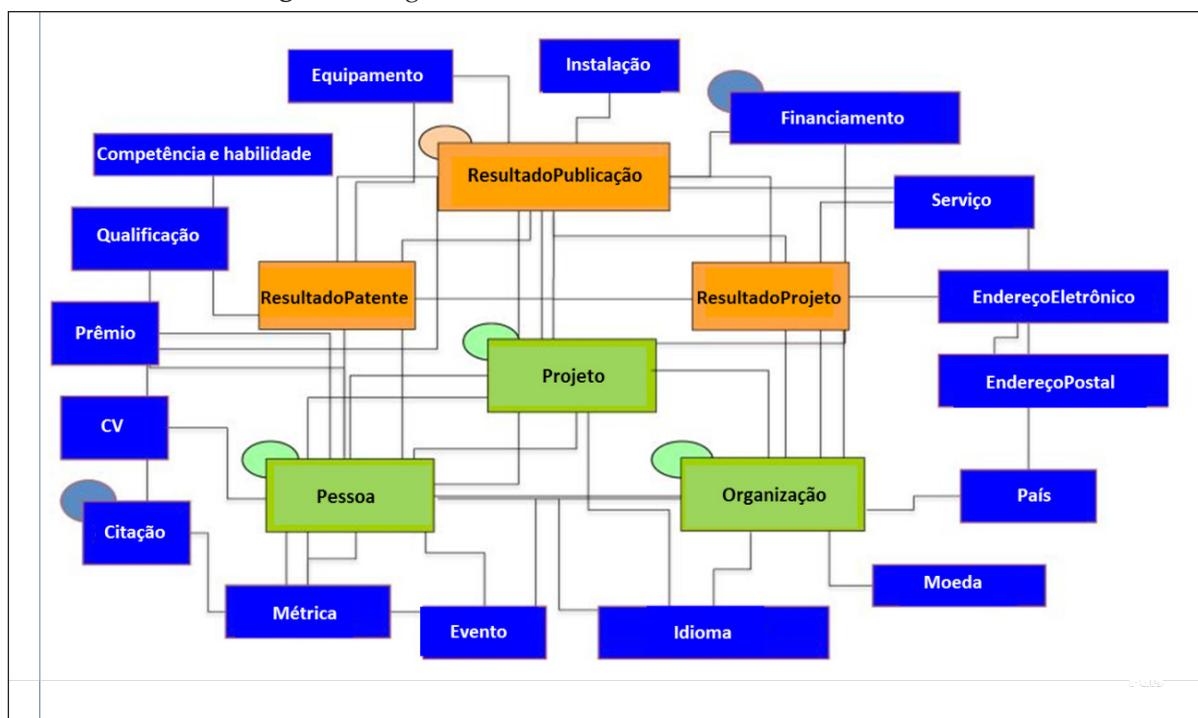
científicas ampliam o seu potencial de intercâmbio e a oferta de serviços baseados em dados padronizados e confiáveis.

Proximamente associados às entidades básicas estão as entidades que expressam os desdobramentos da atividade de pesquisa, ou seja, as entidades resultado: Publicação, Patente e Produto. Publicação é o resultado mais comum nos processos de pesquisa e, portanto, é considerada uma entidade “núcleo”; Patente corresponde às patentes geradas; e Produto corresponde tipicamente ao conjunto de dados de pesquisa, *software* e protótipos e outros produtos

ou *outputs* de atividades de pesquisa (RUSSEL, 2011).

Um segundo nível de entidades é representado no modelo circundando as entidades básicas e as entidades resultados; elas correspondem às entidades que capacitam o modelo relacional a representar o contexto da pesquisa por meio de links originados nas entidades básicas e entidades produtos, por exemplo: país, idioma, financiamento, evento, equipamento, instalação, serviço. A figura 1 representa algumas das entidades do CERIF e seus relacionamentos.

Figura 1 - Algumas entidades CERIF e seus relacionamentos



Fonte: EuroCRIS (2012, tradução nossa)

A representação das relações entre as entidades tem um papel-chave no modelo proposto pela norma CERIF na construção da representação do domínio da pesquisa. A partir daí, uma particularidade importante é enfatizada: no CERIF as relações também são entidades e assim mantêm seus próprios atributos, como, por exemplo, o “carimbo de tempo” que indica a temporalidade das relações. Baseados nesse conceito, as diversas entidades podem ser vinculadas em pares por meio de *links*, como,

por exemplo, um autor à sua publicação. Para dar maior concretude a esta representação, o modelo estabelece a ideia de “entidade *link*”, que são relações ou *links* entre entidades CERIF. Dessa forma, uma entidade *link* conecta duas entidades, sejam elas básicas, resultados ou de segundo nível. Por exemplo: Pessoa “é autor de” ResultadoPublicação; ResultadoPublicação “é financiado por” ProgramaFinanciamento. O conceito de entidade *link* é considerado pelo EuroCRIS como sendo o ponto mais forte

do modelo proposto pelo CERIF, posto que possibilita uma grande expressividade da dinâmica do mundo real da pesquisa.

4.1 Camada semântica do CERIF

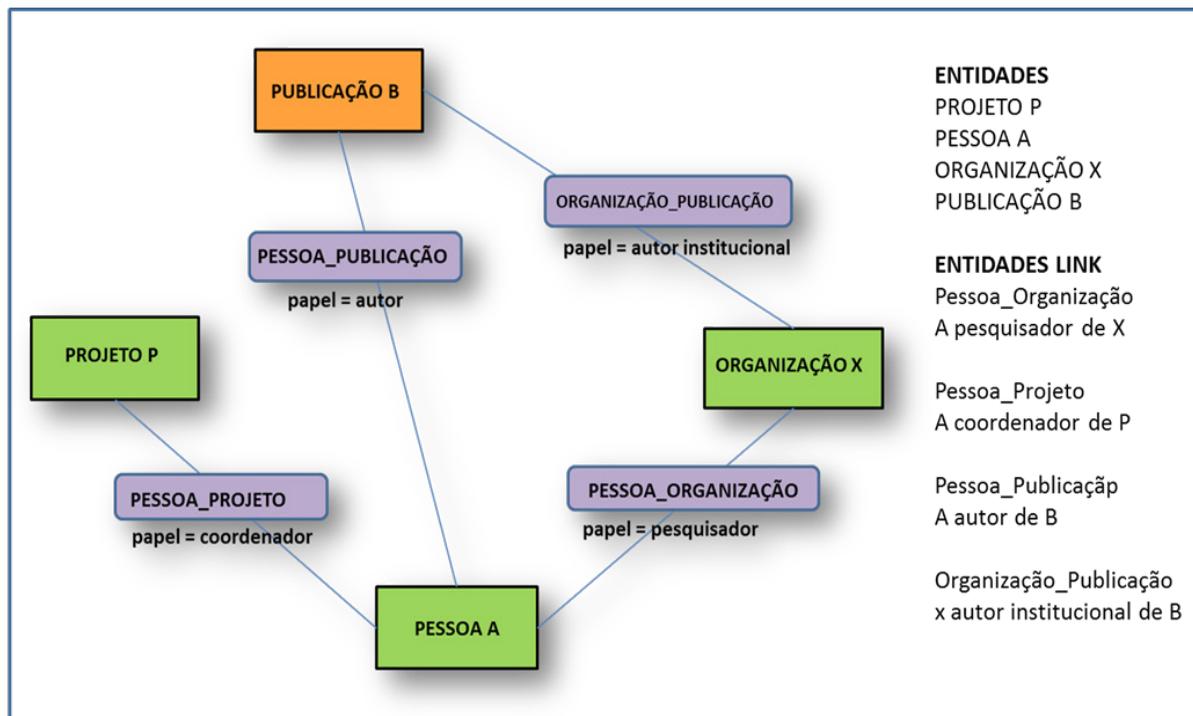
Neste momento, fica clara a necessidade de algum dispositivo semântico que atribua significado aos relacionamentos estabelecidos pelas entidades *links*. Isto é definido pelo “CERIF Semantic Layer”. Esta camada semântica é de fato um construto conceitual para descrever um submodelo do CERIF que permite a gestão eficiente e significativa de vocabulários controlados. Ela é, dessa forma, uma semântica declarada que segue a sintaxe formal estabelecida pelo modelo CERIF. Isto significa dizer que sistemas CRIS compatíveis com o CERIF são construídos sobre uma sintaxe formal e uma semântica declarada (EUROCRIS, 2012).

