# 10. Coleta e reúso de dados de resultados de pesquisa para a constituição de CRIS institucional

*Emanuelle Torino, Caio Saraiva Coneglian e Silvana Aparecida Borsetti Gregorio Vidotti*

## 1 APRESENTAÇÃO

O desenvolvimento tecnológico trouxe diversos avanços ao cenário da comunicação e publicação científica e da disponibilização de dados de pesquisa, que tem sido apoiado por *softwares* desenvolvidos com finalidades específicas, dentre as quais, destacamos nesta pesquisa aqueles utilizados para a publicação e/ou disponibilização de resultados de pesquisa, entendidos neste estudo como, por exemplo, artigos científicos e de dados de pesquisa, trabalhos publicados em eventos científicos, livros e capítulos de livros, e dados de pesquisa.

Nesse contexto, no Brasil, é perceptível que as instituições de pesquisa, sobretudo as universidades, utilizam *softwares* livres, sendo para a publicação, *Open Journal Systems* (OJS), *Open Monograph Press* (OMP) e *Open Conference Systems* (OCS), disponibilizados pelo *Public Knowledge Project* (PKP)[9] "[...] uma iniciativa multiuniversitária que desenvolve *software* de código aberto (gratuito) e realiza pesquisas para melhorar a qualidade e o alcance da publicação acadêmica" (PUBLIC KNOWLEDGE PROJECT, 2022a, tradução nossa). Ao passo que para a disponibilização de resultados de pesquisa em repositórios, há uso expressivo do DSpace[10], que "[...] preserva e permite acesso fácil e aberto a todos os tipos de conteúdo digital, incluindo texto, imagens, vídeos, mpegs e conjuntos de dados." (DSPACE, 2022a, tradução nossa). E, mais recentemente, do Dataverse[11] "[...] um aplicativo da Web de código aberto para compartilhar, preservar, citar, explorar e analisar dados de pesquisa." (DATAVERSE PROJECT, 2022a, tradução nossa).

9 Disponível em: https://pkp.sfu.ca/.

10 Disponível em: https://duraspace.org/dspace/.

11 Disponível em: https://dataverse.org/.

Esses *softwares* são utilizados como ambientes informacionais digitais para a publicação, disponibilização e ampla disseminação dos resultados de atividades acadêmico-científicas, sejam elas oriundas do ensino, da pesquisa e/ou da inovação. Desta forma, atendem à prerrogativa de dar acesso e visibilidade aos resultados de pesquisa.

Contudo, há um aspecto relevante que não é atendido por esses ambientes, o processo de gestão de informações da pesquisa, para o qual, a recomendação é o uso de *Current Research Information System* (CRIS), cuja tradução para o português é Sistema de Informação de Pesquisa Corrente. Nesse sentido, o CRIS se constitui como um ambiente informacional digital para a gestão do ciclo de vida da pesquisa, para tanto, interliga o projeto de pesquisa, os atores neles envolvidos, sejam eles pessoas, instituições e/ou agências de fomento, os recursos investidos, as infraestruturas utilizadas e os resultados desses projetos de pesquisa.

Assim, o CRIS assume um papel preponderante na atividade gerencial da ecologia de pesquisa[12] de uma instituição, sendo capaz de apoiar processos de tomada de decisão acerca da pesquisa e desenvolvimento institucionais, a partir do mapeamento de cenários passado e presente, bem como do planejamento futuro. O CRIS pode ser utilizado ainda para identificar áreas temáticas e pesquisadores relacionados, as relações entre instituições, as redes de coautoria, os projetos similares, os resultados gerados pelos projetos de pesquisa, as necessidades de investimento, dentre outros.

Nesse contexto, diante da necessidade de gerenciar o ciclo de vida da pesquisa, Torino, Coneglian e Vidotti (2020) elaboraram um modelo conceitual de CRIS Institucional a partir da coleta e armazenamento de metadados de resultados de pesquisa oriundos de múltiplas fontes, utilizando-se das estruturas de representação da informação e de infraestrutura semântica. A esta pesquisa, foram incorporados elementos de Inteligência Artificial e Ciência de Dados, visando otimizar os processos e ampliar os valores agregados ao modelo inicial, gerando o Modelo de CRIS institucional com o uso de Inteligência Artificial e Ciência de Dados (CONEGLIAN; TORINO; VIDOTTI, 2021), apresentado na **Erro! Fonte de referência não encontrada**..

12 Entendida como "[...] o relacionamento entre informações dispersas em diferentes sistemas, que gerenciam dados de organização, infraestrutura para pesquisa, projetos de pesquisa, grupos de pesquisa, pesquisadores, fomentos, resultados de pesquisa." (TORINO; CONEGLIAN; VIDOTTI, 2020), sejam eles provenientes de sistemas internos ou externos.

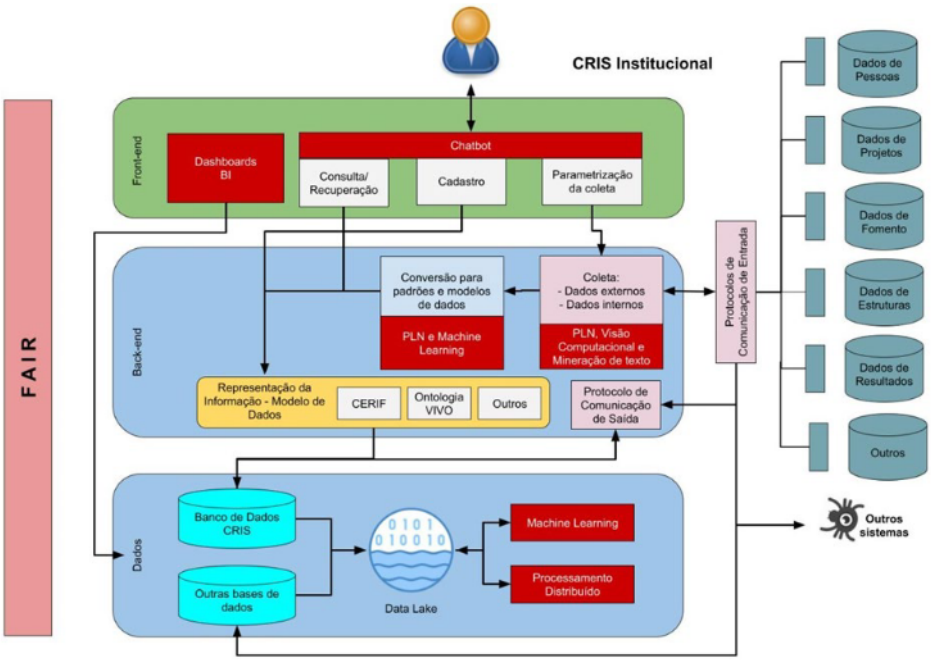


Figura 10-1 - Modelo de CRIS Institucional com o uso de Inteligência Artificial e Ciência de Dados

Fonte: Coneglian, Torino e Vidotti (2021).

