# Interface do gestor: uma dashboard para gestores de repositórios digitais

**JULIANA ARAUJO GOMES DE SOUSA**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA

julianasousa@ibict.br

**LAUTARO MATAS**

LA REFERENCIA

lmatas@gmail.com

**VIVIAN SANTOS SILVA**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA

vivian.ss@gmail.com

**WASHINGTON R. DE CARVALHO SEGUNDO**

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA

washingtonsegundo@ibict.br

**JUAN SEBASTIAN MANITTA**

LA REFERENCIA

manittajuan@gmail.com

**EIXO TEMÁTICO**

Infraestructura tecnológica

**RESUMEN**

O Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) em parceria com a LA Referencia desenvolveu uma aplicação para visualização de dados que permitisse aos gestores de repositórios digitais ter acesso às métricas de coleta de dados realizadas. Para o

desenvolvimento da ferramenta utilizou-se de três tecnologias distintas, a primeira é o Software da Plataforma LA Referencia, que tem como função coletar, validar e enriquecer os metadados, já o keycloak é responsável pelo processo de autenticação, gerenciamento de usuários e apresentação das informações relativas aos repositórios e o Angular que utiliza-se das duas tecnologias anteriores para apresentar as métricas geradas por meio do processo de coleta. Este trabalho se concentra no Eixo 4 (Infraestrutura tecnológica). O objetivo é apresentar a interface do gestor, seu desenvolvimento, tecnologias utilizadas e suas funcionalidades. A metodologia utilizada é quali-quantitativa com natureza aplicada, pois tem a finalidade de solucionar de forma prática problemas específicos. Apesar da ferramenta ainda não estar em produção, identificou-se que ela será um importante aliado para auxiliar os gestores de repositórios a ter autonomia na realização da curadoria de metadados fazendo com que a qualidade dos dados coletados aumente de forma a gerar impacto direto na representação, recuperação e disseminação da informação científica.

**PALAVRAS CHAVE**

Visualização de dados; repositórios digitais; interface do gestor.

## Introdução

No contexto dos portais agregadores de informação científica, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) em parceria com a Red de repositorios de acceso abierto a la ciencia (LA Referencia) desenvolveu um software que permitirá aos gestores de repositórios terem acesso ilimitado e irrestrito aos diagnósticos das coletas entre repositórios digitais e portais agregadores.

O desenvolvimento desse software denomina-se “interface do gestor” e ele permitirá que todas as instituições participantes dos 12 nós da rede LA Referencia possam ter acesso ao sistema de diagnóstico de coleta, que antes só era acessível a instituição responsável por cada nó em um país. Por exemplo, no Brasil apenas os responsáveis pela Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) e pelo Portal brasileiro de publicações e dados científicos em acesso aberto (Oasisbr) tinham acesso a tela de análise dos

dados coletados das fontes de informação associadas à BDTD e ao Portal Oasisbr.

Diante disso, e com o objetivo de oferecer maior autonomia aos gestores de repositórios digitais que fazem parte de um dos doze nós[1] vinculados a Rede LA Referencia, desenvolveu-se um dashboard que possibilita a visualização dos dados gerados por meio da coleta dos repositórios. Portanto, os gestores passarão a ter acesso a quantidade de registros coletados, quantos desses registros foram validados, quantos foram invalidados e o porque o registro foi considerado inválido.

Com isso, o objetivo deste trabalho é apresentar as tecnologias utilizadas para o desenvolvimento da ferramenta, bem como apresentar as funcionalidades que serão disponibilizadas nesta primeira versão aos gestores de repositórios.

Para compreender como os dados são aglutinados, transformados, validados e apresentados em uma dashboard, foram utilizadas três ferramentas distintas e com funções específicas dentro da arquitetura da aplicação. Em primeiro plano, tem-se a plataforma de software LA Referencia, que valida cada um dos registros previamente coletados. O processo de validação tem como resultado uma série de métricas e informações que são armazenadas em um núcleo **SOLR**.

Em segunda camada tem-se a interface da dashboard, que é uma aplicação desenvolvida com o framework **Angular** que apresenta diferentes relatórios de validação que são baseados por meio de uma consulta realizada a uma Application Programming Interface (API), desenvolvida especificamente para

[1] Países que integram os 12 nós da Rede LA Referencia: Argentina, Chile, Brasil, Colômbia, Costa Rica, Equador, El Salvador, México, Espanha, Panamá, Peru e Uruguai.

consultar os dados gerados pelo processo de coleta e validação que é realizado pela plataforma LA Referencia.