Assim, cada entidade *link* porta um significado que é estabelecido pela referência a esta camada semântica do CERIF. Cada registro de vinculação por *links* (linking) entre entidades

necessita também de uma data inicial e uma data final, caracterizando esta vinculação como uma variável temporal. Nessa direção, todas as entidades *link* seguem o mesmo padrão, que é de fato uma tripla: dois identificadores correspondentes às entidades que estão sendo “linkadas”; um carimbo de tempo que informa as datas de início e de fim; e uma referência à camada semântica – ou seja, a um termo de um vocabulário pertencente a um esquema e o *namespace* do esquema (EUROCRIS, 2012).

Resumidamente, a camada semântica armazena os valores semânticos definidos para os atributos das entidades e das entidades *links*. No sentido de capturar os significados desses elementos, os atributos para as entidades assinalam tipos de entidades (por exemplo: projetos, patentes), enquanto para relações assinalam papéis (por exemplo: autor, coordenador). Para tal, qualquer esquema ou estrutura, como dicionários, léxicos, tesouros, ontologias, podem ser incorporados, conforme assinala Russel (2011). A figura 2 ilustra uma situação simples de entidades links com possíveis papéis.

Figura 2 - CERIF: algumas entidades, entidades links e possíveis papéis



Fonte: Propria

.Por fim, é necessário analisar rapidamente os possíveis usos do CERIF sem perder de vista as possibilidades de interoperabilidade CRIS/RI, objeto do presente estudo. Para Russel (2011), o CERIF pode ser usado de três formas:

- Como um modelo para a implementação de um sistema CRIS isolado (porém pronto para a interoperabilidade);
- Como um modelo para definir um invólucro que envolve um sistema legado CRIS, não aderente ao CERIF, que atue como uma interface entre a estrutura interna de dados e o ambiente externo.
- Como uma definição de um formato de troca de dados para criar um armazém de dados proveniente de vários sistemas CRIS.

O histórico do desenvolvimento do CERIF registra um esforço para estendê-lo “para acomodar o Dublin Core (DC), por reconhecer as exigências da comunidade envolvida com a literatura cinzenta e o crescente número de repositórios institucionais baseados em OAI-PMH/DC” (RUSSEL, 2011, p.1). A versão CERIF 2008 - 1.2 (novembro de 2010) - estende a entidade publicação em resposta à interoperabilidade com repositórios institucionais.

A gestão e o compartilhamento de dados sobre pesquisa e de coleções de dados, publicações e outros artefatos que surgem como resultados da atividade pesquisa ocorrem em ambientes, que de forma ideal, estão permeados por recursos tecnológicos avançados. Este arcabouço tecnológico cria um novo patamar que se sobrepõe aos recursos infraestruturais convencionais e reconfigura o modo de fazer ciência delineando as bases da e-Science. É esse ambiente tecnológico que será discutido rapidamente a seguir.

5 CIBERINFRAESTRUTURA DE DADOS

O Relatório da National Science Foundation¹⁶ (NSF), conhecido como “Relatório

Atkins” (ATKINS, 2003), esclarece que o termo infraestrutura vem sendo usado desde 1920 para se referir coletivamente aos serviços públicos necessários ao funcionamento de uma economia industrial, tais como rede elétrica, sistema telefônico, malhas rodoviárias e ferroviárias, assim como tantos outros.

No contexto atual, um novo termo desponta para denotar o arcabouço tecnológico subjacente a uma sociedade transformada pela informação e pelo conhecimento: “ciberinfraestrutura”. O termo se refere a um agregado coordenado de tecnologias da informação, redes de computadores e sistemas - incluindo especialistas e organizações - que possibilitam novas formas de trabalho, entretenimento, pesquisa, educação e estilos próprios de vida (BERMAN 2008). “Se infraestrutura é determinante para uma economia industrial, então nós podemos dizer que ciberinfraestrutura é essencial para uma economia do conhecimento” (ATKINS, 2003, p.5).

No campo da pesquisa científica, ciberinfraestrutura é uma solução tecnológica e sociológica para o problema de conectar de forma eficiente laboratórios, dados, computadores e pessoas, com o propósito de permitir a dedução de novas teorias e conhecimentos científicos. Nessa direção, ciberinfraestrutura 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. Estes 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.

Um breve olhar sobre a evolução do uso de computadores para a manipulação de grandes conjuntos de dados mostra que, nos fins da década de 1990, os procedimentos para este tipo de processamento já estavam bem consolidados nos países mais avançados. As infraestruturas nacionais de computação para a pesquisa se dividiam primariamente em três vias: centros

¹⁶ Disponível em: <<http://www.nsf.gov/>> Acesso em 01 abr. 2015

locais de computação de alto desempenho, localizados em universidades e institutos de pesquisa; recursos de supercomputação sediados em laboratórios nacionais; e um pequeno número de nós de supercomputação distribuídos geograficamente, disponíveis para o pesquisador sob demanda. No final do século passado, essas configurações começaram a evoluir na direção do que começou a se chamar de “computação em grade” (*grid computing*) (GOLD, 2007). Essa nova arquitetura computacional possibilitou o compartilhamento de recursos entre plataformas e redes institucionais distribuídas e heterogêneas, abrindo perspectivas que estão mudando – em conjunto com outros avanços – profundamente o fazer científico. Um exemplo recorrente foi o surgimento da Genômica que ofereceu um novo enfoque para os problemas de pesquisa na área de Biociências, baseado no esforço colaborativo de juntar peças de dados de pesquisa de uma forma padronizada e bem entendida. Essa forma de compartilhar dados culminou com sequenciamento do genoma humano em abril de 2003 (GOLD, 2007)

Os sistemas de informação, incluindo as bibliotecas de pesquisa, têm um papel importante na composição desses novos ambientes de pesquisa, posto que o aparato tecnológico que possibilita curadoria de dados de pesquisa e a gestão de dados e informações sobre pesquisa se caracteriza como uma ciberinfraestrutura. Berman (2008), a respeito disso, nos explica que: no âmbito da pesquisa e da educação os seus protagonistas desejam ambientes coordenados que gerenciem dados digitais - da criação a preservação -, que possibilitem o arquivamento de dados provenientes de instrumentos, sensores, computadores, laboratórios, pessoas e outras fontes; que incluam ferramentas e recursos de gestão de dados, armazenamento seguro e dispositivos voltados para o uso dos dados, como computadores para análises, simulação, modelagem e visualização; além do mais, os pesquisadores desejam que seus dados possam estar disponíveis para seus pares, de forma

inteligível, por meio de portais e banco de dados, fomentando o trabalho colaborativo.

