

# Produção de indicadores de empregabilidade com base em técnicas de mineração de Big Data e Business Intelligence

## **Daniel Alves da Silva**

Doutorando em Engenharia Elétrica pela Universidade de Brasília (UnB) - Brasília, DF - Brasil. Mestre em Engenharia Elétrica pela Universidade de Brasília (UnB) - DF - Brasil. Pesquisador do Laboratório de Tecnologia da Tomada de Decisão LATITUDE, vinculado a Faculdade de Tecnologia da Universidade de Brasília (UnB) - DF - Brasil.

<http://lattes.cnpq.br/9324394974130734>

E-mail: [daniel.alves@redes.unb.br](mailto:daniel.alves@redes.unb.br)

## **Paulo Lima Machado**

Mestre em Engenharia Elétrica pela Universidade de Brasília (UnB) - Brasília, DF - Brasil. Pesquisador do Laboratório de Tecnologia da Tomada de Decisão LATITUDE, vinculado a Faculdade de Tecnologia da Universidade de Brasília (UnB) - DF - Brasil.

Professor do Instituto de Educação Superior do Brasília (IESB) - Brasília, DF - Brasil.

<http://lattes.cnpq.br/0551306329993153>

E-mail: [paulo.machado@redes.unb.br](mailto:paulo.machado@redes.unb.br)

## **Vinícius Coutinho Guimarães Coelho**

Mestre em Engenharia Elétrica pela Universidade de Brasília (UnB) - Brasília, DF - Brasil. Pesquisador do Laboratório de Tecnologia da Tomada de Decisão LATITUDE, vinculado a Faculdade de Tecnologia da Universidade de Brasília (UnB) - DF - Brasil.

<http://lattes.cnpq.br/7897174695186091>

E-mail: [vinicius.coelho@redes.unb.br](mailto:vinicius.coelho@redes.unb.br)

## **Renata Visoná Barbosa**

Graduanda em Engenharia de Redes de Comunicação pela Universidade de Brasília (UnB) - Brasília, DF - Brasil.

Pesquisadora do Laboratório de Tecnologia da Tomada de Decisão LATITUDE, vinculado a Faculdade de Tecnologia da Universidade de Brasília (UnB) - DF - Brasil.

<http://lattes.cnpq.br/6125451246526753>

E-mail: [renata.barbosa@redes.unb.br](mailto:renata.barbosa@redes.unb.br)

## **Fábio Lúcio Lopes de Mendonça**

Doutorando em Engenharia Elétrica, Redes e Telecomunicações pela Universidade de Brasília (UnB) - Brasília, DF - Brasil. Mestre em Engenharia Elétrica pela Universidade de Brasília (UnB) - DF - Brasil. Pesquisador do Laboratório de Tecnologia da Tomada de Decisão LATITUDE, vinculado a Faculdade de Tecnologia da Universidade de Brasília (UnB) - DF - Brasil.

<http://lattes.cnpq.br/6801714487611163>

E-mail: [fabio.mendonca@redes.unb.br](mailto:fabio.mendonca@redes.unb.br)

## **Dário Pereira dos Santos**

Mestrando em Sistemas Mecatrônicos pela Universidade de Brasília (UnB) - Brasília, DF - Brasil. Graduado em Sistemas de Informação pela Faculdade Anhanguera - Brasília, DF - Brasil. Pesquisador do Laboratório de Tecnologia da Tomada de Decisão LATITUDE, vinculado a Faculdade de Tecnologia da Universidade de Brasília (UnB) - DF - Brasil.

<http://lattes.cnpq.br/9158179035727721>

E-mail: [dario.santos@redes.unb.br](mailto:dario.santos@redes.unb.br)

## **Rafael Timóteo de Sousa Júnior**

Pós-Doutorado pela Ecole Supérieure D'électricité Supelec (SUPELEC) - França. Doutor em Processamento de Sinais e Telecomunicações pela Université de Rennes I Rennes (RENNES I) - França. Professor Pesquisador da Universidade de Brasília (UnB) - Brasília, DF - Brasil. Coordenador do Programa de Pós-graduação profissional em Engenharia Elétrica (PPEE) e do Laboratório de Tecnologia da Tomada de Decisão LATITUDE, vinculado a Faculdade de Tecnologia da Universidade de Brasília (UnB) - DF - Brasil.