A API da dashbord implementa funções de autorização para acesso de distintos usuários e/ou distintos grupos de repositórios. Para isso utilizou-se a ferramenta Keycloak, a qual é a terceira camada da aplicação, que é um provedor de identidades que possibilita designar níveis de autorização a cada perfil de usuário que é criado. Com isso um usuário com perfil de administrador pode agregar outros usuários e também ter administração de repositórios distintos. Ademais, a API permite múltiplos pontos de acesso aos relatórios de validação e histórico de coletas de um repositório cadastrado.

## Procedimentos metodológicos

Para a construção metodológica deste trabalho utilizou-se uma abordagem qualitativa de natureza aplicada. Esse método se justifica, pois a abordagem qualitativa foi utilizada na definição e realização de entrevistas guiadas e aplicada pois ao final do estudo tem-se como produto um novo serviço que objetiva atender as necessidades identificadas por meio das entrevistas realizadas.

As entrevistas foram realizadas em 11/05/2020, por meio da plataforma de reunião online Google Meet, com quatro gestoras de repositórios digitais brasileiros que fazem parte tanto da rede da BDTD quanto do Portal Oasisbr. A entrevista foi composta de uma única pergunta que foi “qual a necessidade de informação que elas, enquanto gestoras de repositórios, tinham em relação ao processo de coleta de dados?”

Por meio das respostas obtidas, a equipe pode traçar uma direção para atender às principais demandas apontadas pelas entrevistadas. Como se trata de uma metodologia qualitativa de natureza aplicada, primeiramente iremos apresentar os resultados obtidos por meio da entrevista e em sequência os

passos para o desenvolvimento da dashboard, a qual foi guiada pelas entrevistas.

## Resultados

Por meio de entrevista realizada em 11/05/2020 com 04 gestoras de repositórios, observou-se que os tópicos citados foram:

- Ter acesso aos dados quantitativos de uma coleta: quantos documentos foram coletados, quantos foram validados, quantos foram invalidados, quantos documentos novos são coletados a cada mês.
- Verificar quais as tipologias documentais estão sendo coletadas.
- Exportar relatórios de coletas de maneira personalizada.
- Para atender as necessidades observadas nas entrevistas, a equipe de desenvolvimento se baseou nos tópicos acima para permitir que a dashboard tenha funcionalidades desejadas. As funcionalidades serão melhores descritas no tópico sobre a interface do gestor.

## Software da Plataforma LA Referencia

O software da Plataforma LA Referencia está em sua quarta versão e é desenvolvido em linguagem Java, sob o framework Spring e o banco de dados PostgreSQL. A plataforma é responsável por realizar a coleta dos metadados via protocolo Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) das bibliotecas e repositórios digitais, além de realizar o enriquecimento e a validação dos metadados coletados. A plataforma organiza os dados cadastrais de cada instituição participante da Rede, tais como: nome da instituição, nome do repositório, se a rede é visível ao público, natureza administrativa da instituição, telefone, e-mail, tipo da fonte de coleta, tipos de

documentos a serem coletados, qual o software utilizado, sigla da instituição, Uniform Resource Locator (URL) e a URL URL OAI-PMH.

Pode-se dizer que dentro do formulário de cadastro de coleta de uma instituição a informação mais importante é a URL OAI-PMH, pois é por meio dela que os metadados serão coletados, transformados e validados. Cada uma das três etapas são essenciais para que os registros de uma instituição sejam apresentados no próprio portal da LA Referencia, que atualmente agrega mais 4 milhões de documentos[2] científicos em acesso aberto, bem como para apresentar os registros nos agregadores de cada um dos doze países que fazem uso da tecnologia.