A realocação dos sistemas CRIS e Repositórios Institucionais como componentes dos ambientes tecnológicos voltadas para a pesquisa redimensiona o valor desses sistemas na medida em que seus ativos informacionais se tornam matéria prima para serviços inéditos, de maior alcance em termos de usuários (por exemplo: gestores e administradores) e mais avançados. Para tal, é importante entender com um grau a mais de detalhes as diferenças e semelhanças entre CRIS e RI e os papéis que podem desempenhar nestes novos e promissores cenários.

6 O QUE DIFERENCIA OS SISTEMAS CRIS E RI

Os sistemas CRIS e os RI's atuam em domínios informacionais convergentes, significando que ambos os sistemas estão interessados na coleta de informações sobre resultados dos fluxos de pesquisa em âmbito institucional, regional, nacional ou em torno de uma área de conhecimento. Eles têm, porém, objetivos distintos que, entretanto, muito frequentemente, se sobrepõem, da mesma forma como compartilham inúmeras funções.

O que se observa na literatura é que há um consenso absoluto entre os especialistas da área de que os sistemas de gestão de informação para a pesquisa e os repositórios institucionais não são sistemas incompatíveis. Verdadeiramente, eles oferecem soluções informacionais para diferentes necessidades das organizações voltadas para a pesquisa científica, como universidades e institutos de pesquisa. Ambos estão preocupados com a captura de informações sobre a atividade de pesquisa. Enquanto o CRIS cobre os processos de investigação científica numa perspectiva mais gerencial e abrangente, voltando-se para a instituição ou o conjunto de instituições que formam o seu domínio, a ótica dos repositórios

é mais acadêmica cobrindo publicações e artigos científicos em texto completo, voltando-se para o mundo exterior à instituição, disseminado a sua memória científica por meio de protocolos abertos.

Ambos os sistemas devem estar apropriadamente integrados aos ciclos de vida da pesquisa, registrando dados bibliográficos e documentando as informações relacionadas ao fomento, para o caso do CRIS, e arquivando de forma adequada as versões em texto completo das publicações acadêmicas, para o caso de um repositório. Os sistemas CRIS estão mais focados em monitorar a atividade de pesquisa do que maximizar o impacto, que é uma função do repositório; este, por sua vez, está mais centrado na preservação do material em texto completo do que somente em dados bibliográficos. É preciso considerar também, nessa comparação, as diferenças significantes entre os fluxos de trabalho administrativos e os de pesquisa, que podem ser gerenciados por sistemas mais adequadamente talhados para necessidades específicas de cada atividade (SHEPPARD, 2010).

Reforçando essa argumentação, Elbæk (2012) esclarece que em muitas organizações acadêmicas e de pesquisa esses dois sistemas estão vinculados a diferentes unidades organizacionais, que torna mais evidente suas respectivas missões: os repositórios estão geralmente ligados às bibliotecas científicas e a outras unidades de informação; enquanto os sistemas CRIS, na qualidade de sistemas de gestão, estão subordinados às instâncias administrativas das instituições de pesquisa. Porém “ambos os sistemas confiam [...] nos dados fornecidos pelos pesquisadores e administradores de pesquisa, diretamente ou via serviços mediadores como bibliotecas, escritórios de pesquisa, editores e bases de dados” (CASTRO, 2014, p.1)

Entretanto, a trajetória evolutiva desses sistemas mostra uma tendência para uma convergência funcional, significando que os sistemas de um tipo começam a incorporar

funções do outro. Castro (2014) ressalta que, não obstante serem sistemas distintos, na medida em que as plataformas CRIS avançam na remodelação de suas funcionalidades, elas começam a cobrir um número crescente de particularidades que antes eram tradicionalmente disponibilizadas somente pelos repositórios. Nessa direção, os pacotes de programa CRIS se tornam compatíveis com o protocolo de interoperabilidade OAI-PMH, começam a armazenar arquivos em texto completo acessíveis publicamente, gerenciam tempos de embargo de publicações acadêmicas entre outras funcionalidades típicas dos repositórios.

Por outro lado, os repositórios começam também a ampliar seus modelos de dados de forma que possam abrigar metadados que descrevem atividades gerenciais de pesquisa, como descrição de projetos de pesquisa e identificação de órgãos financiadores de projetos.

Essa crescente sobreposição de funções entre sistemas CRIS e RI, aliada a restrições econômicas, tem privilegiado configurações onde os RI são substituídos por CRIS. Castro (2004) enfatiza ainda que embora os sistemas CRIS não ofereçam atualmente todas as funcionalidades dos repositórios, esses sistemas são percebidos por um número considerado de instituições como avançados o suficiente para justificar a sua adoção para apoiar a totalidade das atividades relacionadas à gestão de informações para a pesquisa, incluindo as atividades desempenhadas pelos RI's.

Com o objetivo de tornar as diferenças entre os dois sistemas mais perceptíveis e caracterizar com maior rigor a natureza de cada um deles, foi proposta o Quadro 1, que tenta estruturar as diferenças que coletivamente identificam conceitualmente cada sistema, mesmo reconhecendo que essa diferenciação se esvanece à medida que os sistemas evoluem e se integram. Uma fonte importante para a formulação dessa síntese foram os trabalhos de Castro (2014) e Castro; Shearer; Summann (2014).

