

## 8. OPENREFINE COMO FERRAMENTA PARA TRATAMENTO DE REGISTROS BIBLIOGRÁFICOS

*Tiago Rodrigo Marçal Murakami*

*Ingrid Torres Schiessl*

*Diego José Macêdo*

*Milton Shintaku*

### 8.1 INTRODUÇÃO

Bibliotecas são cenários férteis para pesquisas científicas, pois são consideradas unidades de informação milenares que atuam de forma transversal e estão presentes nos mais diversos contextos. Nesse sentido, podem ser fonte de pesquisas em várias disciplinas, como a arqueologia com as bibliotecas do mundo antigo, a ciência da computação com o uso de novas tecnologias para atendimento às necessidades informacionais, entre outras. Entretanto, é na biblioteconomia e na ciência da informação que se fomenta a maior parte dos estudos, com pesquisas sobre os catálogos, usuários, formação de coleções, evolução etc.

Nesse ponto, métricas foram criadas para estudar a informação gerenciada pela biblioteca, como a métrica basal *bibliometria* e suas derivações como a *cientometria*, *infometria* e as mais novas *webometria* e *altimetria*. Com as métricas, com o uso forte dos preceitos da estatística, é possível obter indicadores sobre o uso da informação em canais tradicionais como os livros, periódicos e anais de eventos e dos canais mais novos como os *sites* e *mídias sociais*.

A biblioteca como objeto de pesquisa tem ampla possibilidade de estudos, desde o seu papel social na formação dos cidadãos, em si só ampla e complexa, até estudos pontuais como os estudos de caso. Em muito, pela grande possibilidade da tipologia e transversalidade da biblioteca, podendo ser pública ou restrita, educacional em todos os níveis, voltada ao atendimento ao público adulto ou infantil, ser especializada ou de

acervo amplo, ou seja, as múltiplas possibilidades revelam terreno fértil para investigações.

Entre as possibilidades de pesquisas científicas em bibliotecas, repousa sobre o seu catálogo, composto por registros bibliográficos, não apenas obras físicas como também digitais. Hancock-Beaulieu (1990), por exemplo, estudando sobre o comportamento de busca dos usuários, revelava desafios e problemas, principalmente quanto ao assunto. Nesse ponto, são criadas oportunidades de estudos em várias áreas, como para a indexação e criação de vocabulários controlados, criação de novas tecnologias para a recuperação da informação, em letramento informacional para uso de ferramentas informatizadas e outros. Todos esses estudos são baseados no catálogo, na sua formação, representação e uso.

Nesse contexto, as bibliotecas, com seu catálogo formado por registros bibliográficos, podem ser alvos de pesquisas, mas requerem ferramentas que possam atuar com as suas peculiaridades. Grande parte dos catálogos utiliza o padrão *Machine Readable Cataloging* (MARC) para registros das informações. Por isso, exige ferramentas que consigam processar essa tipologia de informação em toda a sua complexidade. Entre as tecnologias existentes está o *OpenRefine*, opção pela simplicidade, flexibilidade e robustez.

## 8.2 REGISTROS BIBLIOGRÁFICOS

Na composição dos catálogos de bibliotecas, os registros bibliográficos desempenham um papel central. Conforme definido por Cunha e Cavalcanti (2008, p. 313), esses registros têm três aspectos fundamentais: primeiro, são armazenados em formatos informatizados e contêm informações bibliográficas que descrevem um ou mais segmentos de registro; segundo, constituem coleções de itens relacionados, tratados como uma unidade e fixados em suporte automatizado; terceiro, nas bases de dados bibliográficos, esses registros substituem ou representam artigos, livros ou outras formas documentais.

A origem da palavra “registros” remonta ao latim “*registrum*”, que significa uma lista ou catálogo de coisas registradas. Com o tempo, essa palavra foi adotada em várias línguas para denotar a ação de registrar ou documentar

informações. Por outro lado, a palavra “bibliográfico” tem suas raízes no grego antigo, com “*biblion*” (livro) e “*grapho*” (escrever), referindo-se, assim, a tudo relacionado à escrita e ao estudo de livros e documentos. Quando combinadas, essas duas palavras formam o termo “Registro Bibliográfico”, que reflete a essência da atividade de documentação e catalogação de materiais, sejam eles impressos ou digitais, em bibliotecas e sistemas de informação.

Compreender o significado dos registros bibliográficos exige uma análise da evolução da catalogação ao longo da história. A criação de registros bibliográficos está intrinsecamente vinculada à prática da catalogação em bibliotecas. Ao longo do tempo, foram desenvolvidas diretrizes e regras para orientar esse processo, como evidenciado em estudos de Ferraz (1991), Fiuza (1987), Machado e Zafalon (2020), Mey (2003) e Selbach *et al.* (2020). Diante do avanço tecnológico, novas regras e diretrizes continuam a ser elaboradas para a catalogação. Isso resultou na evolução dos registros bibliográficos, que se adaptaram às mudanças na forma como as informações são registradas e acessadas nas bibliotecas e sistemas de informação.