Para unificar diversas fontes de informação distribuídas em diferentes contextos tecnológicos e organizacionais, foi necessário definir um padrão de metadados para que, independente da fonte de informação, os dados estivessem padronizados. Além da definição de um padrão comum de metadados, foi necessário definir quais destes metadados seriam de preenchimento obrigatório. Para isso utilizou-se as diretrizes Open Access Infrastructure for Research in Europe (OpenAIRE) para repositórios de literatura, institucionais e temáticos e também as **diretrizes OpenAIRE** para **arquivos de dados**, a ser utilizados para os casos de Repositórios de dados de pesquisa. De acordo com o próprio guia

As **diretrizes OpenAIRE** fornecem orientação científica para que gestores de repositórios possam definir e implementar suas políticas de gestão de dados local...além do mais os gestores de repositórios não estarão apenas permitindo que autores, que depositam publicações em seu repositório, cumpram com os

[2] Os documentos coletados são artigos, teses, dissertações, livro, capítulo de livro, dados de pesquisa e relatórios.

requisitos de Acesso Aberto da Comissão Europeia, mas também e, eventualmente os requisitos de outros financiadores.

O impacto direto no uso das diretrizes OpenAIRE é que dentro de suas recomendações é definido um conjunto de requisitos para o preenchimento de determinados metadados. Essas recomendações atestam ainda quais metadados são de preenchimento obrigatório, obrigatório se aplicável, recomendado e opcional.

De acordo com essas determinações, a plataforma de software LA Referencia optou por utilizar como regra obrigatória apenas os metadados definidos como de uso obrigatório. É importante explicitar que o preenchimento em branco do metadado não deve ser dado como válido. Portanto, os metadados de preenchimento obrigatório para validação são: título, autor, data de publicação, tipo de documento, direitos de acesso e identificador persistente. É importante destacar que dentro do sistema de coleta e validação dos registros cada um desses metadados é definido como uma regra de validação.

Nos campos tipo de documento e direitos de acesso, o software utiliza o vocabulário controlado da Confederation of Open Access Repositories (COAR) . O vocabulário controlado para tipo de documento aceito no processo de validação é o definido pelo OpenAIRE Guideline versões 3 e 4 e também pelos vocabulários COAR, além de utilizar o vocabulário controlado da COAR para definir a tipologia documental, utiliza-se também para determinar quais são os direitos de acesso ao documento.

Portanto, para que os metadados que representam a informação de tipologia documental e direitos de acesso sejam sancionados no processo de validação a informação preenchida no campo deve estar de acordo com o que determina o vocabulário controlado citado.

No entanto, notou-se que muitas instituições não faziam uso dos vocabulários controlados para o preenchimento das informações, para que isso não causasse um impacto direto na validação dos itens coletados os registros

Até aqui apresentamos como é realizado o processo para coleta e validação dos dados, porém é necessário discorrer sobre o processo de transformação ou enriquecimento dos metadados.

De acordo com Carvalho et al (2021, p.2) “o software LA Referencia também tem uma função de enriquecimento e curadoria dos dados” e esse processo é feito durante a execução da transformação dos dados coletados, ou seja, o processo de transformação nada mais é do que um ato de padronização de informações semelhantes. Exemplo de um processo de transformação ocorre para o campo de tipo de documento. Como mostrado acima, para ser validado o campo deve ser preenchido de acordo com as normas do vocabulário controlado indicado.  Caso isso não ocorra, o registro será invalidado e por seguinte não será apresentado na ferramenta de busca dos portais agregadores. Para dirimir esses problemas, criou-se o processo de transformação, que traduz informações despadronizadas para o padrão definido (ver Figura 1).