Diante do exposto, vale destacar que a constituição de CRIS Institucional, considerando os modelos supramencionados, requer a análise e a compatibilização das estruturas de representação da informação existentes nas plataformas das quais os dados serão coletados, sejam sistemas internos ou externos à instituição. Tal processo pode requerer, ainda, a compatibilização das estruturas de representação e a conversão de registros, visando à manutenção adequada dos dados devidamente estruturados no Banco de Dados CRIS.

Nesse contexto, o presente estudo apresenta o processo de coleta e reúso de dados e objetos digitais dos ambientes informacionais digitais utilizados para a publicação e disponibilização dos resultados de pesquisa, notadamente aqueles utilizados para a publicação de periódicos científicos e de dados de pesquisa, anais de eventos científicos, livros e capítulos de livros, bem como daqueles utilizados para a disponibilização de repositórios digitais de produção acadêmico-científica e de dados de pesquisa para a constituição de CRIS Institucional. Vale destacar que, na Figura 1, esse processo compreende, por um lado, ambientes externos à instituição, representados como 'Dados de resultados' e, por outro, no CRIS Institucional, a parametrização da coleta, os protocolos de comunicação de entrada, a coleta, a conversão de registros e o armazenamento no 'Banco de Dados CRIS'.

Para tanto, como procedimento metodológico, utilizou-se a revisão bibliográfica para o embasamento teórico-conceitual a fim de contextualizar as temáticas abordadas, bem como a pesquisa documental relacionada aos *softwares* analisados. Tais aportes teóricos contribuíram para a discussão de como o processo de coleta e reúso de dados de múltiplas fontes para a constituição de CRIS Institucional pode ser realizado.

Como resultados, são apresentados os elementos necessários para a análise, coleta, conversão de registros e armazenamento de dados para a constituição de CRIS Institucional. Tal estudo apresenta uma contribuição significativa para a comunidade da Ciência da Informação, a partir de uma visão da utilização de preceitos já consolidados na área para uma prática emergente e, ao mesmo tempo, colabora com a comunidade científica em geral ao apresentar elementos importantes para a constituição de CRIS Institucional.

## 2 SISTEMAS DE PUBLICAÇÃO DE RESULTADOS DE PESQUISA CIENTÍFICA

O processo editorial para a publicação de resultados de pesquisa científica tem sido apoiado por diversos *softwares*, sejam eles abertos ou proprietários. O movimento de Acesso Aberto, iniciado no final dos anos 1990, trouxe luz à demanda de abertura das publicações, visando tornar os resultados de pesquisa científica mais acessíveis a pesquisadores e à sociedade, com isso, houve o incentivo na adoção de *softwares* livres, o que ocorreu sobretudo nas publicações vinculadas a Universidades.

Nesse sentido, os sistemas utilizados para a publicação de resultados de pesquisa científica, em geral, são os desenvolvidos e disponibilizados, de forma aberta, pelo (PKP): OJS, OMP e OCS, utilizado para a gestão de eventos científicos e para a publicação de trabalhos publicados em eventos científicos, cuja composição final são os anais de eventos científicos; e o *Open Harvester Systems* (OHS), um provedor de dados que possibilita a indexação de metadados disponíveis nos *softwares* de publicação, atuando como sistema de descoberta (PUBLIC KNOWLEDGE PROJECT, 2022a). Neste estudo, abordaremos o OJS, o OMP e o OCS, que visam apoiar o processo de gestão de publicações científicas em formato digital.

O OJS é utilizado para apoiar todo o processo de gestão editorial de periódicos científicos e de periódicos de dados de pesquisa, desde a submissão até a

publicação. Sua adoção por periódicos, sobretudo para publicações vinculadas a Universidades, tem sido crescente.

O OJS atualmente está na versão 3x, cuja documentação[13] que descreve as convenções e conceitos de código está disponível na web.

Os metadados do OJS podem ser armazenados e exportados no formato Dublin Core, porém, é possível adotar outros padrões e modelos.

O OMP tem uso indicado para Editoras, para a gestão do processo editorial, que envolve submissão, publicação e comercialização de livros. O OMP atualmente está na versão 3x, cuja documentação[14] que descreve as convenções e conceitos de código está disponível na web.

Os metadados do OMP estão em *ONIX for Books Product Information Format*, um padrão de metadados que objetiva fornecer aos editores e demais envolvidos na publicação e comercialização de livros, uma estrutura consistente para a representação e o intercâmbio de dados (EDITEUR, 2022).

O OCS é utilizado para a gestão de eventos científicos e para a publicação de trabalhos publicados em eventos científicos, cuja composição final são os anais de eventos. O OCS atualmente está na versão 2x, cuja documentação[15] que descreve as convenções e conceitos de código está disponível na web.

Os metadados do OCS podem ser armazenados e exportados no formato Dublin Core, porém, é possível adotar outros padrões e modelos.

O desenvolvimento do OCS foi suspenso em 2018[16], para que os esforços fossem concentrados em outros sistemas do PKP. Isso tornou possível a modernização do OJS e do OMP.

Tanto no OJS quanto no OMP, a entrada de dados pode ser realizada de forma manual, por meio da submissão realizada pelo autor e de *plugins* disponíveis no próprio sistema, ou, ainda, por linha de comando. Além disso, alguns *plugins*

13 Disponível em: https://docs.pkp.sfu.ca/dev/documentation/en/.

14 Disponível em: https://docs.pkp.sfu.ca/#appomp3.

15 Disponível em: https://docs.pkp.sfu.ca/#appocs2.

16 Disponível em: https://docs.pkp.sfu.ca/faq/en/*software*-features#can-i-use-ojs-to-publish-conference-proceedings-what-happened-to-ocs.

que possibilitam a exportação de dados com finalidades específicas, entre as quais o registro de *digital object identifier* (DOI) e a exportação de artigos para alguns indexadores. Vale destacar que os *plugins* estão disponíveis e podem ser facilmente habilitados, para que possam ser utilizados para a importação e a exportação de dados utilizando a interface gráfica (PUBLIC KNOWLEDGE PROJECT, 2022b).