Em cenários de migração de sistemas, o tratamento dos registros bibliográficos desempenha um papel crucial. Isso se deve à capacidade de adequação às novas regras de catalogação e aos novos modelos conceituais e recursos tecnológicos que permeiam o ambiente da biblioteca. Isso se alinha com a quinta lei de Ranganathan, que postula que “A biblioteca é um organismo em crescimento” (Ranganathan, 2009, p. 241). De acordo com essa perspectiva, a biblioteca como instituição deve acompanhar as tendências da sociedade em aspectos sociais, educacionais, econômicos, tecnológicos e políticos para evoluir de maneira satisfatória e continuar a atender às necessidades de seus usuários.

### 8.3 SOBRE O OPENREFINE

O *OpenRefine* foi criado pela *Metaweb Technologies, Inc.*, com base no produto originalmente escrito e concebido por David Huynh, voltado para limpeza e transformação de registros. Posteriormente, a *Metaweb Technologies, Inc.* foi adquirida pelo *Google, Inc.* Criado originalmente como *Freebase Gridworks*, em Julho 2010, foi renomeado para *Google Refine*. Em outubro de 2012, foi

novamente renomeado para *OpenRefine*, em sua transição para um produto suportado pela comunidade (OpenRefine, 2023) .

O *OpenRefine* é uma ferramenta de código aberto projetada para ajudar na *limpeza, transformação e reconciliação de dados*, oferecendo *interface amigável* e uma ampla gama de recursos que podem ser especialmente úteis para bibliotecários e profissionais de informação ao lidar com a migração de dados bibliográficos.

Verborgh e DeWilde (2013) relatam que pela quantidade de dados existentes, parte sem muita organização, o *OpenRefine* oferta uma grande gama de funcionalidades voltadas para o tratamento desses dados, como limpeza para corrigir problemas, remover falhas e corrigir despadronização. Com isso, pode ser utilizado para preparar dados para análises automatizadas ou não, principalmente quando se tem uma grande quantidade, como no caso dos registros bibliográficos de uma ou mais bibliotecas.

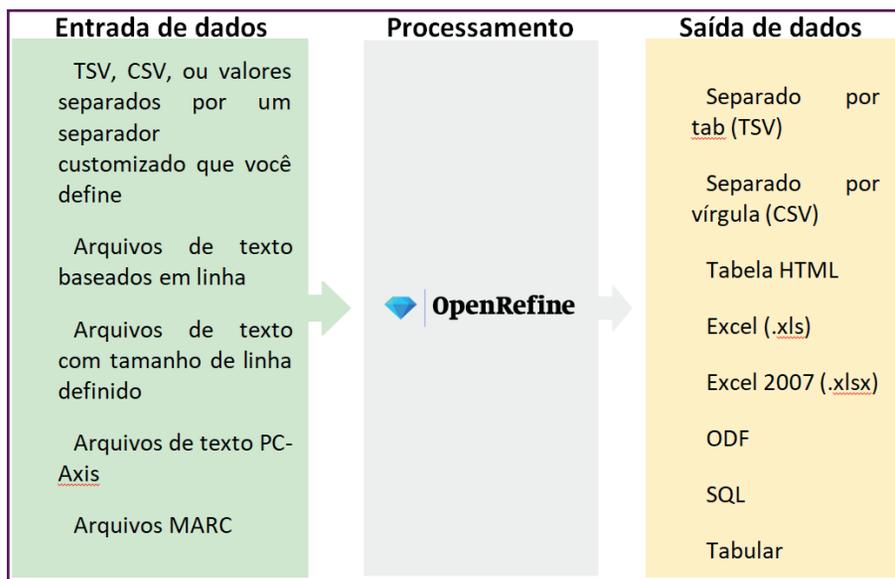
O *OpenRefine* é um *software livre* e pode ser baixado gratuitamente no *site oficial*<sup>94</sup> para ser instalado nos sistemas operacionais mais comuns, como o *Windows, Mac e Linux*. Na sua instalação, o *OpenRefine* possui dependências do *Java Runtime Environment (JRE)*, um ambiente para execução de aplicativos escritos em *Java*. A sua instalação é simples, baseada em pacotes a serem baixados, diferentes para cada tipo de sistema operacional, descompactados e instalados por comando.

Seu funcionamento é semelhante a *páginas web*, com funcionamento simples, em um fluxo de entrada de dados, processamento e saída de dados, como apresentado na Figura 1. Para tanto, pode receber dados em uma grande variedade de formatos, incluindo conexões com gerenciados de banco de dados. Da mesma forma, pode ter os dados extraídos em vários formatos, facilitando o uso posterior em análise ou importação em outras ferramentas. O processamento requer a aplicação de fórmulas que processam os dados.