<http://lattes.cnpq.br/3196088341529197>

E-mail: [desousa@unb.br](mailto:desousa@unb.br)

## RESUMO

O atual cenário da empregabilidade no Brasil é frequentemente evidenciado na mídia nacional, devido a sua relevância como indicador de desenvolvimento econômico e social. Nesse contexto, diversas ações governamentais têm sido propostas a fim de viabilizar a redução do índice de desemprego e a otimização do mercado de trabalho. A utilização de tecnologias para análise de grandes massas de dados é uma realidade inegável e atualmente é aplicada a quase todas as atividades. Assim, a disponibilização de dados para a formulação de cenários que possibilitem a geração de informações, visando auxiliar a tomada de decisões, mostra-se como uma ferramenta de grande valia. O principal objetivo deste trabalho é apresentar como o cruzamento das bases de dados que caracterizam o mercado de trabalho pode gerar informações pertinentes para formação de políticas públicas e sociais. Trata-se de uma pesquisa aplicada, de caráter experimental e exploratório, utilizando a prototipação de um sistema para validação do cenário. As bases de dados utilizadas para o cruzamento de informações são compostas pela Relação Anual de Informações Sociais (Rais) e pelo Cadastro Geral de Empregados e Desempregados (Caged). Como resultado, observou-se que os indicadores e o mapeamento do mercado de trabalho possibilitam o entendimento e identificação das atividades e ocupações laborais, conforme métricas específicas para o planejamento e implementação de políticas públicas e sociais. Foi possível concluir que as técnicas utilizadas são eficientes para a geração de informações acessíveis aos gestores, empresários e cidadãos, viabilizando a formação de políticas com maior efetividade.

**Palavras-chave:** Empregabilidade. Políticas públicas. Tecnologia. Business intelligence.

## ***Production of employability indicators based on Big Data and Business Intelligence mining techniques***

### **ABSTRACT**

*The current scenario of employability in Brazil is often evidenced in the national media, due to its relevance as an indicator of economic and social development. In this context, several governmental actions have been proposed in order to enable the reduction of the unemployment rate and the optimization of the labor market. The use of technologies for analysis of large masses of data is an undeniable reality and is currently applied to almost all activities. Thus, the provision of data for the formulation of scenarios that enable the generation of information, aiming to assist decision making, is shown as a valuable tool. The main objective of this work is to present how the cross-referencing of databases that characterize the labor market can generate pertinent information for the formation of public and social policies. It is an applied research, of experimental character, using the prototyping of a system for validation of the scenario. The databases used for the cross-referencing of information are composed of the Annual Report of Social Information (Rais) and the General Register of Employees and Unemployed Persons (Caged). As a result, it was observed that indicators and labor market mapping make it possible to understand and identify labor activities and occupations, according to specific metrics for the planning and implementation of public and social policies. It was possible to conclude that the techniques used are efficient for the generation of information accessible to managers, entrepreneurs and citizens, making feasible the formation of policies with greater effectiveness.*

**Keywords:** *Employability. Public policy. Technology. Business intelligence.*

## ***Producción de indicadores de empleabilidad basados en técnicas de minería de Big Data y Business Intelligence***

### **RESUMEN**

El actual escenario de la empleabilidad en Brasil es evidenciado en los medios nacionales como indicador de desarrollo económico y social. En este contexto, diversas acciones del gobierno han sido propuestas para hacer viable la reducción del índice de desempleo y la optimización del mercado de trabajo. La utilización de tecnologías para análisis de grandes masas de datos es una realidad y actualmente se aplica a casi todas las actividades. La disponibilidad de datos para la formulación de escenarios que posibiliten la generación de informaciones, con el fin de ayudar a la toma de decisiones, se muestra como una herramienta de gran valor. El objetivo de este trabajo es presentar cómo el cruce de la base de datos, que caracterizan el mercado de trabajo, puede generar informaciones pertinentes para la formación de políticas públicas y sociales. Se trata de una investigación aplicada, de carácter experimental, utilizando el prototipado de un sistema para la validación del escenario. Bases de datos utilizadas para el cruce de informaciones están compuestas por la Relación Anual de Informaciones Sociales (Rais) y por el Catastro General de Empleados y Desempleados (Caged). Como resultado, se observó que los indicadores y el mapeo del mercado de trabajo posibilita el entendimiento e identificación de las actividades y ocupaciones laborales, como métricas específicas para la planificación e implementación de políticas públicas y sociales. Es posible concluir que las técnicas utilizadas son eficientes para la generación de informaciones accesibles a los gestores, empresarios y ciudadanos, permitiendo la formación de políticas con mayor efectividad.