Além dos plugins de exportação, cujo uso está limitado aos usuários com permissão de administrador ou gerente dos sistemas, os registros armazenados no OJS, OCS e OMP podem ser disponibilizados, por meio da configuração, para que sejam coletados por qualquer interessado, por meio de protocolos, sendo o mais utilizado o *Open Archives Initiative - Protocol for Metadata Harvesting* (OAI-PMH). No caso de integração de dados de sistemas disponíveis na mesma instituição, seja para a coleta e disponibilização em repositório digital ou ainda para a constituição de CRIS, é possível utilizar ainda outros protocolos disponíveis no OJS e OMP, a exemplo do SWORD e do REST. No que tange ao OCS, considerando que o desenvolvimento está parado há alguns anos, há limitações no uso de APIs.

Ambientes informacionais digitais como periódicos, anais de eventos, e editoras são fontes primárias de publicação de resultados de pesquisa, consideradas provedores de conteúdo, com isso, por meio do uso de protocolos de comunicação de saída podem fornecer dados para provedores de dados e de serviços, a exemplo dos indexadores, repositórios digitais e do CRIS.

## 3 SISTEMAS DE DISSEMINAÇÃO DE RESULTADOS DE PESQUISA CIENTÍFICA

Os resultados de pesquisa, uma vez publicados, por meio dos sistemas de publicação, a exemplo do OJS, OMP e OCS, podem ser disponibilizados em repositórios digitais, sejam eles institucionais ou temáticos.

No contexto do Acesso Aberto o DSpace é um *software* amplamente utilizado para a implementação de repositórios digitais, sobretudo os que disponibilizam objetos digitais resultantes da pesquisa, quando publicados em uma fonte primária. O DSpace possui código aberto, foi desenvolvido inicialmente pelo Massachusetts Institute of Technology (MIT) e pela Hewlett-Packard (HP) e atualmente mantido pela LYRASIS, teve como responsáveis até a versão 6x um grupo de desenvolvedores designados como *DSpace Committers Group*, responsáveis

por manter o código, receber e analisar contribuições ao código fonte, analisar e corrigir bugs e fornecer suporte. Contudo, a partir da versão 7x esse Comitê foi extinto, passando a figurar no código em *Release Notes* os contribuidores.

O DSpace atualmente está na versão 7x, cuja documentação[17] que descreve as convenções e conceitos de código está disponível na web.

No que tange ao armazenamento de dados, o DSpace suporta metadados em formato Dublin Core. A documentação do *software* se refere ao suporte desse esquema de metadados "[...] significa que os metadados podem ser inseridos no DSpace, armazenados no banco de dados, indexados adequadamente e tornados pesquisáveis por meio da interface de usuário pública. Atualmente, isso se aplica principalmente a metadados descritivos, embora, à medida que os padrões surjam, também possam incluir metadados técnicos, de direitos, de preservação, estruturais e comportamentais." (DSPACE, 2022b, tradução nossa).

A entrada de dados pode ser realizada de forma manual ou automática. A entrada manual pode ocorrer por autoarquivamento por parte do autor ou ser mediada por um terceiro que, autorizado por ele, se responsabilize pela inserção do conteúdo no repositório, atividade esta, em geral, realizada por bibliotecários. Enquanto a entrada automática pode ser realizada por meio dos protocolos e/ou serviços disponíveis no DSpace, além da possibilidade da inserção por linha de comando.

No avanço do movimento de Acesso Aberto expandiu as discussões e, a chamada Ciência Aberta, dentre outras coisas, trouxe luz à necessidade de gestão, compartilhamento e reúso de dados de pesquisa. Neste contexto, o Dataverse tem sido o *software* de código aberto mais utilizado para a implementação de repositórios digitais de dados de pesquisa.

O Dataverse é desenvolvido pelo Institute for Quantitative Social Science (IQSS) de Harvard e possui muitos colaboradores em todo o mundo (DATAVERSE PROJECT, 2022a). Atualmente o Dataverse está na versão 5x, cuja documentação[18] que descreve as convenções e conceitos de código está disponível na web.

17 Disponível em: https://wiki.lyrasis.org/display/DSPACE/Documentation.

18 Disponível em: https://wiki.lyrasis.org/display/DSPACE/Documentation.

Para a entrada de dados, o Dataverse suporta metadados em diferentes esquemas padronizados, dentre eles, DataCite, Dublin Core e Schema.org e, busca assegurar que os metadados possam ser mapeados para outros esquemas e exportados em JSON, visando interoperabilidade e preservação (DATAVERSE PROJECT, 2022b). Neste contexto, o próprio Dataverse Project disponibiliza um documento no qual há um *crosswalk* entre os esquemas suportados[19].

A entrada de dados pode ser realizada de forma manual ou automática, sendo que a segunda pode ser realizada por meio dos protocolos e/ou serviços disponíveis no Dataverse, dentre eles, é possível utilizar o OAI-PMH para configurar coletas de outros repositórios para um Dataverse. Além disso, há a possibilidade da inserção por linha de comando.

Os registros armazenados no DSpace e no Dataverse podem ser disponibilizados para a coleta por meio do Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH), e dispõem ainda de Application Programming Interface (API) que possibilita a interoperabilidade com outros *softwares*.

Vale destacar que uma das finalidades dos repositórios digitais é a ampliação da visibilidade dos conteúdos por eles armazenados. Contudo, deve-se ter em mente que os resultados de pesquisa disponibilizados em repositórios digitais podem ter sido publicados anteriormente, por exemplo em periódicos, anais de eventos ou por editoras. Porém, os repositórios digitais podem ser fontes de publicação, por exemplo, no caso da documentação que acompanha dados de pesquisa (TORINO, ROA-MARTINEZ, VIDOTTI, 2020) ou ainda quando é o ambiente de publicação de livros de editora, e, ainda, ser fonte de disponibilização/publicização, no caso de trabalhos acadêmicos, recursos educacionais abertos e dados de pesquisa que não acompanhem documentação.

Destaca-se ainda que, os repositórios digitais podem ser provedores de dados, para os registros que possuem representação e *link* para acesso ao original na fonte de publicação e de conteúdos, quando, além disso, disponibiliza o objeto digital contendo o texto completo. Desta forma, é importante que os repositórios digitais realizem o enlace de registros relacionados.

Com isso, por meio do uso de protocolos de comunicação de saída podem fornecer dados para provedores de dados e de serviços, a exemplo dos indexadores, repositórios digitais e CRIS.

19 Disponível em: https://docs.google.com/spreadsheets/d/10Luzti7svVTVKTA-px27oq3RxCUM-QbiTkm8iMd5C54/edit#gid=587531992.

# 4 PROTOCOLOS DE COMUNICAÇÃO

## 4.1 OPEN ARCHIVES INITIATIVE - PROTOCOL FOR METADATA HARVESTING (OAI-PMH)

Um dos protocolos mais relevantes para a interoperabilidade de metadados de publicação científica é o OAI-PMH. Esse protocolo foi proposto no âmbito da *Open Archives Initiative*, com o intuito de favorecer a interoperabilidade entre os distintos recursos.