---

94 Disponível em: <https://openrefine.org/>. Acesso em: 27 set. 2023.

**Figura 1 - Fluxo de entrada e saída de dados no *OpenRefine*.**



Fonte: Elaboração dos autores (2023).

A versatilidade do *OpenRefine* repousa nas grandes possibilidades de entrada e saída de dados e, principalmente, nas possibilidade de aplicação das fórmulas para o tratamento deles. Sendo assim, as fórmulas conseguem processar em todos os dados da linha, escritas por meio de linguagem *General Refine Expression Language* (GREL), em *Jython* (implementação da linguagem *Python* para *Java*) e em *Clojure*.

#### 8.4 EXEMPLO DO USO DO OPENREFINE PARA MANIPULAÇÃO DE REGISTROS BIBLIOGRÁFICOS

A utilização do *OpenRefine* no contexto de processamento de registros bibliográficos pode ser dividida em duas grandes etapas:

- 1. Limpeza de Dados:** Muitas vezes, os registros bibliográficos existentes podem estar sujeitos a inconsistências, erros de digitação e formatos variados. O *OpenRefine* permite a detecção e correção eficaz desses problemas, garantindo consistência para os registros bibliográficos.

**2. Transformação de Dados:** Alguns *softwares* para análise de registros bibliográficos podem os requerer em um determinado padrão. O *OpenRefine* permite que os bibliotecários transformem facilmente os dados para atender a esses requisitos, seja na formatação de datas, padronização de campos ou qualquer outra necessidade específica.

A seguir, apresenta-se algumas das funcionalidades da ferramenta. A Figura 2 apresenta a funcionalidade de geração automática de facetas, tal recurso envolve a criação de categorias ou grupos com base nos valores de uma coluna de dados específica, facilitando o tratamento dos dados.

**Figura 2 - Geração de facetas.**

The screenshot shows the OpenRefine interface with a data table and a facets panel. The facets panel on the left shows a list of facets, with 'tipo\_material' selected and labeled 'Resultado'. The data table on the right has columns for 'Todos', 'registro\_sistema', 'titulo', 'sub\_titulo', 'assunto', 'autor', 'tipo\_material', 'quantidade', 'ano', and 'edicao'. A red arrow points from the 'Facetas' panel to the 'Facetas' column in the table, and another red arrow points to the 'Facetas' column header with the text 'Selecione aqui!'.

Fonte: Captura de tela (2023).

As facetas permitem organizar e analisar dados de maneira mais eficaz, especialmente quando você tem uma grande quantidade de informações não estruturadas. Por exemplo, imagine que você tem uma coluna com nomes de autores de livros em um conjunto de dados. Você pode criar facetas a partir dessa coluna para agrupar os autores por sua primeira letra inicial, facilitando a navegação e a pesquisa. Isso pode ser útil para bibliotecários e pesquisadores, pois ajuda na organização e recuperação de informações.

No *OpenRefine*, você pode criar facetas usando as operações de transformação, como dividir valores em partes ou aplicar expressões regulares para extrair informações específicas de uma coluna. Isso permite que você crie uma estrutura mais organizada para seus dados, o que pode ser valioso na área de Ciência da Informação e em tarefas de tratamento de informações.

Outra funcionalidade é a criação de filtros, conforme apresentado na Figura 3. Os filtros são ferramentas que permitem selecionar e exibir um subconjunto específico de dados com base em critérios definidos. Eles são úteis para analisar, limpar e transformar dados de maneira mais eficiente, permitindo que você foque apenas nos dados relevantes para suas tarefas.

**Figura 3 - Geração de filtros.**

The screenshot shows the OpenRefine interface with a table of 11 rows. The columns are: registro\_sistema, titulo, sub\_titulo, assunto, autor, tipo\_material, quantidade, ano, edicao, and editora. The table contains various entries related to psychology, sociology, and economics. A red arrow points to the 'Selecione aqui' button in the sidebar.