Quadro 1 - Principais diferenças entre o sistema CRIS e RI

	SISTEMAS	
	CRIS	RI
FUNÇÃO PRINCIPAL	Gestão das informações sobre o ciclo total das atividades de pesquisa científica	Formação da memória acadêmica digital das instituições de pesquisa
ESCOPO	Lida com as informações de todo o ciclo de atividade de pesquisa, incluindo informações sobre projetos, financiamentos, pessoas, organizações, produtos de pesquisa, instalações e equipamentos.	Ênfase no arquivamento de produtos de pesquisa, principalmente publicações acadêmicas. Entretanto, o escopo tem se estendido para incluir outros produtos, tais como dados de pesquisa.
OBJETIVO	Coleta informações com o objetivo de reportar as atividades institucionais de pesquisa, considerando também informações sobre ambientes externos como governo e agências de fomento à pesquisa.	Arquivar, tratar, preservar e disseminar produtos de pesquisa, com forte ênfase em publicações acadêmicas, tendo como orientação os pressupostos do movimento de Livre Acesso.
ABRANGÊNCIA	Na qualidade de ferramenta voltada para coletar informações sobre todas as áreas da instituição envolvidas em atividades de pesquisa, incluindo informações econômicas, os sistemas CRIS são basicamente projetados para uso interno. Esses sistemas não estão, tradicionalmente, preocupados em disseminar externamente as informações que eles armazenam.	Os repositórios por sua vez, são orientados para o mundo exterior, ou seja, disseminar, dar visibilidade, permitir o acesso livre e aumentar o impacto dos produtos de pesquisa da instituição.
GESTÃO	Os sistemas CRIS são tipicamente comandados pelos mesmos setores que fazem a gestão dos projetos de pesquisa na instituição e que cuidam da avaliação, planejamento, finanças etc.	Os repositórios são geralmente gerenciados pelas unidades de informação das instituições, representados, na maioria dos casos, pelas bibliotecas de pesquisa.
CONTEÚDO	Como ferramenta acurada para reportar atividades de pesquisa, os sistemas CRIS lidam exclusivamente com metadados que contextualizam os fluxos de pesquisa da instituição.	Os repositórios também fazem uso de metadados, porém sua atenção está centrada na disponibilidade de texto completo e na criação, arquivamento, curadoria e reuso dos objetos digitais.
METADADOS	Os sistemas CRIS operam com padrões de metadados complexos e não suficientemente harmonizados, mas que são necessários para a descrição das inúmeras atividades de pesquisa. O CERIF é o padrão mais usado. Embora a sua implementação varie consideravelmente entre países, ele está se tornando rapidamente o padrão <i>default</i> que poderá contribuir para a interoperabilidade de alto nível.	Os repositórios operam com modelos bem mais simples de metadados, que são, por sua vez, transversalmente bastante consistentes entre instituições e países, isto permite um alto grau de interoperabilidade. O padrão mais usado é o Dublin Core qualificado.
PLATAFORMAS DE SOFTWARE	Embora muitas instituições tenham optado por desenvolver suas próprias soluções de CRIS no sentido de atender aos seus requisitos específicos, as soluções comerciais tendem a prevalecer, principalmente devido à complexidade intrínseca dos modelos e à possibilidade de configurar as comunidades de usuários.	As plataformas de <i>software</i> para Repositórios têm uma arquitetura mais simples e são, tradicionalmente, pacotes do tipo <i>open source</i> que envolve no seu desenvolvimento uma ampla comunidade internacional. Os pacotes mais usados mundialmente são o Dspace e o EPrints.
COMPARTILHAMENTO DE INFORMAÇÕES	As comunidades CRIS só recentemente iniciaram o compartilhamento sistemático de informações.	As comunidades responsáveis pelos repositórios, orientadas pelo padrão de intercâmbio bibliotecas de pesquisa, praticam uma troca intensiva de informações.
INTEROPERABILIDADE	Ainda insipiente, devido à questão de metadados, modelos adotados e padronização.	Via protocolo OAI-PMH e o uso de metadados DC.

Fonte: autores, adaptado de Castro (2014)

7 CONFIGURAÇÕES DE SISTEMAS DE GESTÃO DE INFORMAÇÕES PARA A PESQUISA

Castro e seus colaboradores (2014), como resultado de um levantamento realizado em instituições de pesquisa do Reino Unido, sintetizam as possibilidades de configurações de sistemas CRIS nos seguintes casos:

- **Repositórios Institucionais somente** - Muitas instituições científicas, principalmente as menores, não operam sistemas CRIS e utilizam os RI's como sua principal ferramenta de gestão de informação sobre pesquisa, adotando o padrão conhecido como "RI como CRIS" ou "RI ampliado" em que algumas funcionalidades características de sistemas CRIS são obtidas por meio da extensão do modelo de dados do RI. Esse mecanismo permite que os repositórios coletem informações adicionais sobre a atividade de pesquisa além das circunscritas pelas publicações acadêmicas. "Não há nada que impeça um modelo de metadados baseado no Dublin Core de ser estendido para cobrir áreas adicionais de pesquisa, como projetos de pesquisa" (CASTRO, 2014, p.6). A ampliação dos repositórios por meio de ferramentas tais como *addon* para repositórios E-Prints ou a solução de *software* livre Dspace-CRIS "capacitam as instituições a gerenciarem suas informações de pesquisa tradicionalmente associadas com sistemas CRIS, como dados de pessoas e projetos". (CASTRO, 2014, p.6).
- **CRIS somente** - Este é o caso em que uma instituição concentra todas as informações para pesquisa no ambiente CRIS, que incorpora também as funções específicas de repositório institucional. Em relação à sustentação tecnológica para essa configuração, Castro e seus colaboradores (2014) observam que algumas plataformas comerciais CRIS vêm desenvolvendo atributos de RI para os seus produtos, incluindo aderência ao protocolo OAI-PMH. Isto significa, ainda de acordo com estes autores, que

há uma clara tendência, principalmente nos países que possuem infraestruturas avançadas de gestão de informação para a pesquisa, na direção de tornar sistemas RI e CRIS numa plataforma única. Nesse desenho, os produtos de pesquisa são geridos no ambiente CRIS, onde também são disponíveis via acesso aberto e para coleta automática (*harvesting*) de agregadores OAI-PMH. Entretanto é necessário considerar que os sistemas CRIS não foram originalmente projetados para disseminar externamente produtos de pesquisa, e que não obstante a gradual aderência ao protocolo OAI-PMH, ainda estão longe de possuírem especificidades oferecidas, por exemplo, pelo protocolo SWORD¹⁷, como lembram Castro, Shearer e Summann (2014, p.45)

- **Operação conjunta CRIS e repositório institucional** - nessa categoria os pesquisadores reuniram as configurações adotadas pelas instituições que operam concomitantemente um sistema CRIS, desenvolvido por elas mesmas, ou sistemas comerciais que sejam baseados no padrão CERIF - como, por exemplo, o *Pure* ou *Converis* -, e também um repositório institucional. Os dois sistemas podem rodar de forma independente, dando suporte aos relatos de atividades de pesquisa e de disseminação de produtos de pesquisa como processos apartados. Entretanto, como destacam Castro, Shearer e Summann (2014, p.43), "devido considerações sobre a eficiência e a gradual disponibilidade de tecnologias [...], ambos os sistemas estão geralmente interligados de forma que ocorre um processo sistemático de troca de dados entre eles". Tipicamente, o CRIS submete automaticamente os metadados de publicações ao RI por meio de mapeamento CERIF/Dublin Core ou CERIF/MODS¹⁸. Nessa direção, o CRIS/CERIF atua como um sistema de gestão

17 SWORD (Simple Web-service Offering Repository Deposit) é um padrão de interoperabilidade que permite que os repositórios digitais aceitem depósitos de conteúdos de fontes múltiplas em diferentes formatos. Para saber mais: <<http://swordapp.org/>>

18 MODS (Metadata Object Description Schema) é um esquema de metadados bibliográficos que pode ser usado para vários propósitos. Para saber mais: <<http://www.loc.gov/standards/mods/>>

interno para os produtos de pesquisa, enquanto o RI/OAI-PMH dissemina esses produtos para o mundo externo, oferecendo acesso aberto aos textos completos e a esses produtos sempre que possível. Completando, Castro, Shearer e Summann (2014) assinalam que a disponibilidade de *gateway* CERIF/DC ou CERIF/MODS nos sistemas CRIS oferecem uma grande oportunidade de interoperabilidade CRIS/IR, entretanto existem numerosos exemplos de interoperabilidade entre plataformas CRIS e RI, onde o CRIS não é aderente ao padrão CERIF.