**Palabras clave:** Empleabilidad. Políticas públicas. Tecnología. Business intelligence.

### **INTRODUÇÃO**

A atual crise de empregabilidade impõe a necessidade de ações integradas na busca para geração de empregos e renda. As alterações nas relações trabalhistas, resultado da recente reforma trabalhista<sup>1</sup>, reafirmam a necessidade de programas voltados para a melhoria do mercado de trabalho em âmbito nacional. Diante destes novos desafios, surgem iniciativas como as parcerias entre os órgãos do governo federal e universidades públicas, uma vez que elas são, por essência, polos de pesquisa e desenvolvimento científico e tecnológico.

Nesse sentido, destaca-se o projeto Pesquisa de desenvolvimento metodológico e capacitação para governança e gestão pública da empregabilidade no Brasil, firmado entre o Ministério do Trabalho e a Universidade de Brasília (UnB)<sup>2</sup>, denominado “Escola do Trabalhador”.

<sup>1</sup> LEI Nº 13.467, DE 13 DE JULHO DE 2017. Disponível em [http://www.planalto.gov.br/ccivil\\_03/\\_Ato2015-2018/2017/Lei/L13467.htm](http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2017/Lei/L13467.htm) - Acesso em 05/01/2019.

<sup>2</sup> TED NR 001/2017 MTb/SPPE - FUB/UnB.

Esse projeto visa à oferta de cursos de qualificação profissional, em âmbito nacional, por meio do uso de tecnologias da informação (TICs).

Nesse projeto, com o objetivo de viabilizar a integração de dados e a sistematização de informações sobre trabalho e emprego, foram desenvolvidas atividades para a prototipação de uma plataforma de inteligência organizacional com indicadores pertinentes à capacitação dos trabalhadores brasileiros visando à empregabilidade, correlacionando fontes de dados governamentais, a fim de gerar painéis e relatórios de apoio à tomada de decisão e formação de políticas públicas.

Foi desenvolvido um protótipo de software que possibilita a análise de cenários específicos para formulação de políticas públicas voltadas ao mercado de trabalho. Este protótipo é utilizado para publicação de dados gerais sobre o mercado de trabalho através de painéis na Internet com informações gráficas e numéricas (*dashboards*), conforme listados a seguir e detalhados na seção Elaboração dos Painéis.

Cada painel está disposto em uma página da Internet específica e com acesso facilitado através de um portal centralizador de informações onde é descrito cada um deles.

- **Perfil do trabalhador:** informações acerca da quantidade de vínculos ativos e remuneração média por região, gênero, faixa etária, profissão, etc.
- **Caracterização da mão de obra:** informações mensais sobre as características da mão de obra no mercado de trabalho formal brasileiro.
- **Mapa de demanda por qualificação profissional:** extrato sobre a empregabilidade formal através de visualização georreferenciada.

No presente artigo, além desta seção de Introdução, são apresentadas a metodologia e os conceitos utilizados para a formulação dos painéis e indicadores formulados, informações sobre as bases de dados que fundamentaram a pesquisa, além da apresentação dos painéis com suas funcionalidades e as conclusões sobre as possíveis aplicações e utilização das informações tratadas.

## METODOLOGIA

*Business Intelligence* (BI) é um termo que engloba um conjunto de ferramentas de software, com metodologia, arquitetura, bases de dados, *data warehouse*<sup>3</sup>, etc., que apoiam a tomada de decisão. Tem como objetivo possibilitar aos analistas, gerentes de negócio e altos executivos acessar quaisquer dados da organização de maneira fácil e rápida, de preferência em tempo real, bem como formular análises e manipulações apropriadas (TURBAN, E. *et al*, 2009).

O BI possui capacidades resultantes da integração de suas metodologias e tecnologias que permitem a geração de relatórios, análises complexas, mineração de dados, previsões, entre outros tipos de apresentação.