registro_sistema	titulo	sub_titulo	assunto	autor	tipo_material	quantidade	ano	edicao	editora
1	desenvolvimento sem judge /		Psicologia	Adão Cattella Operário	Livro	3	1967		Secretaria Regional do Nordeste
2	Manual de Psicologia /		Psicologia	Atkinson, C. J.	Livro	1	1965		Zahar
3	Statistical Abstract of The United States /		Estatística #EUA	Bureau Of Census	Livro	1	1959		s.n.l.
4	Verdade contra Fraud /		Psicanálise #S&P	Andrade, Almir de	Livro	1	1933		Schmitt
5	Historia de La U.R.S.S. /	época del socialismo (1917-1957) /	União Soviética #S&H	Academia de Ciências de La U.R.S.S.	Livro	1	1958		Editorial Grigobo
7	Estórias e lendas de Goiás e Mato Grosso /		Lendas#S&G	Alcides Jardi	Livro	1	[1943] 2 ed.		EDGRAF
8	Curso de análise matemática /		Análise matemática #S&M	Alcides Jardi	Livro	0	1955	3 ed.	Ed. Científica
9	Ecologia do grupo afro-brasileiro /		Ecologia #S&A	Alves, Rogério	Livro	0	1986		Rio de Janeiro #S&M
10	Habitación, desenvolvimento e urbanização /		Habitación #S&E	Alvares, Carlos	Livro	0	1967		Ministério da Educação e Cultura, Serviço de Documentação, EE O Cruzeiro
11	Tratado de direito /		Direito civil #S&D	Americanos, Jorge	Livro	0	1926		s.n.l.

Fonte: Captura de tela (2023).

Existem diferentes tipos de filtros disponíveis no *OpenRefine*, incluindo:

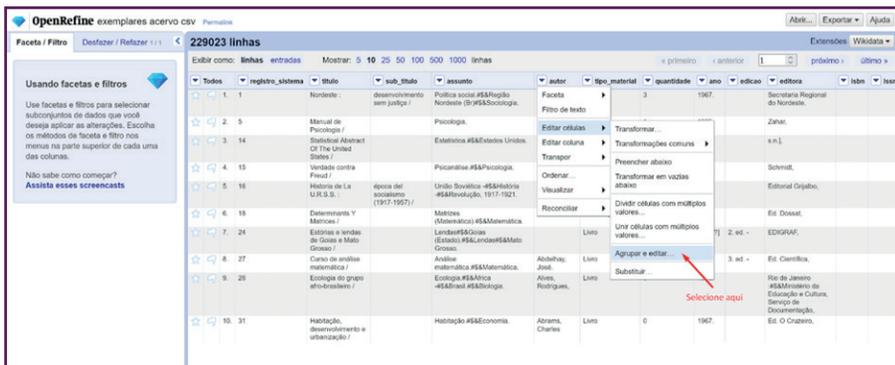
- **Filtro de Texto:** Esse tipo de filtro permite que você pesquise e filtre dados com base em valores de texto em uma coluna. Você pode procurar por palavras-chave ou usar expressões regulares para refinar os resultados.
- **Filtro Numérico:** Usado para filtrar dados em colunas numéricas com base em valores específicos, como intervalos numéricos, valores maiores ou menores que um número dado etc.
- **Filtro de Data:** Esse tipo de filtro é útil quando você lida com datas. Permite que você filtre dados com base em datas específicas, intervalos de datas e muito mais.
- **Filtro de Faceta de Texto:** Como mencionado anteriormente, as facetas são categorias criadas a partir dos valores de uma coluna. Você pode usar filtros de facetas de texto para selecionar rapidamente grupos de dados relacionados a essas categorias.

- **Filtro Personalizado:** Além dos filtros padrão mencionados acima, você também pode criar filtros personalizados usando expressões *GREL* (*General Refine Expression Language*) para aplicar condições específicas aos seus dados.

Os filtros são uma parte essencial do processo de limpeza e transformação de dados no *OpenRefine*. Eles ajudam a focar nos dados relevantes, permitindo que você realize operações específicas em subconjuntos de dados, incluindo edições em lote, o que é útil para bibliotecários e pesquisadores que trabalham com tratamento de informações em Ciência da Informação.

A terceira funcionalidade é chamada de *Clusterização*. Tal termo não é uma palavra da língua portuguesa. Ela é derivada do termo em inglês “*clustering*”, que se refere ao processo de agrupar ou criar *clusters* em análises de dados. No *OpenRefine*, essa funcionalidade permite identificar e agrupar automaticamente valores semelhantes em uma coluna de dados, conforme Figura 4.