As configurações de CRIS e RI apresentadas se tornam mais complexas na medida em que entram em cena os repositórios de dados de pesquisa – também chamados por alguns autores de repositório de *e-Science*, repositório de e-pesquisa ou ainda e-Repositório. Embora haja algumas experiências interessantes na aplicação de repositórios de *e-prints* para a arquivamento de dados e conjunto de dados de pesquisa – como é o caso do CarpeDien¹⁹ do Instituto de Engenharia Nuclear (SALES, 2014) –, via de regra, as instituições adotam soluções distintas para dados e publicações. As razões para tal, conforme explicitado por Jeffery e Asserson (2009), recaem principalmente nos padrões de acesso e nos requisitos de representação via metadados, que são claramente diferentes. Os repositórios de e-pesquisa requerem um conjunto de metadados muito mais detalhado e mais preciso para o controle e a gestão de seus ativos informacionais. Estes requisitos superam os limites mais simples necessários à descoberta e ao acesso aos recursos dos repositórios de *e-prints*. Por exemplo, as anotações sobre a prática laboratorial de análise de uma determinada substância química vão requerer um conjunto específico de metadados, enquanto outra coleção de dados, para ser interpretada, gerida e recuperada, vai exigir outro e diferente grupo de metadados (SALES, 2014, p. 63).

Uma das maiores dificuldades na criação de repositórios de dados recai sobre a

natureza heterogênea e diversificada do que é chamado dados de pesquisa. Cada disciplina, cada instituição tem visões distintas sobre o que é dado de pesquisa, como deve ser feita a curadoria, a recuperação e a apresentação de suas coleções de dados. Além do mais, há muito mais restrições sociais, éticas e legais que devem ser levadas em contas do que quando consideremos os repositórios abertos de publicações acadêmicas; e, por fim, questões de autenticidade, integridade, proveniência e validade no tempo exigem diferentes formas de registro. Tudo isso somado implica uma gestão mais apurada e tecnologias mais sofisticadas para dar conta do ciclo de vida dos dados de pesquisa arquivados em repositórios de e-pesquisa.

Enquanto artefato tecnológico, os repositórios de dados estão estruturados, na maioria dos casos, na forma de base de dados factuais. Entretanto, devido às condicionantes de cada disciplina, o panorama geral desses repositórios é extremamente diversificado em termos de conteúdo, concepção e tratamento dos dados. Sendo assim, na prática, a escolha do conjunto de tecnologias a ser aplicado no desenvolvimento e na operação dos repositórios de e-pesquisa está relacionada à forma como se processa o fluxo das pesquisas e como se configura o ciclo de curadoria dos dados, ou seja, que tipos de dados serão arquivados, como eles serão recuperados, visualizados e reusados.

8 INTEROPERABILIDADE CRIS/RI

A convergência de interesse e a complementaridade de objetivos e funções dos sistemas CRIS e dos Repositórios Institucionais apontam para a necessidade de se construir pontes tecnológicas, sintáticas e semânticas que reforcem as características de cada um deles em prol das instituições de pesquisa a que servem. Numa primeira abordagem, isso se dá através da transferência de informação entre essas diferentes plataformas informacionais.

As ações técnicas e tecnológicas que permitem a troca inteligível de informações entre os sistemas CRIS e RI's localizados no mesmo domínio informacional – seja uma

¹⁹ Disponível em: < <http://carpedien.ien.gov.br/> > Acesso em: 01 abr. 2015

instituição ou uma rede de instituições – determinam o nível de interoperabilidade entre esses sistemas. De acordo com Castro (2014), o fluxo de troca de informações entre os sistemas irá envolver graus de transferência de metadados entre eles, o que por sua vez, implica, via de regra, na necessidade do estabelecimento de regras de mapeamento entre os diferentes padrões de metadados operados pelas plataformas CRIS e pelos RI's.

A integração dos ativos informacionais - antes autocontidos em seus sistemas específicos - proporcionada pela interoperabilidade, aumenta o espectro de possibilidades que se pode vislumbrar, em termos de redução de esforços, oferecimento de novos serviços e troca de informações com sistemas externos ao domínio a que pertencem.

Uma vantagem direta da possibilidade de interoperabilidade é a redução de esforços na entrada de informações nas áreas onde os sistemas apresentam sobreposições importantes, como por exemplo, na descrição de um projeto de pesquisa; e também no enriquecimento, proporcionada pela integração, das informações recuperadas em ambos os sistemas através do *linkage* dos itens de informação. Por exemplo, recuperados os metadados sobre um projeto na plataforma CRIS, a integração permite o acesso imediato ao texto completo das publicações relacionadas, como artigos e relatórios, via URI, na base de dados do repositório.

Uma sinalização eloquente sobre a importância da interoperabilidade CRIS/IR em patamares interinstitucionais e regionais está explicitada na publicação, bastante recente, "OpenAIRE Guidelines for CRIS Managers based on CERIF XML" (OpenAIRE, 2014). Tomando como princípio que o modelo de dados adotado pelo OpenAIRE²⁰ está em conformidade com o padrão CERIF, e que, além do mais, o CERIF XML foi adotado pelo OpenAIRE como base para coleta automática (*harvesting*) e importação de metadados dos sistemas CRIS, os *Guidelines* oferecem aos

gestores de sistemas CRIS orientação para expor os seus metadados para *harvesting* de forma que eles sejam compatíveis com a infraestrutura OpenAIRE. "Implementando os *Guidelines*, os gestores CRIS apoiam a inclusão e, portanto, o reuso de metadados dos seus sistemas no âmbito da infraestrutura OpenAIRE" (Open AIRE, 2014, P.1).

Atualmente o OpenAIRE atua sobre uma rede interoperável e validada de mais de 520 repositórios e de periódicos *Open Archives* (AO), integrando mais de 9 milhões de publicações livres e cerca de 1000 conjuntos de dados. "A extensão do provedor de dados OpenAIRE para incluir os sistemas CRIS resultará numa quantidade significativa de conteúdos coletados automaticamente [...]" CASTRO (2014, p.6). Isto torna evidente as vantagens da implementação da interoperabilidade CRIS/IR também no plano interinstitucional, reforçando, ainda mais, os pressupostos de uma ciência aberta.