Atualmente as plataformas de BI geralmente utilizam a Internet como interface gráfica, disponibilizando as consultas através dos mais diversos dispositivos como computadores, tablets e celulares.

Utilizando como base o paradigma da informação, o ecossistema de Big Data pode ser definido como “o estudo do ambiente de facilidades técnicas para armazenamento, processamento e entrega de dados, pressuposto para a gestão da informação e do conhecimento nas organizações” (ARAÚJO JÚNIOR, R.H. de; SOUSA, R.T.B. de, 2016).

O entendimento do mercado de trabalho através dos dados pode ser utilizado como apoio a políticas públicas, conforme registrado por Frey e Osborne (FREY; OSBORNE, 2017), que utilizaram dados sobre o mercado de trabalho para análise e entendimento do impacto das ocupações que estão suscetíveis à informatização. Os autores desenvolveram, através de dados sobre o mercado de trabalho, uma metodologia para estimar a probabilidade de informatização de cerca de 700 ocupações e como esse fato poderá impactar no mercado de trabalho nos Estados Unidos.

As bases governamentais brasileiras geralmente são disponibilizadas através de arquivos com textos tabulados, sendo necessário o tratamento dos dados até seu formato final através da adoção de ferramentas específicas (ARAÚJO, L.R; SOUZA, J.F. de, 2011). Ante a natureza dos dados e do processamento demandado, as ferramentas que mais se adequam ao projeto são contempladas pelo BI. Ou seja, adotou-se a técnica de Extração, Transformação e Carga, do inglês Extract, Transformation and Load (ETL), para tratamento dos arquivos de carga e uma ferramenta de Online Analytical Processing (OLAP) para apresentação dos dados através dos painéis (VASSILIADIS; SIMITSIS; SKIADOPOULOS, 2002).

Existe atualmente uma grande quantidade de opções de software, inclusive na modalidade *freeware*, para suporte a projetos de BI.

<sup>3</sup> Data Warehouse é definido como uma coleção de dados orientada ao assunto, integrada, não volátil e variante em tempo (INMON, 2005).

Para suportar o armazenamento dos dados, foi definido o SGBD PostgreSQL<sup>4</sup>. Este Sistema Gerenciador de Banco de Dados (SGBD) é responsável pelo armazenamento e gerenciamento das tabelas que compõem o modelo dimensional de dados, a área de *staging*<sup>5</sup> para o processo de ETL e as eventuais tabelas de agregação para melhoria de desempenho nas consultas através dos filtros disponíveis nos painéis.

As ferramentas da suíte Pentaho<sup>6</sup> possuem versões *freeware* (Community Edition) e versões comerciais, que contam com suporte técnico da fabricante e algumas funcionalidades adicionais. Após uma análise sobre os recursos oferecidos pela versão comercial comparados com os da versão *freeware*, verificou-se que a versão Community Edition atende plenamente às necessidades apresentadas pelo projeto.

Para o processo de ETL, a ferramenta utilizada foi o Pentaho Data Integration (PDI). Este *software* é um dos mais utilizados atualmente, contemplando uma interface gráfica para construção das rotinas e suporte para execução manual e automatizada. Oferece ainda vasta opção de componentes que auxiliam as atividades de transformação de dados, além do suporte a boa variedade de fontes de dados, como bancos de dados, planilhas, arquivos de textos formatados, tabulados, entre outras (MAJCHRZAK; JANSEN; KUCHEN, 2011).

Para apresentação dos dados e gráficos que compõem os painéis, a ferramenta escolhida foi o Pentaho BI Server, produto que utiliza a tecnologia Web com boa flexibilidade de customização (personalização) e escalabilidade para suportar infraestruturas para cenários de grandes movimentações de dados e acessos simultâneos.

Tanto o desenvolvimento dos painéis (*dashboards*) quanto a apresentação final é realizada através de um navegador de Internet (*browser*). Assim como no caso do PDI, também se optou por pela versão Community Edition para o Pentaho BI Server.