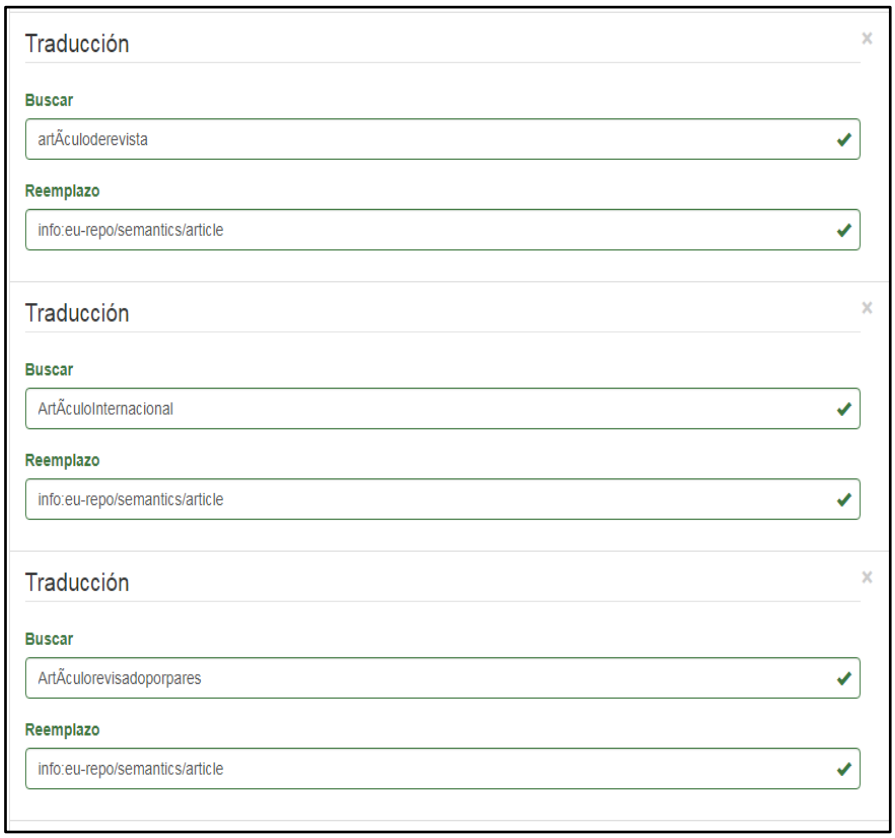


FIGURA 1. Processo de transformação para o campo de tipo de documento

Fonte: PrintScreen de uma das regras de transformação do La Referencia Harvester

A figura 1 apresenta um exemplo de um processo de transformação, o qual funciona da seguinte forma:

- No campo “Buscar” deve-se inserir a informação que a instituição preenche originalmente no campo.
- No campo “Reemplazo” deve-se inserir a informação correspondente no vocabulário controlado.

Realizado os passos acima, o sistema automaticamente substituirá a informação preenchida pela instituição e irá inserir a instituição definida por padrão. Dessa forma, não só o campo de tipologia documental, que os campos têm suas informações padronizadas e tratadas diretamente no software da Plataforma LA Referencia.

Ao final do processo de coleta, transformação e validação dos registros são gerados dados quantitativos referente a cada uma das regras de coleta. No entanto, essas métricas só são geradas se não ocorrer falha em nenhuma das três etapas.

Antes do desenvolvimento da dashboard, somente pessoas com acesso ao software LA Referencia podiam ter acesso  aos dados. No entanto, com o lançamento da interface do gestor, todas as instituições passarão a ter acesso a esses dados. Para tornar isso possível, foi necessário utilizar uma ferramenta que permitisse o gerenciamento de usuários, e para esse propósito, optou-se pelo Keycloak.

## Keycloak

Os usuários devem estar autenticados para ter acesso à Interface do Gestor. Para isso foi utilizado o Keycloak, uma ferramenta de autenticação e autorização que permite a definição de permissões de acesso de forma flexível, no nível de usuário, grupos de usuários, ou papéis (roles), ou ainda através de regras definidas de acordo com as necessidades da aplicação. No contexto da Interface do Gestor, a permissão de acesso a repositórios é baseada em uma regra que determina que um usuário comum (não administrador) deve ser membro de um grupo identificado pelo acrônimo de um repositório para ter acesso às informações deste repositório. Usuários administradores, por outro lado, têm acesso irrestrito a todos os repositórios. O Keycloak se encarrega de verificar as informações do usuário autenticado, aplicar as regras convenientes e decidir se concede ou não o acesso.

A Ferramenta Keycloak possui uma interface web de administração (para cadastro de aplicações, inclusão de usuários, definição de permissões, etc.), mas também permite a administração de forma programática, permitindo que a criação e atualização de usuários, individualmente ou em massa, possa ser feita diretamente dentro da Interface do Gestor da Biblioteca/Repositório

Digital. Para isso, utilizando a API de administração disponibilizada pelo Keycloak, a API REST da plataforma LA Referencia foi estendida para prover também funcionalidades de gestão de usuários. Dentro da Interface do Gestor, um módulo de administração de usuários, visível apenas para usuários com perfil de administrador, permite a criação, edição e remoção de usuários e grupos, e inclusão e exclusão de usuários em grupos (o que irá determinar suas permissões de acesso, ou seja, que repositórios poderá visualizar). Além disso, também é possível carregar informações de uma lista de usuários de um arquivo Comma-separated values (CSV) e sincronizar a lista de grupos com a lista de repositórios na base de dados da plataforma LA Referencia. Desta forma, qualquer administrador na Interface do Gestor pode gerenciar usuários de forma simples e transparente, sem precisar lidar diretamente com o Keycloak.