Neste contexto, aponta-se que a iniciativa do protocolo OAI-PMH possui papel importante no aspecto do compartilhamento de metadados. A sua primeira versão foi lançada em 2001, sendo depois lançada uma versão 2.0, que é a última versão, datada de julho de 2002 (GARCIA; SUNYE, 2003). Tal protocolo utiliza o Dublin Core como padrão de metadados para possibilitar o intercâmbio das informações, permitindo, assim, que a interoperabilidade possa ocorrer dentro dos ambientes atendidos.

Um aspecto importante de se destacar é que por meio do uso do protocolo OAI-PMH, um sistema pode obter os metadados dos objetos que estão contidos em um repositório. Para tal, basta realizar uma requisição para o endereço (URL) onde o repositório se encontra, além de inserir as configurações daquela coleta. Como resultado, é retornada uma lista que contém todos os metadados daquele repositório ou ambiente de publicação científica.

## 4.2 OPEN ARCHIVES INITIATIVE - OBJECT REUSE AND EXCHANGE (OAI-ORE)

Outro protocolo de destaque, que está vinculado ao contexto da *Open Archives Initiative*, é o *Object Reuse and Exchange* (OAI-ORE). Tal protocolo tem como objetivo apoiar e definir como os recursos podem ser interoperáveis entre os distintos ambientes digitais.

Destaca-se que este protocolo busca criar "[...] padrões que generalizem todas as informações baseadas na web, incluindo as redes sociais cada vez mais populares da "web 2.0." (OPEN ARCHIVES, 2014, tradução nossa). O que demonstra que tal protocolo está focado em permitir a troca de dados em contextos mais genéricos e amplos do que o OAI-PMH.

Um dos aspectos mais relevantes do OAI-ORE é solucionar o problema da agregação e da compreensão por parte dos usuários humanos de páginas web. Por meio deste protocolo, estaria claro quais aspectos em uma página tratam de ligações e conteúdos vinculados ao conteúdo apresentado nesta mesma página, e quais os outros aspectos que tem como objetivo apenas apoiar a navegação do usuário para outras páginas e ambientes.

Desta forma, tal protocolo apoia a coleta de dados de ambientes digitais com a contextualização dos links e informações descritas em tal ambiente.

## 4.3 REPRESENTATIONAL STATE TRANSFER (REST)

O *Representational State Transfer* (REST) é uma arquitetura que tem como objetivo realizar a transferência de dados, por meio do seu estado representacional. Tal arquitetura foi criada por Roy Fielding no ano de 2000, e hoje é a principal forma para realizar a transferência de informações entre aplicações.

Em uma documentação técnica sobre o REST, aponta-se que "Quando um cliente faz uma solicitação usando uma API RESTful, essa API transfere uma representação do estado do recurso ao solicitante ou endpoint. Essa informação (ou representação) é entregue via HTTP utilizando um dos vários formatos possíveis [...]" (RED HAT, 2020).

Em linhas gerais, a aplicação entrega uma representação de um objeto, com todas as suas características, permitindo assim, que essa outra aplicação consiga ter acesso ao estado geral de tal aplicação. Para isso, pode-se utilizar diversos protocolos e linguagens, como *JavaScript Object Notation* (JSON), *eXtensible Markup Language* (XML), *HyperText Markup Language* (HTML), entre outros. No entanto, vale apontar que o JSON é o formato mais utilizado nos protocolos que implementam o REST.

## 4.4 SIMPLE OBJECT ACCESS PROTOCOL (SOAP)

O *Simple Object Access Protocol* (SOAP) é um dos protocolos mais utilizados para a transferência de recursos dentro da Web. Ainda que nos últimos anos, o SOAP perdeu relevância devido a adoção cada vez maior das aplicações na utilização do REST, o SOAP continua sendo utilizado por uma série de aplicações.

Em uma documentação técnica, o SOAP é apresentado como: "[...] um protocolo leve e baseado em XML para troca de informações em um ambiente descentralizado e distribuído. Ao combinar solicitações e respostas baseadas em SOAP com um protocolo de transporte, como HTTP, a Internet se torna um meio para os aplicativos publicarem serviços *on-line* baseados em banco de dados" (ORACLE, 2001, tradução nossa).

Por meio do SOAP, duas aplicações construídas em linguagens e com arquiteturas distintas podem trocar informações, o que favorece a interoperabilidade dos dados. Tal aspecto é essencial para que os sistemas possam ser integrados e possam colaborar um com o outro.

Das características do SOAP, destaca-se que a comunicação acontece com independência de protocolo, de idioma, de plataforma e de sistema operacional. Outra característica é que o SOAP funciona a partir do XML e está estruturado a partir do envio de solicitações de um cliente para um servidor, tendo como o retorno os dados solicitados, chamados de resposta SOAP.

Diante dos tópicos apontados, verifica-se que o SOAP é mais um protocolo que permite a transferência de dados, sendo utilizado em diversas aplicações no âmbito da Web.

## 4.5 SUBMISSION INFORMATION PACKAGE (SIP)

O *Submission Information Package* (SIP) é um protocolo focado em repositórios e outros ambientes de armazenamento digital, para que os ambientes consigam os metadados, mas também os próprios recursos, como imagens e vídeos. Destaca-se ainda que esse protocolo foca em aspectos de preservação digital, ao permitir que os objetos digitais possam ser armazenados de forma integral, seguindo as recomendações de preservação.

"[O] SIP inclui os dados de áudio a serem armazenados e todos os metadados relacionados necessários sobre o objeto e seu conteúdo. [Além disso, realiza o processo de] Ingest, no modelo OAIS, que aceita o conteúdo e todos os seus metadados relacionados (SIP), verifica o arquivo, extrai os dados relevantes e prepara o AIP para armazenamento e garante que os AIPs e suas informações descritivas de suporte sejam estabelecidas dentro de o OAIS" (IASA, 2022, tradução nossa).

### 4.6 SIMPLE WEB-SERVICE OFFERING REPOSITORY DEPOSIT (SWORD)

Outro protocolo utilizado para a transferência de dados em ambientes informacionais digitais é o Simple Web-service Offering Repository Deposit (SWORD), que está buscando aprimorar os aspectos de interoperabilidade em distintos ambientes. A atual versão é a 3.0 e trouxe uma série de avanços para facilitar a troca de conteúdos entre objetos digitais complexos.