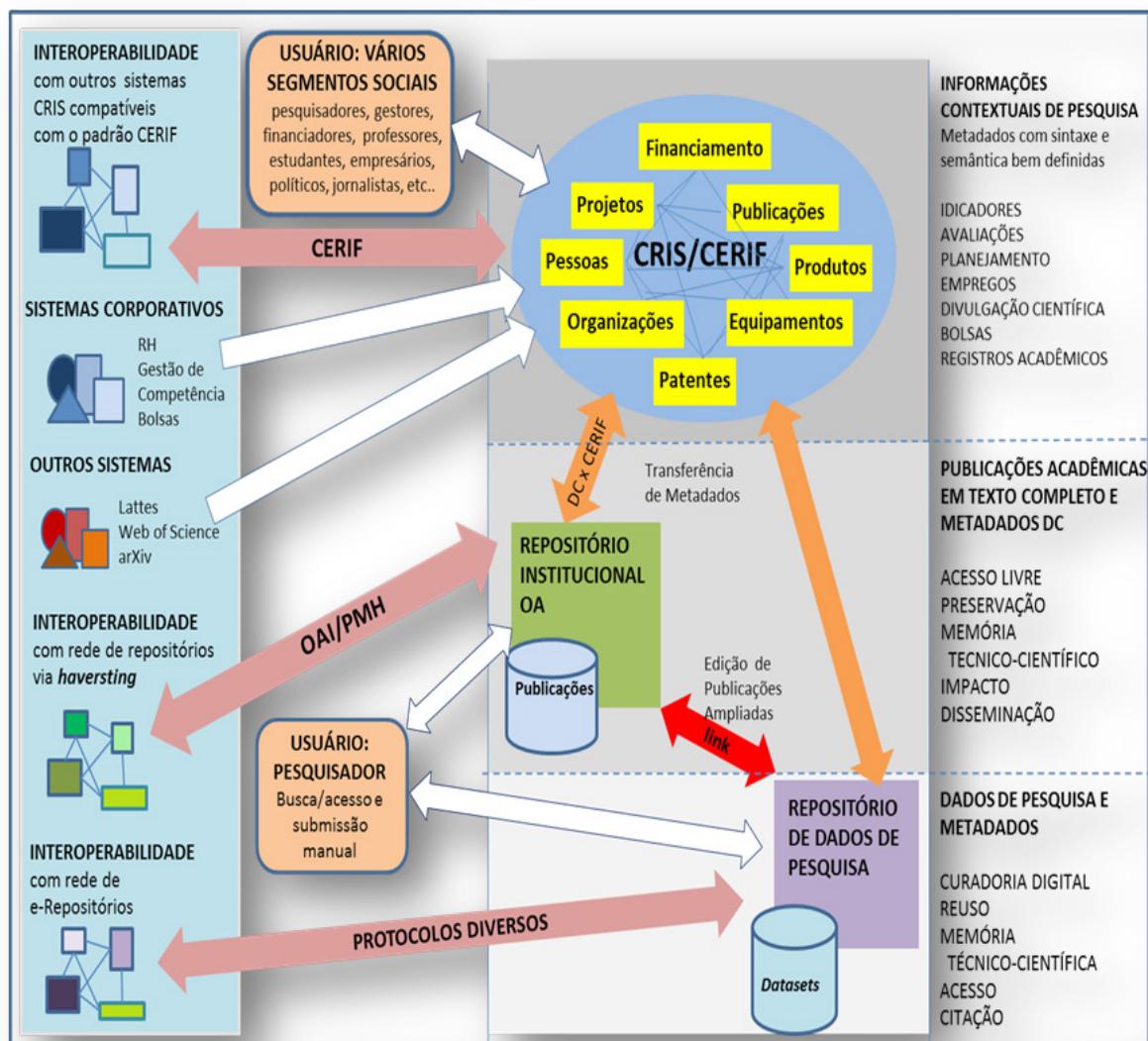
Indo mais adiante, o *gateway* CERIF-XML para intercâmbio de metadados entre sistemas CRIS e RI pode ser também adaptado para ser aplicado a sistemas CRIS não compatíveis com o CERIF. Isto é viabilizado por meio do mapeamento dos modelos de dados específicos com o padrão CERIF, e abre a possibilidade do surgimento de redes que vão além dos propósitos do OpenAIRE (DE CASTRO; SHEARER; SUMMANN 2014).

9 PROPOSTA DE UMA ARQUITETURA

A partir dos elementos discutidos anteriormente, é delineado um modelo de integração sintetizado na figura 3 e no quadro 2. A arquitetura proposta estabelece um ambiente completo de informações para a pesquisa, identificada como parte da ciberinfraestrutura de uma organização focada em atividades acadêmicas. Esta arquitetura está composta pelos seguintes sistemas: Repositório institucional (RI); Repositório de dados de pesquisa (e-Repositório) e Sistema CRIS que opera com um modelo de dados aderente à norma CERIF.

20 O OpenAIRE é um esforço colaborativo no âmbito da Comunidade Europeia que tem como objetivo geral apoiar a implementação do acesso livre na Europa. Para saber mais: <<https://www.openaire.eu/>>.

Figura 3 - Modelo de interoperabilidade RI X RD X CRIS



Fonte: Própria.

O quadro 2 sintetiza os principais componentes da arquitetura, usuários e os fluxos de troca de informações entre os sistemas e entre os demais sistemas administrativos e acadêmicos internos e externos à instituição.

Quadro 2 – Sistemas, usuários e troca de informações entre os sistemas internos e externos à instituição