## A interface do gestor de repositórios

Para o desenvolvimento da interface de visualização dos dados foram utilizados conceitos de Single-Page Application (SPA) com a utilização do framework Angular em sua 15ª versão, dando ênfase em interfaces de usuário interativas, contando com o poder de tipagem do typescript e facilidade de reuso de componentes e acoplamento em módulos. Uma API REST, desenvolvida como um módulo da plataforma LA Referencia com auxílio de Spring Boot, é consumida pela dashboard para recuperação dos dados relativos ao histórico de coletas e de validação dos registros.

Esta API permite a recuperação da lista completa de repositórios coletados, das informações de um repositório específico, da lista de coletas de um determinado repositório, incluindo data e horário de início e fim da coleta e quantidade de registros coletados, e dos resultados da validação para cada coleta, incluindo o número de registros válidos e inválidos, e, para os registros inválidos, quais regras foram violadas.A internacionalização, por meio do

Angular Internationalization (i18n), foi utilizada para a tradução da interface em diferentes idiomas e a formatação de dados para locais específicos.

O desenvolvimento da interface do gestor foi pensado para permitir que os gestores tenham acesso aos processos de coletas e validação dos repositórios. A pertinência do desenvolvimento dessa ferramenta deu-se ao entrevistar alguns gestores para entender qual a necessidade de informação que eles tinham que fossem relacionadas ao processo de coleta. As principais dúvidas estão associadas aos dados quantitativos, principalmente, quando o número de documentos apresentados no repositório era sumariamente maior do que o quantitativo apresentado nos portais agregadores. Essas dúvidas cotidianas dos gestores ocasionaram demandas significativas aos responsáveis pelos nós para analisar o diagnóstico de cada instituição e respondê-las de forma que eles pudessem compreender o problema e trabalhar na resolução, visto que o interesse em coletar e apresentar dados de qualidade é de ambos os lados.

Isto posto, deu-se início ao processo de criação da interface, tendo em mente que os principais usuários dessa ferramenta serão bibliotecários ou cientistas da informação, de modo que os dados deveriam ser expostos de maneira compreensível para que estes profissionais tenham autonomia em entender os dados apresentados na dashboard.

Para que isso fosse possível, foi necessário realizar um trabalho de verificação das informações apresentadas e criar novos textos para que os dados apresentados ficassem claros para o grupo de usuários.

Ao acessar a dashboard a primeira tela que o gestor terá acesso apresentará um o histórico de coletas, ou seja, a cada período é realizada uma coleta completa do repositório, cada uma dessas coletas gera um identificador único. Portanto, a primeira métrica que o usuário da interface do gestor irá visualizar será o histórico de coletas, onde será apresentado o identificador único da coleta, a quantidade de registros coletados, a quantidade de registros válidos e

inválidos. Por meio do histórico de coletas, o gestor poderá verificar quantos documentos foram coletados a cada mês ou a cada coleta. Além disso, será possível exportar esses dados em formato de tabela. Conforme apresenta a figura abaixo.

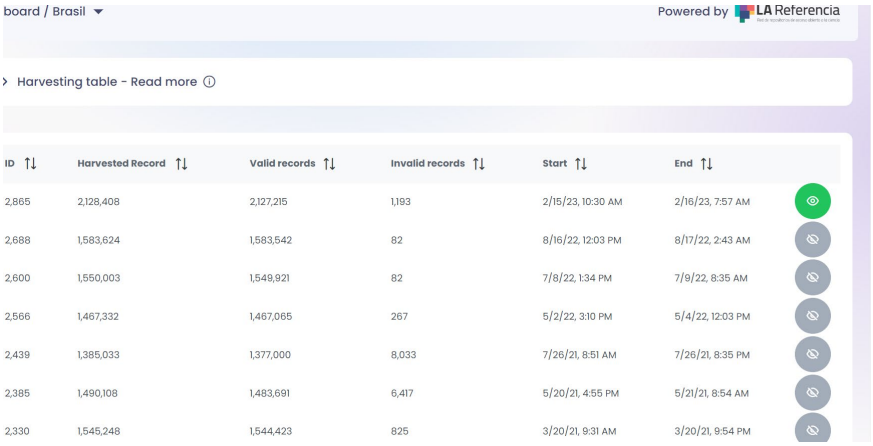


FIGURA 2. Apresentação do histórico de coletas de uma instituição na interface do gestor