"O SWORD 3.0 é um protocolo que permite que clientes e servidores se comuniquem em torno de objetos digitais complexos, especialmente no que diz respeito ao suporte ao depósito desses objetos em um serviço como um repositório digital" (JONES, 2021).

De forma complementar, aponta-se que o SWORD compreende objetos digitais complexos, como metadados, junto do próprio arquivo, dos mais distintos formatos e tamanhos. Ademais, o protocolo busca definir os elementos vinculados à criação, anexos, substituição, exclusão e recuperação das informações.

## 5 COLETA DE DADOS DE RESULTADOS DE PESQUISA PARA CRIS INSTITUCIONAL

A partir da reflexão sobre o Modelo CRIS Institucional (**Erro! Fonte de referência não encontrada.**) e da discussão acerca dos protocolos de saída de dados dos ambientes de publicação e disseminação de resultados de pesquisa, disponíveis para propiciar a coleta de dados para diversas finalidades, é necessário aprofundamento nos temas que dizem respeito ao processo da coleta nos diversos ambientes relacionados ao CRIS.

Partindo do Modelo de CRIS Institucional, refletiu-se sobre as etapas e as práticas necessárias para que o processo de coleta de dados aconteça de forma satisfatória. Para tanto, além da coleta, é necessária a conversão dos registros coletados para o modelo de dados utilizado no âmbito do CRIS Institucional.

Neste estudo, a enfatiza-se a coleta de dados de fontes externas à Instituição, apresentadas no Modelo de CRIS Institucional como "Dados de resultados de pesquisa", que consistem nos ambientes informacionais digitais utilizados para a publicação e a disseminação de objetos resultantes da pesquisa, a exemplo de periódicos científicos, anais de eventos científicos, editoras de livros e repositórios acadêmico-científicos e de dados de pesquisa.

Para elucidar o processo de coleta de dados utilizando os protocolos disponíveis, realizou-se uma coleta de dados utilizando o OAI-PMH no periódico científico Ciência da Informação, um periódico tradicional mantido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict). Para tal, realizou-se a coleta de todos os registros da revista no padrão Dublin Core por meio da URL <https://revista.ibict.br/ciinf/oai?verb=ListRecords&metadataPrefix=oai_dc>. Assim, apresenta-se na Figura 10-2 um fragmento dos dados obtidos.

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="https://revista.ibict.br/lib/pkp/xml/oai2.xsl" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
        http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
    <responseDate>2022-05-31T01:42:31Z</responseDate>
    <request verb="ListRecords" metadataPrefix="oai_dc">https://revista.ibict.br/ciinf/oai</request>
    <ListRecords>
        <record>
            <header>
                <identifier>oai:ojs.revista.ibict.br:article/1</identifier>
                <datestamp>2019-04-05T13:15:57Z</datestamp>
                <setSpec>ciinf:ART</setSpec>
            </header>
            <metadata>
<oai_dc:dc
    xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
    xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
    http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
    <dc:title xml:lang="pt-BR">Da Bibliografia à Ciência da Informação: Um Histórico e Uma Posição</dc:title>
    <dc:creator>Zaher, Célia Ribeiro</dc:creator>
    <dc:creator>Gomes, Hagar Espanha</dc:creator>
    <dc:subject xml:lang="pt-BR">Bibliografia. Ciência da informação</dc:subject>
    <dc:description xml:lang="pt-BR">Novas formas de registro de informações ampliaram o âmbito da Bibliografia, levando ao aparecimento da Documentação. Necessidades socia
    <dc:publisher xml:lang="pt-BR">Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict)</dc:publisher>
    <dc:date>1972-06-30</dc:date>
    <dc:type>info:eu-repo/semantics/article</dc:type>
    <dc:type>info:eu-repo/semantics/publishedVersion</dc:type>
    <dc:type xml:lang="en-US">Peer-reviewed articles</dc:type>
    <dc:type xml:lang="es-ES">Evaluado por pares</dc:type>
    <dc:type xml:lang="pt-BR">Avaliado por pares</dc:type>
    <dc:format>application/pdf</dc:format>
    <dc:identifier>https://revista.ibict.br/ciinf/article/view/1</dc:identifier>
    <dc:identifier>10.18225/ci.inf..v111.1</dc:identifier>
```


Figura 10-2 - Fragmento dos dados coletados na Revista Ciência da Informação por meio do Protocolo OAI-PMH

Fonte: Obtido a partir da consulta via OAI-PMH em https://revista.ibict.br/ciinf/oai?verb=ListRecords&metadataPrefix=oai_dc.

O processo de coleta de dados via OAI-PMH é chamado de *harvesting*, em que basicamente é possível obter todos os metadados disponíveis em uma determinada fonte de informação estando estruturados em um determinado padrão, como o Dublin Core. Ademais, aponta-se que os dados podem ser ordenados seguindo a data da publicação, permitindo que a coleta para o CRIS possa obter os metadados dos registros disponíveis em um período temporal. No caso do registro apresentado na Figura 3, foram obtidos todos os registros disponíveis naquela revista.

Outra forma muito utilizada para a coleta de dados é via API REST, que permite com que dados relativos a todo o processo de submissão, publicação e envio possam ser obtidos. O exemplo apresentado na Figura 10-3 demonstra como os dados de todas as submissões de um autor foram obtidos a partir de uma requisição via API REST. No caso, utilizou-se como exemplo a revista

Encontros Bibli[20], por meio da seguinte requisição <https://periodicos.ufsc.br/index.php/eb/api/v1/submissions>.