## TRATAMENTO DOS DADOS

Abordaremos aqui as principais fontes de dados utilizadas no projeto e como os dados são processados para alimentar os painéis. A Relação Anual de Informações Anuais (Rais) e o Cadastro Geral de Empregados e Desempregados (Caged) são as principais fontes de dados. Vale observar que outras fontes de dados foram utilizadas como apoio ao modelo utilizado, uma vez que são diretamente referenciadas pela Rais e Caged, como por exemplo o cadastro de Municípios do IBGE e o Cadastro Brasileiro de Ocupações (CBO) do próprio Ministério do Trabalho.

## RELAÇÃO ANUAL DE INFORMAÇÕES SOCIAIS (RAIS)

A fonte de dados mais relevante para produção dos painéis é a compilação da Relação Anual de Informações Sociais (Rais), que é enviada eletronicamente pelas empresas ao Ministério do Trabalho, conforme Decreto nº 76.900 de 23/12/1975, e tem por objetivo suprir as necessidades de controle de atividades trabalhistas no país, prover dados para elaboração de estatísticas do trabalho e disponibilizar informações do mercado de trabalho às entidades governamentais.

A Rais registra o movimento e informações referentes aos seguintes tipos de vínculos empregatícios: empregados sob o regime da Consolidação das Leis do Trabalho (CLT), servidores da administração pública direta ou indireta, federal, municipal e fundações supervisionadas, trabalhadores avulsos e temporários, aprendizes, servidores e trabalhadores licenciados e servidores públicos cedidos ou requisitados.

<sup>4</sup> Disponível em <https://www.postgresql.org/> - Acesso em 10/01/2019.

<sup>5</sup> Área de Staging (Área de Representação de Dados) realiza a ligação entre os dados brutos e o Data Warehouse (RAHM; DO, 2000).

<sup>6</sup> Plataforma Pentaho para BI – disponível em <https://community.hitachivantara.com/community/products-and-solutions/pentaho>. Acessado em 10/01/2019

Não estão relacionados na Rais os seguintes perfis: diretores sem vínculos empregatícios sem recolhimento do Fundo de Garantia por Tempo de Serviço (FGTS), autônomos e eventuais, ocupantes de cargos eletivos, estagiários, empregados domésticos e cooperados ou cooperativados. Vale ressaltar que os painéis desenvolvidos abrangem somente os trabalhadores sob o regime da CLT.

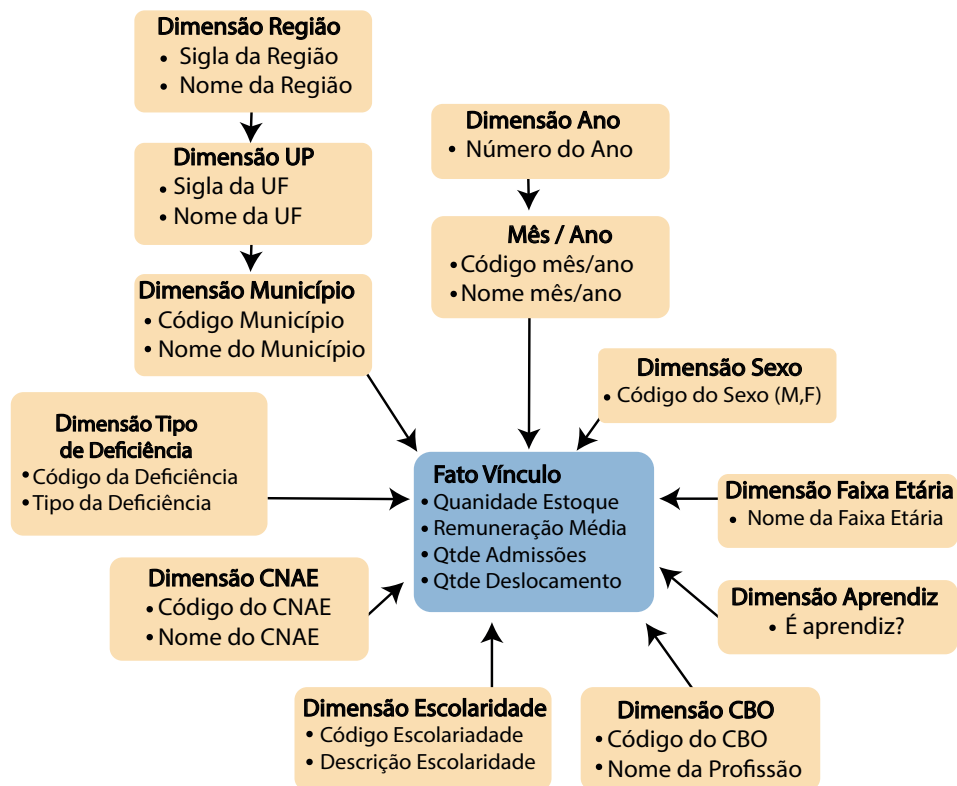
Dos dados informados na Rais, são listados a seguir os absorvidos neste projeto:

- Sexo do trabalhador (gênero);
- Escolaridade;
- Data de nascimento (para cálculo de idade e faixa de etária);
- Código do município, conforme a tabela do IBGE;
- Código Brasileiro de Ocupações (CBO), conforme a tabela do IBGE;

- Código Classificação Nacional de Atividades Econômicas (Cnae 2), conforme a tabela do IBGE;
- Tipo de deficiência (se for o caso);
- Indicação de aprendiz;
- Remuneração no último mês do ano de referência e
- Indicação se o vínculo estava ativo no último dia do ano de referência.

Para que os dados anteriores sejam vistos de forma sintetizada, e ainda permitir a implantação de filtros nos painéis para selecionar valores correspondentes a faixas específicas de dados, foi elaborada a estrutura de dados seguindo a Modelagem dimensional mostrada na figura 1.

Figura 1 – Modelo dimensional base



Fonte: os autores

A figura 1 indica o modelo utilizado na solução proposta. Utiliza a arquitetura *snow flake*, onde o elemento em azul representa a tabela fato, enquanto os demais elementos em amarelo representam as dimensões.

As dimensões município, UF e região estão hierarquicamente estruturadas devido à dependência de uma em relação à outra, permitindo operações de *drill up* e *drill down*. Na prática, é possível filtrar os registros da tabela fato por município, UF ou região. Ao filtrar por UF, por exemplo, significa selecionar o conjunto de todos os municípios daquela UF, e por consequência, todos os vínculos trabalhistas na tabela Fato que estejam nessa localidade.

A estrutura traz a dimensão faixa etária, que representa o agrupamento dos registros pela idade do trabalhador. Esta abordagem exige uma classificação etária prévia sobre os registros para vinculação à faixa etária correta, a fim de agilizar as consultas e construção dos gráficos.

## CADASTRO GERAL DE EMPREGADOS E DESEMPREGADOS (CAGED)

O Caged foi criado pela Lei nº 4923/65 e instituiu o registro permanente de admissões e desligamentos de empregados sob o regime da CLT. Como o Caged traz muitos dados complementares à Rais e tem emissão mensal, seu processamento torna-se importante para que os painéis possam exibir dados mais atualizados, haja vista que as empresas enviam dados da Rais relativos ao exercício anterior ao ano atual. Após sua consolidação, a Rais acaba por ser disponibilizada apenas na segunda metade do ano seguinte ao exercício.

Considerando apenas a Rais, os painéis teriam um problema de desatualização (dados relativos ao cenário de mais de um ano atrás), dando aos painéis a impressão de obsolescência.

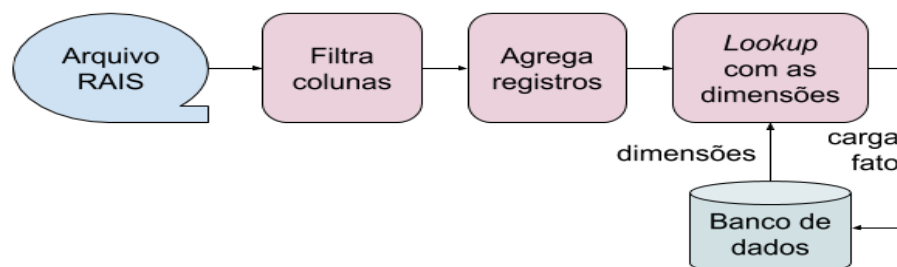
É possível combinar parcialmente os dados da Rais com os dados do Caged, permitindo que os dados dos painéis se mostrem mais atualizados. Tal combinação é possível devido à similaridade entre várias colunas dos leiautes da Rais e do Caged.