Fonte: PrintScreen da aplicação no sistema desenvolvido para a Interface do Gestor

Os dados da figura 2 também poderão ser visualizados em forma de gráfico, onde cada coleta será representada por uma barra com a data e a quantidade de registros coletados, validados e invalidados. Esta visualização está representada na Figura 3.


FIGURA 3. Apresentação do histórico de coletas de uma instituição na interface do gestor

Fonte: PrintScreen da aplicação no sistema desenvolvido para a Interface do Gestor

Também é possível visualizar as métricas relativas aos registros válidos e inválidos separando os dados quantitativos por tipo de regra, ou seja, ao coletar um repositório é possível visualizar quantos dos registros coletados foram validados e invalidados para o campo de autoria, título, tipo de documento, data de publicação, direitos de acesso e identificador persistente. Para isso, os dados são apresentados em uma tabela em que cada linha corresponde a uma regra, conforme apresenta a Figura 3:

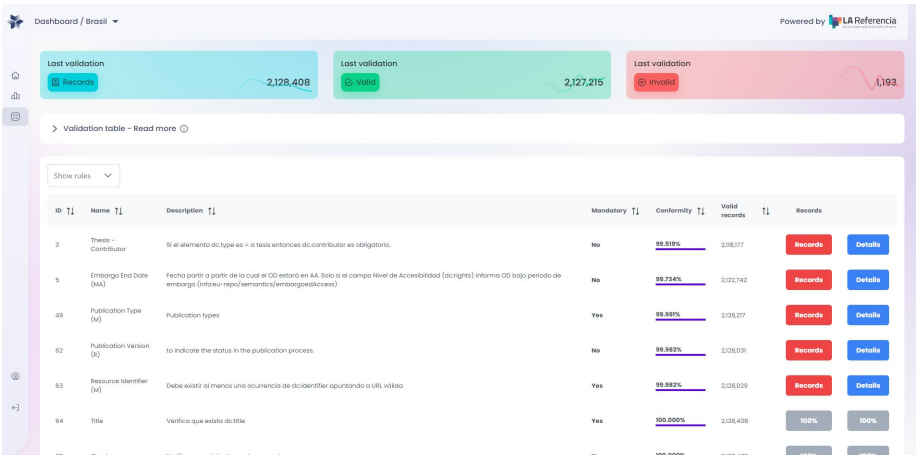


FIGURA 4. Tela de diagnóstico por tipo de regra

Fonte: PrintScreen da aplicação no sistema desenvolvido para a Interface do Gestor

Há um contador de ocorrências de termos para cada regra aplicada, possibilitando uma análise de similaridade de ocorrências. Além disso, é possível ter acesso individual sobre cada registro da coleta, analisar cada regra aplicada, verificar cada ocorrência examinada, e visualizar o XML do registro transformado.