```
{"itemsMax":4,"items":[{"_href":
"https:\/\/periodicos.ufsc.br\/index.php\/eb\/api\/v1\/submissions\/67665","contextId":2,
"currentPublicationId":55285,"dateLastActivity":"2020-05-08 15:19:46","dateSubmitted":"2019-09-19
20:22:24","doiSuffix":null,"id":67665,"lastModified":"2020-09-03 14:48:03","pub-id::doi":null,
"publications":[{"_href":
"https:\/\/periodicos.ufsc.br\/index.php\/eb\/api\/v1\/submissions\/67665\/publications\/XXXXXX",
"authorsString":"","authorsStringShort":"","categoryIds":[],"coverImage":{"en_US":null,"es_ES":
null,"pt_BR":null},"datePublished":"2020-05-08","doiSuffix":null,"fullTitle":{"en_US":"XXXXXXXX",
"pt_BR":"XXXXXXXXXX","es_ES":""},"galleys":[{"dependentFiles":[],"doiSuffix":null,"file":{
"fileName":"XXXX","id":99999,"revision":2,"fileStage":10,"genreId":85},"fileId":258586,"id":43138,
"isApproved":true,"label":"Artigo PDF\/A","locale":"pt_BR","pub-id::doi":null,
"pub-id::publisher-id":null,"publicationId":55285,"seq":0,"urlPublished":
"https:\/\/periodicos.ufsc.br\/index.php\/eb\/article\/view\/xxxxxxx","urlRemote":null},{
"dependentFiles":[],"doiSuffix":null,"file":{"fileName":"xxxx-Texto do Artigo-xxxx.pdf","id":
258744,"revision":1,"fileStage":10,"genreId":85},"fileId":258744,"id":43151,"isApproved":true,
"label":"Parecer do Artigo","locale":"pt_BR","pub-id::doi":null,"pub-id::publisher-id":null,
"publicationId":55285,"seq":1,"urlPublished":
"https:\/\/periodicos.ufsc.br\/index.php\/eb\/article\/view\/xxxxxxx","urlRemote":null}],"id":
99999,"locale":"pt_BR","pages":"01-25","prefix":{"en_US":"","es_ES":"","pt_BR":""},
"primaryContactId":999999,"pub-id::doi":"10.5007\/99999999","pub-id::publisher-id":"150999999",
"sectionId":5,"status":3,"submissionId":999999,"subtitle":{"en_US":"","es_ES":"","pt_BR":""},
"title":{"en_US":"XXXXXXXX","pt_BR":"XXXXXXXXXXX","es_ES":""},"urlPublished":
"https:\/\/periodicos.ufsc.br\/index.php\/eb\/article\/view\/9999999.e67665\/version\/999999",
"version":1}],"stageId":4,"status":3,"statusLabel":"Publicado","submissionProgress":0,
"urlAuthorWorkflow":
```


Figura 10-3 - Fragmento dos dados coletados na Revista Encontros Bibli por meio da requisição em um API REST

Fonte: Obtido a partir da consulta via API REST em https://periodicos.ufsc.br/index.php/eb/api/v1/submissions.

Na Figura 10-3, os dados são recebidos no formato JSON, e apontam diversas informações acerca das submissões realizadas por determinado autor. Destaca-se que o processo de coleta via API REST permite o acesso a maior quantidade de informações, mas com a necessidade de estar logado no sistema, ou tendo a chave de acesso à API.

Logo, no caso das revistas que são da própria instituição do CRIS, é possível obter o acesso à chave da API do administrador, o que possibilita a obtenção dos dados apenas dos autores que estão vinculados à instituição. Ademais, para se obter dados de outros autores, é necessário que tenha-se o acesso do usuário à API, o que permite o acesso apenas dos registros do autor.

Vale destacar que esse é o funcionamento da API REST do OJS, e que isso é definido a partir da concepção e da implementação de cada API de acordo com o *software* e com os dados a serem obtidos.

No que tange ao SWORD, é possível utilizá-lo entre ambientes da mesma instituição, visando automatizar o processo de envio dos dados, no caso de

20 Disponível em: https://periodicos.ufsc.br/index.php/eb.

um ambiente digital de publicação de resultados de pesquisa para um repositório digital ou CRIS Institucional. Nesse contexto é possível acessar na Ibict Wiki um documento que explicita o processo de configuração do protocolo SWORD nos sistemas DSpace e OJS[21].

Vale ainda destacar que, para realizar a coleta e o reúso de dados disponíveis em outras fontes, é necessário considerar que existe um ciclo de vida de dados (CVD) relacionado a todos os ambientes informacionais digitais, apresentados em quatro fases: coleta, armazenamento, recuperação e descarte (SANT'ANA, 2013; 2016). Nesse sentido, é primordial considerar que o detentor do CVD é o responsável pela tomada de decisões com relação ao ambiente informacional que gerencia. Assim, compete a ele as definições relacionadas aos processos de coleta, armazenamento, recuperação e descarte.

No contexto dos dados de resultados de pesquisa, como tratados neste estudo, o detentor é a organização ou sujeito responsável pela estruturação e processos no CVD, considerando cada fonte de publicação e/ou disseminação dos resultados de pesquisa. Assim, a coleta e o reúso de dados para a constituição de CRIS Institucional requerem que o detentor do CVD tenha estruturado a recuperação de forma a possibilitar a consulta, requisição e coleta dos dados armazenados.

Torino, Vidotti e Sant'Ana (2019) discutem os ciclos de vida de dados e atores no processo de publicação e acesso à produção científica, que contribuem para o entendimento das questões elencadas.

Essa compreensão do CVD dos ambientes informacionais é parte do processo inicial da coleta, que consiste na análise dos ambientes de forma individual, a fim de compreender a disponibilidade de dados para a coleta, os protocolos/serviços de comunicação de saída disponíveis, o modelo de dados utilizados e os metadados que são fornecidos, considerando as especificidades no fornecimento de cada protocolo.

Por outro lado, no ambiente CRIS Institucional, é necessário domínio dos protocolos/serviços de comunicação de saída disponíveis no ambiente do qual os dados serão coletados e, a partir da compreensão das especificidades desse ambiente, configurar e parametrizar a coleta no CRIS.

21 Disponível em: http://wiki.ibict.br/index.php/Configurar_protocolo_SWORD_nos_sistemas_DSpace_e_SEER/OJS.

Essa parametrização de coleta, no Modelo de CRIS Institucional, pode ser realizada por usuário com permissão específica, diretamente no *front-end*, para que ocorra de modo independente dos códigos de programação. Contudo, nada impede que esse processo ocorra diretamente no *back-end*, ou seja, realizar a definição das fontes e de seus parâmetros diretamente em linha de comando

A parametrização da coleta de dados no CRIS consiste na definição dos diferentes elementos a serem utilizados no processo de coleta de um ambiente informacional específico, que inclui os seguintes parâmetros: endereço de acesso, protocolo/serviço de comunicação a ser utilizado, variável(eis) utilizadas para a requisição e modelo de dados e *crosswalk* de metadados, considerando os dados do CRIS Institucional.

Quando disponível no ambiente dos quais os dados são coletados, e considerando o protocolo de saída de dados disponível no ambiente do qual os dados serão coletados, é possível ainda incluir na requisição a coleta do objeto digital; ou, caso o ambiente não possua um protocolo que permita a coleta de objetos digitais, pode-se coletar o metadado que exibe a URL de acesso a ele para, posteriormente, realizar o *download* e o armazenamento.