**Figura 4 - Funcionalidade de Clusterização.**



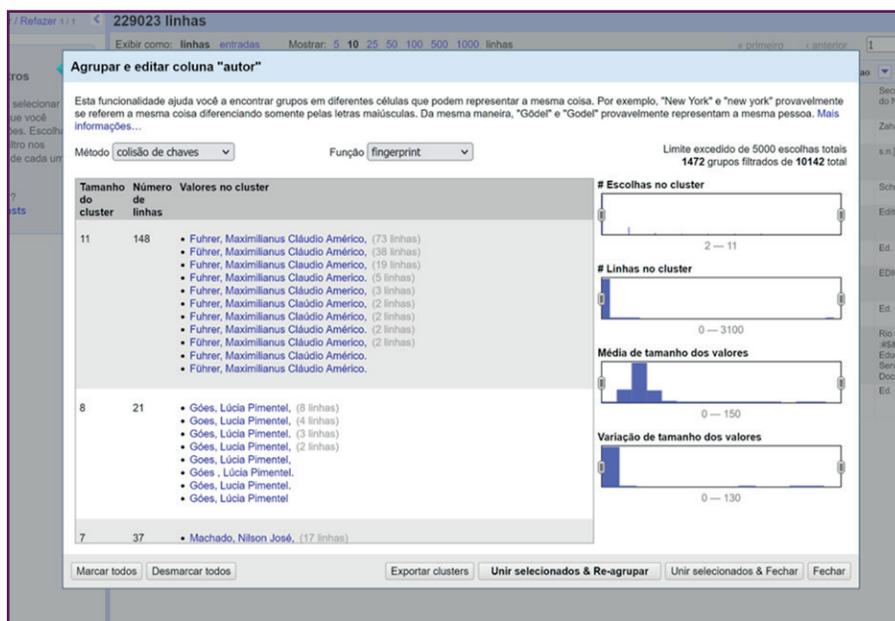
Fonte: Captura de tela (2023).

Essa funcionalidade funciona da seguinte forma:

1. **Identificação de valores semelhantes:** O *OpenRefine* analisa os valores únicos em uma coluna e identifica aqueles que são semelhantes com base em critérios como ortografia, distância de edição e outros métodos de comparação.

2. **Agrupamento:** Os valores semelhantes são agrupados automaticamente em *clusters*. Cada *cluster* representa um grupo de valores que são considerados equivalentes com base nos critérios de semelhança definidos.
3. **Revisão e Fusão:** Após a identificação dos *clusters*, você tem a oportunidade de revisar os agrupamentos propostos e, se necessário, fundir valores de diferentes *clusters* em um único valor corrigido.

**Figura 5 - Funcionalidade de Clusterização aplicada à coluna "autor".**

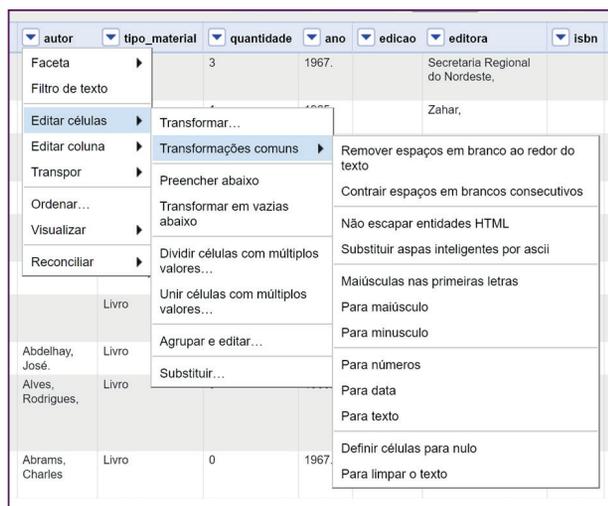


Fonte: Captura de tela (2023).

Essa funcionalidade é especialmente útil para tarefas de limpeza de dados, na qual você pode ter variações nos nomes de autores, por exemplo, devido a erros de digitação ou abreviações. O *clustering* ajuda a consolidar esses valores, tornando o processo de limpeza mais eficiente.

Existem outras funcionalidades que ajudam a edição das células, como Transformações comuns em células, conforme a Figura 6.

**Figura 6 - Função de transformação comum de células.**



Fonte: Captura de tela (2023).

Essa funcionalidade permite aplicar transformações simples e comuns aos dados de uma célula. Tais transformações são frequentemente usadas para limpar, padronizar ou formatar os dados de maneira consistente, por exemplo, as entradas de autoria ou assuntos. Alguns exemplos de transformações comuns que podem ser aplicadas usando essa função incluem:

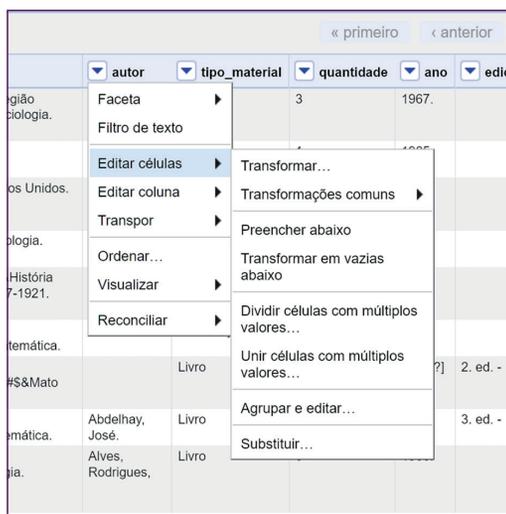
- Para maiúsculo: Converte todo o texto em maiúsculas.
- Para minúsculo: Converte todo o texto em minúsculas.
- Maiúsculas nas primeiras letras: Torna a primeira letra de cada palavra maiúscula e as demais em minúsculas.
- Remover Espaços em Branco: Remove espaços em branco extras no início ou no final de um valor.
- Substituir Texto: Substitui um valor específico por outro em toda a coluna.
- Editar Célula: Permite editar o conteúdo de uma célula individualmente.