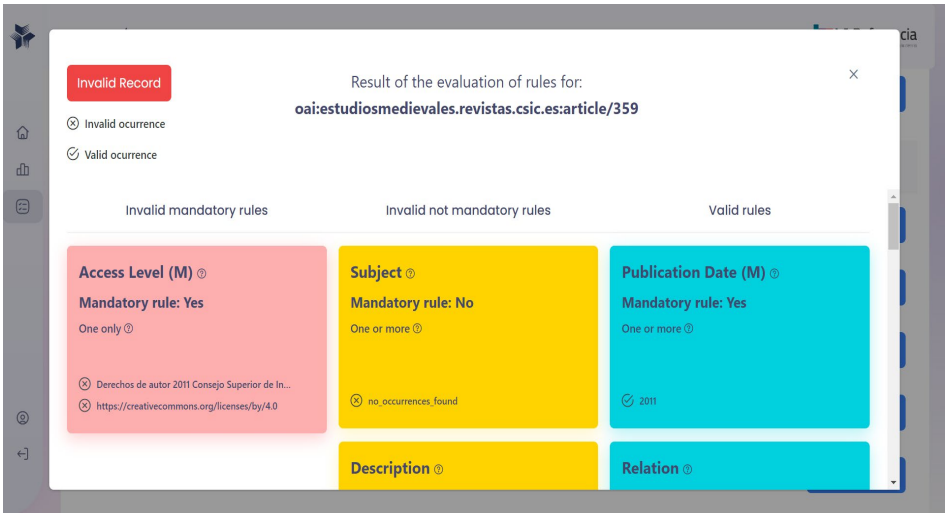


FIGURA 5. Aplicação das regras ao registro e acesso ao XML transformado

Fonte: PrintScreen da aplicação no sistema desenvolvido para a Interface do Gestor[3]

As métricas acima serão disponibilizadas na primeira versão da interface do gestor Apesar de ter informações claras sobre os dados apresentados, a equipe desenvolvedora entende que será necessário treinar os futuros usuários da ferramenta para que o objetivo principal da ferramenta possa ser alcançado.

## Considerações finais

O objetivo principal para o desenvolvimento da interface do gestor se concentrou em dar acesso aos relatórios de coleta de dados aos gestores dos repositórios digitais, no entanto, a tecnologia desenvolvida também auxilia os gestores a realizarem a curadoria dos metadados diretamente na fonte, sem que seja necessário utilizar a funcionalidade de transformação da Plataforma de software LA Referencia.

[3] As imagens e os dados apresentados nas figuras 2, 3, 4 e 5 foram retirados do ambiente de teste e no momento as definições de layout ainda não tinham sido aplicadas.

Esse processo de autonomia na realização da curadoria dos metadados impacta diretamente na qualidade dos dados coletados; na representação e recuperação da informação; na diminuição dos registro invalidados no processo de coleta e, principalmente, contribui de maneira indireta na disseminação, compreensão e uso das boas práticas para o compartilhamento da informação científica na web por meio de portais agregadores.

A interface do gestor já está em etapa de conclusão, tendo os seus módulos de integração já consolidados. O trabalho atual consiste em realizar ajustes no código da dashboard e também refinar os recursos relacionados à experiência do usuário e também ao design da aplicação.

No que se refere à confiabilidade dos dados apresentados na dashboard, durante o ano de 2021 foram realizados diversos testes de visualização dos dados que estavam sendo apresentados, portanto, neste quesito a ferramenta já está estável.

Diante disso, espera-se que ainda em 2023 a interface do gestor seja finalizada e distribuída para a implementação, gerência e uso dos doze nós integrantes da Rede LA Referencia.

## Bibliografía


Carvalho, J., Matas, L., Segundo, W., Graça, P., & Lopes, P. (2021). Dos repositórios aos agregadores, o metamodelo de relações entre entidades: o caso LA Referencia e RCAAP. 11ª Confoa. <http://aleph.letras.up.pt/index.php/paginasaeb/article/view/10244/9511>

Confederation of Open Access Repositories (COAR). (2022). Controlled vocabularies for repositories. <https://vocabularies.coar-repositories.org/access_rights/>

Open Access Infrastructure for Research in Europe (OpenAIRE). (2018). OpenAIRE Guidelines for Literature Repository Managers v4. <https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/>

OPEN ACCESS INFRASTRUCTURE FOR RESEARCH IN EUROPE (OpenAIRE). (2020). Diretrizes OpenAIRE para Repositórios de publicações científicas v4. <https://livroaberto.ibict.br/handle/123456789/1087>

OPEN ACCESS INFRASTRUCTURE FOR RESEARCH IN EUROPE (OpenAIRE). OpenAIRE Guidelines for Data Archives. (2022). <https://guidelines.openaire.eu/en/latest/data/index.html>

SCHMIDT GODOY, A. (1995). Pesquisa qualitativa. *Revista de Administração de Empresas,* *35*(3), 20-29. <https://www.scielo.br/j/rae/a/ZX4cTGrqYfVhr7LvVyDBgdb/?format=pdf&lang=pt>