SISTEMAS QUE INTEGRAM A ARQUITETURA	
Embora estes sistemas estejam integrados, eles podem funcionar de maneira independente uns dos outros, dando apoio, de forma separada, aos processos institucionais de reportar as atividades de pesquisa – geralmente de responsabilidade do escritório de pesquisa – e de disseminação dos produtos e resultados de pesquisa, que é responsabilidade característica da biblioteca científica.	
SISTEMAS	REPOSITÓRIO INSTITUCIONAL Repositório digital compatível com o protocolo OAI-PMH/ Dublin Core; base de dados que arquiva, preserva e dissemina publicações acadêmicas em texto completo e seus metadados e permite o acesso livre e autossucessão. ACESSO LIVRE; PRESERVAÇÃO; MEMÓRIA TÉCNICO-CIENTÍFICA; IMPACTO E DISSEMINAÇÃO
	REPOSITÓRIO DE DADOS DE PESQUISA (e-REPOSITÓRIO) Base de dados de coleções de dados de pesquisa que tem políticas específicas de acesso e interoperabilidade. Utiliza metadados compatíveis com o DC/DataCite ¹ e metadados específicos da área. CURADORIA DIGITAL; REUSO DE DADOS; MEMÓRIA TÉCNICO-CIENTÍFICA, ACESSO E CITAÇÃO
	SISTEMA CRIS Ambiente que opera com um modelo de dados aderente à norma CERIF; caracterizado como uma ferramenta de apoio à gestão de pesquisa por meio do fornecimento de informações contextuais sobre atividades de pesquisa da instituição. INDICADORES; AVALIAÇÕES; PLANEJAMENTO; EMPREGOS; FINANCIAMENTO; REGISTROS ACADÊMICOS
USUÁRIOS: CONSULTA E SUBMISSÃO	
O ambiente tem como perspectiva oferecer recursos informacionais para uma ampla gama de usuários internos e externos à instituição que podem ser categorizados em três tipos:	
USUÁRIOS	PESQUISADORES Consultam o ambiente integrado via interface CRIS para obterem informações contextualizadas; utilizam as interfaces das plataformas RI e e-Repositório para alguns tipos de busca e para submissão manual de publicações acadêmicas e coleções de dados de pesquisa.
	GESTORES, ADMINISTRADORES E FINANCIADORES DE PESQUISA Utilizam o CRIS como ferramenta de apoio aos processos de gestão de pesquisa (indicadores, métricas etc.).
	USUÁRIOS FORA DO MUNDO DA PESQUISA (PROFESSORES, ESTUDANTES, JORNALISTAS, GOVERNANTES, CIDADÃO COMUM ETC.) Consultam o ambiente via interface CRIS para obterem informações sobre pesquisa para suas atividades profissionais ou de estudos
INTEROPERABILIDADE ENTRE OS SISTEMAS	
Os processos sistemáticos de troca de informação que se estabelecem entre os sistemas podem ser descritos da seguinte forma:	
TROCA DE INFORMAÇÕES ENTRE CRIS, RI E e-Repositório	CRIS à RI A plataforma CRIS submete automaticamente todos os metadados de publicações coletados de fontes externas ao RI por meio de um mapeamento CERIF/Dublin Core; o CRIS armazena links para as publicações em texto completo no RI e um amplo conjunto de informações contextuais;
	RI à CRIS O RI transfere metadados capturados através de <i>harvesting</i> ou entradas manualmente por autossucessão via mapeamento DC/CERIF para o CRIS; metadados correspondem a publicações (Resultado/Publicação)
	e-Repositório à CRIS O repositório de dados transfere metadados sobre coleções de dados (Resultado/Produto) para o CRIS via mapeamentos específicos.
	e-Repositório à RI O repositório de dados faz o link dos conjuntos de dados com os textos completos pertinentes ao RI, compondo publicações ampliadas
INTEROPERABILIDADE COM OUTROS SISTEMAS	
Além da troca de informações entre si, os três sistemas também podem, para cumprir os seis objetivos, apresentar graus de Interoperabilidade com outros sistemas administrativos e acadêmicos internos e externos à instituição, ampliando as bases informacionais do ambiente.	
TROCA DE INFORMAÇÕES COM OUTROS SISTEMAS ACADÊMICOS E ADMINISTRATIVOS	RI COMO PROVEDOR DE DADOS E AGREGADOR OAI O RI expõe seus metadados para o <i>harvesting</i> via protocolo OAI-PMH, ao mesmo tempo em que pode, na qualidade de provedor de dados, coletar metadados de outros repositórios para a constituição de novos serviços.
	CRIS X SISTEMAS CORPORATIVOS Internamente o CRIS acessa dados relativos às atividades de pesquisa armazenadas nos sistemas corporativos tais como recursos humanos, currículos, registros de estudantes, controle de bolsas de pesquisa e finanças
	CRIS X CRIS/CERIF EXTERNOS o CRIS troca dados com outros sistemas CRIS compatíveis com a norma CERIF.
	CRIS X SERVIÇOS DE INFORMAÇÃO EXTERNOS Além da entrada de dados manual, o CRIS pode capturar dados através de coleta automática (<i>harvesting</i>) ou <i>uploading</i> de fontes externas como, por exemplo, o PubMed, arXiv, Web of Science e a Plataforma Lattes.
	e-REPOSITÓRIO X REDE EXTERNA DE e-REPOSITÓRIOS O Repositório de dados troca dados com a rede de e-repositórios segundo protocolos e esquemas de metadados específicos ou padronizados, como o OAI-PMH e o datacite; além disso, os dados e conjunto de dados podem ser "linkados" com as publicações armazenadas em RI distribuídos, compondo publicações ampliadas.

Fonte: Própria.²¹

²¹ DataCite: Padrão para citação de dados de pesquisa. Para saber mais: <<https://www.datacite.org/>>

10 À GUIA DE CONCLUSÃO

Ainda em 2010, mais de 60 especialistas se reuniram em Leeds, no norte da Inglaterra, em torno do evento de nome sugestivo “Learning how to play nicely: repositories and CRIS”. A ideia do evento era explorar os relacionamentos e as sobreposições entre os repositórios de acesso aberto e os sistemas CRIS que estavam cada vez mais sendo implantados nas universidades.

O fato relevante desse evento é que ele inaugurava, em torno da possibilidade de interoperabilidade técnica entre CRIS e RI, um novo patamar de diálogo entre gerentes de pesquisa, financiadores de pesquisa, administradores e bibliotecários científicos.

O que parece claro, a partir da análise da literatura e das implementações que estão documentadas, é que as necessidades informacionais dos escritórios de pesquisa e das bibliotecas científicas começam a convergir, tornando mais tênue a propalada dicotomia entre sistemas CRIS e RI. Esta realidade começa a moldar os novos sistemas de informação para o ambiente de pesquisa que exploram a complementaridade dessas duas plataformas informacionais. Esses sistemas, considerados isoladamente, não são necessariamente incomuns, mas reconfiguram os seus papéis e a sua importância estratégica quando se integram orientados para novos propósitos e necessidades do universo da pesquisa científica contemporânea. Contribui para esse cenário, em que os sistemas crescentemente se fundem para melhor servir a instituição e seus pesquisadores, a visão e os interesses dos usuários que não está focada em sistemas, mais sim nos serviços que eles podem proporcionar.

Essa interlocução pode ser ampliada para o domínio dos estudos em sistemas de informação

e abre horizontes importantes e renovados para a Ciência da Informação e para a chamada Biblioteconomia Científica. Isto porque as novas configurações reveladas pelas possibilidades de interoperabilidade técnica, sintática e semântica entre repositórios institucionais e sistemas de gestão de informações para a pesquisa precisam ser estudados transversalmente, em termos práticos e teóricos, por várias disciplinas, mas especialmente pela Ciência da Informação, em cujo domínio se concentram os principais conceitos envolvidos.

Deve-se considerar, finalmente, que a implantação de ambientes voltados para a gestão das atividades de pesquisa depende fortemente da percepção do seu valor por parte das instâncias governamentais apropriadas e também dos órgãos de fomento à pesquisa. No caso específico do Brasil, há que se considerar também que os sistemas do tipo CRIS ainda não se tornaram uma realidade visível e plenamente estudada; de uma forma geral, as instituições de pesquisa não estão capacitadas tecnicamente, não dispõem de aportes financeiros contínuos e expertise para implantar e manter ambientes integrados de gestão de informações para a pesquisa. Por outro lado, o país possui uma ampla base implantada de repositórios institucionais desenvolvidos sobre a plataforma Dspace e operados pelas principais universidades e institutos de pesquisa. Este fato pode indicar que solução “RI como CRIS” ou “RI ampliado”, centrada na ampliação dos modelos de dados dos repositórios para hospedar algumas entidades CRIS-CERIF, seja a mais apropriada como primeira etapa para o desenvolvimento de sistemas mais completos que contemplem o registro e a troca de dados sobre pesquisa entre instituições do país.