Nesse processo, caso seja necessário, é possível configurar quais elementos de metadados serão coletados, considerando a possibilidade de que haja algum que não seja compatível com o perfil de aplicação do CRIS Institucional. Da mesma maneira, podem ser utilizadas técnicas como processamento de linguagem natural (PLN), visão computacional e mineração de texto para extrair conteúdos e metadados existentes, bem como apoiar a descoberta de termos e conceitos significativos que são parte dos objetos digitais e, com isso, enriquecer a representação da informação.

No que tange à variável de consulta, devem ser priorizados os identificadores persistentes (PID). Nesse sentido, considerando que, neste estudo, abordamos o CRIS Institucional, uma das variáveis de consulta possíveis é o PID institucional, ou a forma padronizada de representação do nome da Instituição mantenedora do CRIS. Uma prática recomendada é a adoção de uma forma padronizada de representação do nome da Instituição e de um identificador persistente de instituição, a exemplo do *Research Organization Registry* (ROR)[22].

22 Disponível em: https://ror.org/.

O ROR atua como PID para instituições relacionadas à pesquisa, que objetiva ampliar o uso de identificadores persistentes para instituições e permitir conexões entre os registros das instituições em vários sistemas, incluindo-os na infraestrutura acadêmico-científica e nos metadados de diferentes ambientes informacionais, propiciando a descoberta e o rastreamento eficiente dos resultados de pesquisas de instituições e agências de fomento (RESEARCH ORGANIZATION REGISTRY, 2022).

Evidentemente, é possível realizar a coleta utilizando outras variáveis de consulta, a exemplo do nome da instituição e, caso não esteja padronizado, suas variáveis, os identificadores persistentes de pesquisadores, a exemplo do *Open Researcher and Contributor ID* (ORCID), os nomes e variantes de nomes de pesquisadores, entre outros. É possível, ainda, a inclusão de mais de uma variável por coleta em uma mesma base de dados. Contudo, a adoção de PID torna os resultados das requisições mais precisos e a coleta mais efetiva.

Na coleta, outro processo que merece destaque é o *crosswalk* de metadados, visando compatibilizar o modelo de dados do ambiente informacional coletado e o CRIS Institucional, o que mantém a precisão da base de dados CRIS. Para tanto, é imprescindível que haja um modelo de dados CRIS especificado em um perfil de aplicação, para o qual os dados oriundos da coleta devem ser mapeados e compatibilizados. Essa etapa pode requerer processamento, que pode ser realizado por meio de PLN, mineração de texto, *machine learning* e/ou o uso de ontologias para apoiar a compreensão do contexto dos textos.

Ademais, devem ser inseridos metadados de proveniência para indicar a fonte original da qual os dados foram coletados e metadados de preservação, para registrar alterações que possam ter sido realizadas, visando contribuir para a consistência e a credibilidade.

Merece destaque a necessidade de manutenção dos metadados estruturados de acordo com o modelo do CRIS Institucional, e que a camada de "Representação da Informação" deve atender aos princípios FAIR.

Considerando a recorrência de coletas às mesmas fontes de dados, após o processo de coleta e conversão de registros, é necessário realizar a deduplicação, a fim de identificar duplicidade de registros entre artigos, trabalhos publicados em eventos e livros coletados da fonte primária de publicação e, posteriormente, de repositórios digitais ou, ainda, dos dados coletados e os que são armazenados na base de dados CRIS.

A deduplicação exige análise dos registros para evitar que sejam descartados ou incorporados registros similares, mas não correspondentes. Nesse sentido, também é possível definir uma ou mais variáveis para o processamento, entre as quais destacam-se novamente os PID, como Digital Object Identifier (DOI), Handle System, ORCID iD, ROR, embora não se limitem a ela. A deduplicação pode ser processada pelo próprio CRIS ou com o uso de um *software* específico para a finalidade.

É imprescindível, ainda, averiguar a qualidade dos dados para se certificar que os registros estão relacionados à Instituição, os metadados estão corretos e completos e, quando coletados, os objetos digitais não estejam corrompidos. Nesse processo, pode ser necessário o cruzamento dos dados coletados e os disponíveis em outros sistemas da instituição. Para tal, devem ser privilegiadas coletas de ambientes digitais confiáveis, sobretudo aqueles que possuem certificação do dado disponível.

Realizados tais processos, ocorre a ingestão dos dados e objetos digitais coletados na base CRIS, disponível na camada de dados. Uma vez adequadamente tratados, estruturados e FAIR, tais dados possibilitam análises e interpretações necessárias à gestão das atividades de pesquisa institucionais. Soma-se a isso a possibilidade de ingestão dos dados coletados de forma bruta, sem o tratamento indicado anteriormente, a fim de elaborar um banco de dados de outras bases.

Assim, a união dos registros do Banco de Dados CRIS e do banco de dados de Outras bases de dados constitui um *data lake*, que possibilita a mineração e a descoberta de conhecimento. Ele também pode ser utilizado para finalidade gerencial, por exemplo, na comparação entre os dados brutos e tratados, que podem apontar parâmetros de representação da informação capazes de impactar no posicionamento do impacto da pesquisa institucional, quer seja, por exemplo, no uso inadequado de afiliação, que pode dividir a produção institucional, atrelando parte das publicações a outra(s) instituição(ões) ou, ainda, em elementos como a quantidade de caracteres no título, a presença de símbolos e caracteres especiais no título e no resumo, que prejudicam a coleta e o ranqueamento por mecanismos de busca.

# 6 CONSIDERAÇÕES FINAIS

O advento das tecnologias e dos aparatos computacionais possui potencial para auxiliar o processamento humano de dados, visando apoiar atividades como a gestão da ecologia de pesquisa de uma instituição.

Nesse contexto, considerando que os dados necessários à gestão da ecologia de pesquisa, torna-se relevante o investimento em técnicas e tecnologias capazes de otimizar o processo de coleta de dados visando apoiar a análise dos dados e informações vitais da parte central de uma instituição que atua com pesquisa.