- Dividir células com múltiplos valores: Divide uma coluna em várias colunas com base em um separador, como vírgulas.
- Unir células com múltiplos valores: Mescla os valores de várias colunas em uma única coluna.
- Para data: Formata datas de acordo com um padrão específico.
- Para números: Converte valores de texto em números.
- Para células em branco: Preenche toda a seleção para valores em branco.

A função “Transformações comuns” é uma ferramenta poderosa para a limpeza e a preparação de dados no *OpenRefine*, tornando mais fácil garantir que os dados estejam em um formato adequado para análises subsequentes. Ela oferece uma série de opções para realizar essas transformações de maneira eficiente e consistente.

Além da função mencionada anteriormente, também há outras edições possíveis nas células, como apresentado na Figura 7.

**Figura 7 - Função de edição das células.**



Fonte: Captura de tela (2023).

Essas funcionalidade são:

- Preencher abaixo: Essa funcionalidade permite preencher valores em branco em uma coluna com base no valor da célula acima. Isso é útil para preencher lacunas em uma coluna com valores repetidos ou sequenciais.
- Transformar em vazias abaixo: Com essa função, você pode limpar o conteúdo das células abaixo de uma célula selecionada. Normalmente utilizado para excluir dados repetidos em células, a fim de facilitar o tratamento em outras colunas.
- Dividir células com múltiplos valores: Essa funcionalidade é útil quando você tem células que contêm múltiplos valores separados por um delimitador, como vírgulas. Ela permite dividir esses valores em células individuais ou em várias colunas diferentes.
- Unir células com múltiplos valores: Ao contrário da divisão, essa funcionalidade permite unir células com múltiplos valores em uma única célula, separando os valores por um delimitador específico. Isso é útil quando você deseja consolidar informações em uma única célula.
- Substituir: Com a função "Substituir", você pode encontrar e substituir valores específicos em uma coluna. Isso é útil para correções em massa ou para padronizar valores em todo o conjunto de dados.

Essas funcionalidades são valiosas para a manipulação e preparação de dados no *OpenRefine*, ajudando a garantir que seus dados estejam em um formato adequado para análises posteriores ou para atender a requisitos específicos.

Além das células, é possível realizar edição nas colunas, conforme Figura 8. É uma funcionalidade que permite criar edições personalizadas nos valores de uma coluna de dados. Ela é útil quando é preciso realizar ajustes específicos nos valores da coluna que não podem ser realizados com as Transformações Comuns.

**Figura 8 - Função de edição de colunas.**

« primeiro < anterior 1						
	▼ autor	▼ tipo_material	▼ quantidade	▼ ano	▼ edicao	▼ ed
a.			3	1967.		Secreta do Nor
			1	1965.		Zahar,
idos.						s.n.],
						Schmid
ia l.						Editoria
ca.						Ed. Do
ato		Livro				EDIGR
a.	Abdelhay, José.	Livro				Ed. Cie
	Alves, Rodrigues,	Livro				Rio de :#\$&Mi Educaç Serviço Docum
	Abrams, Charles	Livro	0	1967.		Ed. O C

Fonte: Captura de tela (2023).

Essas funcionalidades são as seguintes:

- Dividir em diversas colunas: Essa funcionalidade permite dividir os valores de uma coluna em várias colunas com base em um separador específico. É útil quando você tem valores que estão concatenados e deseja organizá-los em colunas separadas.

- Mesclar colunas: Com essa funcionalidade, você pode mesclar os valores de duas ou mais colunas em uma única coluna. É útil para combinar informações de várias fontes em uma única coluna.
- Adicionar coluna baseada nesta coluna: Essa funcionalidade permite criar uma nova coluna com base nos valores de uma coluna existente. Você pode aplicar expressões *GREL* para calcular ou transformar os valores da nova coluna com base nos valores da coluna original.
- Adicionar coluna através de *URLs*: Com essa funcionalidade, você pode criar uma nova coluna recuperando dados de uma *URL* externa. Isso é útil quando você deseja enriquecer seus dados com informações obtidas da *web*, como informações de geolocalização, cotações de moeda etc.
- Renomear esta coluna: Essa funcionalidade permite renomear uma coluna, dando-lhe um novo nome que seja mais descritivo ou adequado ao seu projeto.
- Remover esta coluna: Com essa funcionalidade, você pode remover uma coluna inteira do conjunto de dados. Isso é útil quando você tem colunas desnecessárias ou duplicadas.
- Mover: A funcionalidade “Mover” permite reorganizar a ordem das colunas no conjunto de dados. Você pode arrastar e soltar as colunas para posicioná-las onde desejar, facilitando a visualização e análise dos dados.