CYBERINFRASTRUCTURE INFORMATION FOR RESEARCH: a proposal for architecture for integrating repositories and CRIS systems

ABSTRACT Despite the importance of institutional repositories (IR) and data repositories (e-repositories) as an essential part of today's global information infrastructure for research, the data models of these systems can't support the complexity and diversity of information needed to comprehensive management of research environments. Furthermore, governments, research institutions and funding agencies are deploying more sophisticated systems known collectively by CRIS (Current Research Information System), aimed to manage the large amount of metadata about entities involved in research activities - such as projects, researchers, funding, publications, patents, laboratories, equipment, curricula etc. -, as well as the relationships and information flows that are established between them. In this scenario, a major challenge that arises is the integration and the synchronization of CRIS, IRs and e-repositories in order to increase the range of applications and services that such systems can perform in administrative and academic scope. In this sense, this paper aims to propose an integrated architecture that synchronize the information flow of CRIS systems based on the european standard CERIF (Common European Research Information Format), IR's based on the OAI-PMH (Open Archives Initiative-Protocol for Metadata Harvesting), e-Repositories and others academic and administrative systems, internal and external to a research institution. As methodology have been examined standards and technologies that make possible the interoperability of these systems in the context of a dedicated cyberinfrastructure focused on research.

Keywords: CRIS. CERIF. Institutional repositor. Data repositor. Research ciberinfrastructure.

REFERÊNCIAS

ATKINS, Daniel. **Revolutionizing science and engineering through cyberinfrastructure**: Report of the National Science Foundation. Blue-Ribbon Advisory Panel on Cyber-infrastructure. 2003. Disponível em: <<http://www.nsf.gov/cise/sci/reports/atkins.pdf>>. Acesso em: 14 abr. 2015.

BERMAN, Francine. Got data? A Guide to data preservation in the information age. **Communications of the ACM**, v. 51, Dec. 2008. Disponível em: <<http://dl.acm.org/citation.cfm?id=1409376&CFID=501533798&CFTOKEN=44027378>>. Acesso em: 14 abr. 2015.

CASTRO, Pablo de. **7 things you should know about ... Institutional Repositories, CRIS Systems, and their interoperability**. Sep. 2014. Disponível em: <<https://www.coar-repositories.org/activities/repository-observatory/third-edition-ir-and-cris/7-things-you-should-know->

[about-institutional-repositories-cris-systems-and-their-interoperability/>](#). Acesso em: 15 abr. 2015.

CASTRO, Pablo de; SHEARER, Kathleen; SUMMANN, Friedrich. The gradual merging of repository and CRIS solutions to meet institutional research information management requirements. **Procedia Computer Science**, v. 33, p. 39-46, 2014.

DEMPSEY, Lorcan. **Research information management systems - a new service category**, 26 Oct. 2014. Disponível em: <<http://orweblog.oclc.org/archives/002218.html>>. Acesso em: 14 abr. 2015.

ELBÆK, Mikael Karstensen. **KE CRIS-OAR metadata interoperability Project**: Final report. Knowledge Exchange, 27 Mar. 2012. Disponível em: <https://infoshare.dtv.dk/twiki/pub/KeCrisOar/ProjectDocuments/Knowledge_Exchange_CRIS-OAR_final_report_26032012.pdf>. Acesso em: 14 abr. 2015.

- EUROCRIS. **CERIF Introduction**. EuroCRIS, 2010. Disponível em: <<http://www.eurocris.org/Index.php?page=CERIFintroduction&t=1>>. Acesso em: 14 abr. 2015.
- EUROCRIS. **CERIF 1.3 Full Data Model (FDM): Introduction and Specification**. 2012. Disponível em: <http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_FDM.pdf>. Acesso em: 14 abr. 2015.
- GOLD, Anna. Cyberinfrastructure, data, and libraries, part 1: a cyberinfrastructure primer for librarians. **D-Lib Magazine**, v. 13, n. 9/10. Sep./Oct. 2007. Disponível em: <<http://www.dlib.org/dlib/september07/gold/09gold-pt1.html>>. Acesso em: 14 abr. 2015.
- IMPERIAL COLLEGE LONDON. **Research information management: Developing tools to inform the management of research and translating existing good practice**. London: Imperial College London, 2010. Disponível em: <<http://www.snowballmetrics.com/wp-content/uploads/research-information-management1.pdf>>. Acesso em: 14 abr. 2015.
- INDIANA UNIVERSITY. **What is cyberinfrastructure?** 2014. Disponível em <<https://kb.iu.edu/d/auhf>>. Acesso em: 15 abr. 2015.
- JEFFERY, Keith; ASSERSON, Ane. Institutional Repositories and Current Research Inftionorma Systems. **New Review of Information Networking**, v. 14, n.12, p. 71-83. 2009. Disponível em: <<http://www.tandfonline.com/doi/pdf/10.1080/13614570903359357>>. Acesso em: 14 abr. 2015.
- JOINT, Nicholas. Current research information systems, open access repositories and libraries: ANTAEUS. **Library Review**, v. 57, n. 8, p. 569-575, 2008. Disponível em: <<http://dx.doi.org/10.1108/00242530810899559>>. Acesso em: 14 abr. 2015.
- LAMBERT, S. C. E-infrastructure, science data and CRIS. *Data Science Journal*, v. 9, p. CRIS53-CRIS58, 2010.
- OPENAIRE. **OpenAIRE Guidelines for CRIS Managers based on CERIF XML**. 2014. Disponível em: <https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_CRIS>. Acesso em: 14 abr. 2015.
- PACHECO, Roberto C. S. *et al.* Toward CERIF-ScienTI cooperation and interoperability. 2006. In: ASSERSON, Anne Gams Steine ; SIMONS, Eduard J. (ed.). **Enabling interaction and quality : beyond the Hanseatic**. Leuven : Leuven University Press, 2006, p. 179-188. Disponível em: <<http://repositorium.sdum.uminho.pt/bitstream/1822/5239/1/ArtigoEuroCris2006.pdf>>. Acesso em: 14 abr. 2015.
- PINTO, Carlos Souza; SIMÕES, Claudia; AMARAL, Claudia. CERIF - Is the standard helping to improve CRIS? **Procedia Computer Science**, v. 33, 2014. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1877050914008035>>. Acesso em: 15 abr. 2015.
- RUSSELL, Rosemary. **An introduction to CERIF**. Bath: UKOLN, Jun. 2011. Disponível em: <http://www.ukoln.ac.uk/rim/documents/Introduction_to_CERIF_1.0.pdf>. Acesso em: 14 abr. 2015.
- SALES, Luana Farias. **Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de ciências nucleares**. 2014. 265 f. Tese (Doutorado) - PPGCI, UFRJ/IBICT, Rio de Janeiro, 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>>. Acesso em: 14 abr. 2015.
- SHEPPARD, Nick. Learning how to play nicely: repositories and CRIS. **Ariadne**, n. 64, July 2010. Disponível em: <<http://www.ariadne.ac.uk/issue64/wrn-repos-2010-05-rpt>>. Acesso em: 15 abr. 2015.