Com isso, a coleta de dados de múltiplas fontes tem sido utilizada para diferentes finalidades, como para o povoamento de repositórios digitais, a criação de *dashboards* de métricas e impacto. Porém, o modelo conceitual de CRIS Institucional apresentado na Figura 1 prevê que fontes internas e externas à instituição sejam utilizadas como provedores de dados para a constituição de CRIS Institucional.

Indubitavelmente o processo de coleta de dados de resultados de pesquisa, materializados na forma de artigos, trabalhos publicados em eventos, livros, capítulos de livros, dados de pesquisa e armazenamento em CRIS Institucional contribuem para o monitoramento do cenário da pesquisa desenvolvida em uma instituição, o que pode colaborar, por exemplo, com indicadores, posicionamento em *rankings*, dotação orçamentária e investimento em infraestrutura.

Para tanto, é necessário que as Instituições interessadas na aplicação do Modelo de CRIS Institucional invistam na definição de políticas que assegurem que as publicações de pesquisadores a ela relacionados devem indicar a afiliação institucional de forma adequada e, quando possível, indicando PID de autor, instituição e gerando PID do objeto digital publicado. Ademais, tais publicações devem ser realizadas preferencialmente em fontes de publicação que possibilitem a coleta dos dados para reúso, por exemplo, no povoamento de repositório digital e na constituição de CRIS Institucional.

No contexto dos pesquisadores, faz-se necessário capacitá-los no que tange aos processos de representação da informação quando da submissão de um manuscrito para a publicação, visando ampliar a compreensão dos impactos que podem decorrer disso.

No contexto das fontes de publicação e disponibilização de resultados de pesquisa discutidas neste estudo – periódicos, eventos, editoras, repositórios digitais

de produção acadêmico-científica e de dados de pesquisa –, é necessário ampliar as discussões e a compreensão quanto aos modelos de dados utilizados, à necessidade da adoção e compartilhamento do perfil de aplicação adotado, além da relevância da configuração dos protocolos de comunicação de saída.

No que tange ao perfil de aplicação e ao modelo de dados das fontes de publicação e disponibilização de resultados de pesquisa discutidas neste estudo, é preponderante a atuação dos profissionais da Ciência da Informação, visando assegurar a adoção adequada de esquemas de metadados, a fim de que se possibilite a representação exaustiva dos conteúdos e contextos a serem representados, além do atendimento aos princípios FAIR.

Já no tocante à coleta e reúso de dados de resultados de pesquisa para CRIS Institucional, reafirmamos nossa recomendação de que apenas os metadados sejam armazenados e os objetos digitais acionáveis a partir de um campo específico de metadados que possibilite o acesso na fonte original de publicação, preferencialmente por meio de PID. Contudo, é possível que a Instituição opte pela constituição do CRIS, juntamente ao Repositório Digital. Nesses casos, os objetos digitais podem ser coletados e armazenados no CRIS Institucional.

## REFERÊNCIAS

CONEGLIAN, C. S.; TORINO, E.; VIDOTTI, S. A. B. G. Inteligência Artificial e Ciência de Dados em CRIS institucional: modelo conceitual. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 11., 2021, Rio de Janeiro. **Anais** [...]. Rio de Janeiro, 2021. Disponível em: https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/337. Acesso em: 17 fev. 2022.

DATAVERSE PROJECT. **About**. Disponível em: https://dataverse.org/about. Acesso em: 17 fev. 2022a.

DATAVERSE PROJECT. **User guide**. Disponível em: https://guides.dataverse.org/en/latest/user/index.html. Acesso em: 17 fev. 2022b.

DSPACE. **About DSpace**. Disponível em: https://duraspace.org/dspace/about/. Acesso em: 17 fev. 2022a.

DSPACE. **Support**. Disponível em: https://wiki.lyrasis.org/display/DSPACE/Support. Acesso em: 02 mar. 2022b.

EDITEUR. **ONIX**. Disponível em: https://www.editeur.org/8/ONIX . Acesso em: 17 fev. 2022.

JONES, R. **SWORD 3.0 Specification**. Disponível em: https://swordapp.github.io/swordv3/swordv3.html. Acesso em: 24 maio 2022.

OPEN ARCHIVES. Disponível em: https://www.openarchives.org/ore/. Acesso em 19 maio 2022.

ORACLE. Simple Object Access Protocol Overview. Disponível em: https://docs.oracle.com/cd/A97335_02/integrate.102/a90297/overview.htm. Acesso em 21 maio 2022.

PUBLIC KNOWLEDGE PROJECT. **Data import and export**. Disponível em: https://docs.pkp.sfu.ca/admin-guide/en/data-import-and-export. Acesso em: 02 mar. 2022b.

PUBLIC KNOWLEDGE PROJECT. Disponível em: https://pkp.sfu.ca/. Acesso em: 17 fev. 2022a.

RED HAT. Api Rest. Disponível em: https://www.redhat.com/pt-br/topics/api/what-is-a-rest-api Acesso em 21 maio 2022.

RESEARCH ORGANIZATION REGISTRY. **About**. Disponível em: https://ror.org/about/. Acesso em 21 maio 2022.

SANT'ANA, R. C. G. Ciclo de vida dos dados e o papel da ciência da informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais** [...]. Florianópolis, 2013.

SANT'ANA, R. C. G. Ciclo de vida dos dados: uma perspectiva a partir da ciência da Informação. **Informação & Informação**, Londrina, v. 21, n. 2, p. 116-142, maio/ago. 2016. Disponível em: http://www.uel.br/revistas/uel/index.php/informacao/article/view/27940. Acesso em: 14 set. 2018.

TORINO, E.; CONEGLIAN, C. S.; VIDOTTI, S. A. B. G. Estruturas de representação para reúso de dados no contexto da ecologia de pesquisa: CRIS Institucional. **Informação & Informação**, Londrina, v. 25, n. 3, p. p. 1-27, jul./set. 2020. DOI: http://dx.doi.org/10.5433/1981-8920.2020v25n3p1. Disponível em: http://www.uel.br/revistas/uel/index.php/informacao/article/view/41946. Acesso em: 17 fev. 2022.

TORINO, E.; ROA-MARTINEZ, S. M.; VIDOTTI, S. A. B. G. Dados de pesquisa: disponibilização ou publicação?. In: SHINTAKU, M.; SALES, L. F.; COSTA, M. (org.). **Tópicos sobre dados abertos para editores científicos**. Botucatu, SP: ABEC, 2020. p. 183-201 Disponível em: http://repositorio.utfpr.edu.br/jspui/handle/1/4725. Acesso em: 03 maio 2022.

TORINO, E.; VIDOTTI, S. A. B. G.; SANT'ANA, R. C. G. Ciclo de vida de dados no processo de publicação e acesso à produção científica. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20. 2019, Florianópolis. **Anais** [...]. Florianópolis: Universidade Federal de Santa Catarina, 2019. Disponível em: https://conferencias.ufsc.br/index.php/enancib/2019/paper/viewFile/576/612. Acesso em: 15 maio 2022.

**Como citar este capítulo:**

TORINO, Emanuelle; CONEGLIAN, Caio Saraiva; VIDOTTI, Silvana Aparecida Borsetti Gregorio. Coleta e reúso de dados de resultados de pesquisa para a constituição de CRIS institucional. In: SANTOS, Gildenir Carolino; SHINTAKU, Milton (org.). **Ecossistemas e inovações tecnológicas**: da construção as boas práticas. Campinas: UNICAMP/BCCL; Brasília: Ibict, 2022. Capítulo 10, p. 225-246. DOI: 10.22477/ISBN9786588816363.cap10