Essas funcionalidades são utilizadas para a limpeza, transformação e preparação de dados. Permitindo a edição dos valores de forma eficiente de acordo com as necessidades.

Outra ferramenta que auxilia na organização dos dados é a “Transpor”, que permite reorganizar os dados de uma tabela trocando as linhas por colunas e vice-versa. Isso é especialmente útil quando você deseja *pivotar* ou girar a estrutura dos seus dados para uma melhor visualização ou análise. São elas:

- Transpor células de linhas para colunas: A funcionalidade permite trocar as linhas da tabela pelas colunas e vice-versa. Isso é útil quando seus

dados estão dispostos horizontalmente, mas você precisa deles na vertical ou vice-versa.

- Transpor células de colunas para linhas: Você pode escolher quais colunas deseja transpor. Isso é importante quando você deseja manter algumas colunas inalteradas e transpor apenas um subconjunto delas.
- Você pode especificar o nome da nova coluna que será criada para armazenar os valores transpostos. Isso permite dar um nome descritivo e significativo à nova estrutura de dados.

As funcionalidades mencionadas anteriormente são aplicadas de forma combinada. Durante o processo de migração, essa ferramenta é utilizada quando os dados de origem são transformados em formato tabular. Essa tabela é então formatada de acordo com as normas *MARC* e exportada em um formato tabular para a posterior migração para o formato *MARC*, utilizando uma ferramenta externa, como o *Librecat/Catmandu*<sup>95</sup> ou o *MARCEdit*<sup>96</sup>.

Exemplos de operações realizadas com o *OpenRefine* durante o processo de migração incluem:

Conversão de codificação: O formato *MARC* requer que os dados sejam codificados em campos específicos e siga uma codificação única. Por exemplo, o país de publicação deve ser codificado como "bl" para Brasil no *MARC*. No entanto, em sistemas legados, os dados podem variar, incluindo "Brasil", "Brazil" ou até mesmo conter erros de digitação. O *OpenRefine* oferece uma solução para corrigir esses dados, permitindo a separação dos dados em uma única coluna, o uso de facetas para criar conjuntos de dados e, em seguida, a alteração dos dados para os novos códigos. Essa funcionalidade permite a correção de todas as ocorrências na coluna de forma eficiente, algo que não seria facilmente executado em outras ferramentas.

Separação de campos: Em sistemas anteriores, informações como título e subtítulo podem estar contidas no mesmo campo, separadas por um

---

95 Disponível em: <https://librecat.org/>. Acesso em: 27 set. 2023.

96 Disponível em: <https://marcedit.reeset.net/>. Acesso em: 27 set. 2023.

caractere como ":" (dois pontos). No entanto, o formato *MARC* requer campos específicos para o título (\$245a) e subtítulo (\$245b). O *OpenRefine* oferece a funcionalidade de dividir colunas por separadores, o que possibilita a criação de duas colunas distintas para o título e o subtítulo, seguindo as exigências do formato *MARC*. Isso facilita a organização e a migração dos dados de maneira precisa.

## 8.5 CONSIDERAÇÕES FINAIS

O *OpenRefine* é uma ferramenta valiosa para o tratamento de dados bibliográficos, principalmente, em situações de migração de sistemas de informação. Sua capacidade de limpar, transformar e reconciliar dados de forma eficiente ajuda as bibliotecas a garantirem a qualidade e consistência de seus registros. Além disso, a flexibilidade do *OpenRefine* permite que as instituições personalizem o processo de migração de acordo com suas necessidades específicas, tornando-o uma escolha eficaz para bibliotecas de diversos tamanhos e contextos.

A integração bem-sucedida do *OpenRefine* no fluxo de trabalho para tratamento de registros bibliográficos pode resultar em economia de tempo e recursos, uma vez que automatiza tarefas repetitivas e facilita a correção de erros nos registros. No entanto, é importante que as equipes envolvidas recebam o treinamento adequado para aproveitar ao máximo essa ferramenta. Em resumo, o *OpenRefine* desempenha um papel significativo na simplificação do processo de migração de registros bibliográficos para sistemas de informação, contribuindo para uma gestão mais eficiente da informação nas bibliotecas.

Por fim, o *OpenRefine* se apresenta como mais uma ferramenta que pode ser útil na etapa de tratamento de dados em pesquisas científicas que tenham como dados os registros bibliográficos ou similares. Com isso é possível juntar dados provenientes de várias fontes, assim como transformar para padrões desejados para facilitar as análises. Em alguns casos de pesquisas científicas, nem toda a coleta de dados obtém os registros no formato desejado, requerendo processamento para passar para a próxima etapa, e o *OpenRefine* pode atender a essa demanda.

## REFERÊNCIA

CUNHA, Murilo Bastos da; CAVALCANTI, Cordélia Robalinho de Oliveira. **Dicionário de Biblioteconomia e Arquivologia**. Brasília, DF: Briquet de Lemos, 2008. 451 p. Disponível em: <https://repositorio.unb.br/handle/10482/34113>. Acesso em: 21 set. 2023.

FERRAZ, Iraneuda Maria Cardinali. Uso do catálogo de biblioteca: uma abordagem histórica. **TransInformação**, Campinas, v. 3, n. 1/3, p. 90-114, jan./dez. 1991. Disponível em: <https://brapci.inf.br/index.php/res/download/55433>. Acesso em: 1 mar. 2023.

FIUZA, Marysia Malheiros. A catalogação bibliográfica até o advento das novas tecnologias. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 16, n. 1, p. 45-53, mar. 1987.

HANCOCK-BEAULIEU, Micheline. Evaluating the impact of an online library catalogue on subject searching behavior at the catalog and the shelves. **Journal of Documentation**, Bingley, v. 46, n. 4, p. 318-338, Apr. 1990. DOI 10.1108/eb026863. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/eb026863/full/html>. Acesso em: 21 set. 2023.

MACHADO, Raildo; ZAFALON, Zaira. **Catalogação: dos princípios e teorias ao RDA e IFLA LRM**. João Pessoa, PB: Editora UFPB, 2020. 128 p. Disponível em: <http://eprints.rclis.org/43200/>. Acesso em: 21 set. 2023.

MEY, Eliane Serrão Alves. **Não brigue com a catalogação**. Brasília, DF: Briquet de Lemos, 2003. 186 p.

OPENREFINE. [2023, *online*]. Disponível em: <https://openrefine.org/>. Acesso em: 1 mar. 2023.

RANGANATHAN, Shiyali Ramamrita. **As Cinco Leis da Biblioteconomia**. Brasília, DF: Briquet de Lemos, 2009. 336 p.

SELBACH, Clarissa Jesinska; FERREIRA, Anamaria; KERN, Lucas Martins; NOVAK, Loiva Duarte. Catalogação com Resource Description and Access (RDA): relato de experiência na Biblioteca Central Irmão José

Otão (PUCRS). **Revista ACB**: Biblioteconomia em Santa Catarina, Florianópolis, v. 25, n. 3, p. 729-733, 2020. Disponível em: <https://brapci.inf.br/index.php/res/v/151853>. Acesso em: 1 mar. 2023.

VERBORGH, Ruben; DE WILDE, Max. **Using OpenRefine**. Birmingham, UK: Packt Publishing, 2013.

## DADOS DOS AUTORES:

### Tiago Rodrigo Marçal Murakami



Tiago Rodrigo Marçal Murakami é Bibliotecário na Escola de Comunicações e Artes da Universidade de São Paulo. Graduado em Biblioteconomia pela Universidade de São Paulo.

<https://orcid.org/0000-0003-1942-6434>  
trmurakami@usp.br

### Ingrid Torres Schiessl



Ingrid Torres Schiessl é Mestre em Ciência da Informação e bacharela em Biblioteconomia pela Universidade de Brasília (UnB). Bibliotecária e pesquisadora no Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict).

<https://orcid.org/0000-0001-5815-2574>

[ingridschiessl@ibict.br](mailto:ingridschiessl@ibict.br)

### Diego José Macêdo



Diego José Macêdo é Mestre em Ciência da Informação pela Universidade de Brasília. Bacharel em Sistema de Informação pela Universidade Católica de Brasília. Atualmente é tecnologista do Instituto Brasileiro de Informações em Ciência e Tecnologia - Ibict.

[diegomacedo@ibict.br](mailto:diegomacedo@ibict.br)

<https://orcid.org/0000-0002-5696-0639>

## Milton Shintaku



Milton Shintaku é Doutor em Ciência da Informação pela Universidade de Brasília. Coordenador de Tecnologia para Informação (Cotec) do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict).

shintaku@ibict.br

<https://orcid.org/0000-0002-6476-4953>

### Como referenciar o capítulo 8:

MURAKAMI, Tiago Rodrigo Marçal; SCHIESSL, Ingrid Torres; MACÊDO, Diego José; SHINTAKU, Milton. OpenRefine como ferramenta para tratamento de registros bibliográficos. *In*: SHINTAKU, Milton; MACÊDO, Diego José; MARIN, Luciano Heitor Gallegos (org.). **Tecnologias utilizadas em pesquisas acadêmicas em Ciências Sociais Aplicadas**. Brasília, DF: Ibict, 2023. cap. 8. p. 225-244. ISBN 978-65-89167-94-5. DOI: <http://doi.org/10.22477/9786589167938cap8>.