Logo, os painéis podem apresentar posição numérica de quantidade de vínculos mês a mês, e não ano a ano, o que ocorreria no caso de contar-se apenas com a Rais. A quantidade de vínculos empregatícios contabilizada na Rais é a soma de registros do arquivo da Rais onde há a indicação de que o vínculo estava ativo no final do ano de referência da Rais. Esta posição é considerada como sendo a do mês de dezembro do ano de referência. Todos os demais meses seriam dados por composição com os dados mensais do Caged. Isto justifica as dimensões mês/ano e ano no modelo dimensional da figura 1.

## EXTRAÇÃO, TRANSFORMAÇÃO E CARGA DA RAIS

Esta etapa segue a técnica de ETL abordada anteriormente e tem como objetivo extrair os dados dos arquivos de Rais e Caged, realizar as transformações necessárias para então carregá-los na estrutura correspondente ao modelo dimensional proposto na figura 1. O processamento da Rais é representado na figura 2.

Figura 2 – Processo de ETL do arquivo Rais



Fonte: os autores

O resultado do processo de ETL da Rais é a carga dos registros na tabela Fato Vínculo mostrada na figura 1. A figura 2 expõe as etapas realizadas neste processo. A primeira etapa lê o arquivo Rais filtrando apenas as colunas que são relevantes para o processo. A segunda etapa sumariza os registros na sua menor granularidade. Normalmente essa agregação é dispensada em projetos de BI para que as operações de *drill down* possam chegar até o menor nível de granularidade, que seria neste caso um vínculo empregatício específico. Uma vez que os painéis permitem, de certo modo, *drill down* até o nível de município e CBO, a agregação prévia permite melhor desempenho nas consultas ao banco de dados. A terceira etapa verifica os valores fornecidos no arquivo com os correspondentes nas tabelas dimensionais fazendo a ligação pelas chaves artificiais (*surrogate keys*).

Duas métricas calculadas a partir da agregação dos registros são relevantes para construção dos painéis do trabalhador:

- **Estoque de trabalhadores:** dado pela contagem de vínculos de trabalhadores ativos em 31 de dezembro do ano de referência.
- **Remuneração média:** a média aritmética baseada na remuneração do mês de dezembro do ano de referência da Rais.

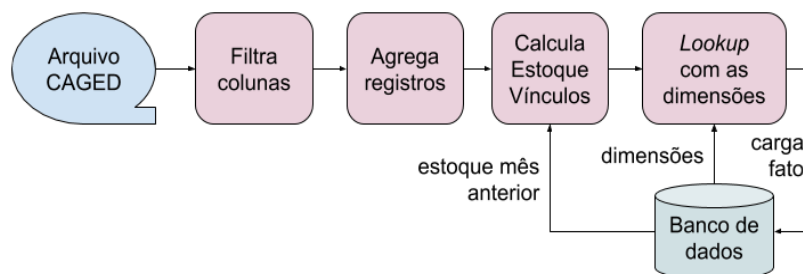
## EXTRAÇÃO, TRANSFORMAÇÃO E CARGA DO CAGED

O processamento do Caged é aplicado mensalmente conforme liberação do arquivo previamente tratado pelo Observatório do Trabalho, órgão do Ministério do Trabalho. Como a Rais consolidada é disponibilizada somente no segundo semestre do ano posterior, o processamento do Caged complementa a métrica de estoque de trabalhadores que compõe os painéis. Esse cálculo é realizado pela adição do número das admissões registradas no arquivo de Caged no estoque do mês anterior, e pela subtração do número de desligamentos.

A quantidade de admissões e desligamentos mensais fornecidas pelo Caged são também registradas com métricas na tabela Fato Vínculo mostrada na figura 1. Essa operação permite compor alguns dos painéis, como veremos adiante. Outra métrica utilizada é o saldo, que corresponde ao número de admissões subtraído pelo número de desligamentos.

Por definição, a métrica relativa à remuneração média não é afetada pelo processamento do Caged em função das diferentes metodologias de apuração entre este último e a Rais. Para efeito de apresentação, o valor de remuneração média mostrado nos painéis corresponde ao processamento da Rais imediatamente anterior.

Figura 3 – Processo de ETL do arquivo Caged



Fonte: os autores



Para cálculo de admissões e demissões, é utilizada uma das colunas da base de dados denominada “saldo\_mov”. A coluna indica, para cada registro de empregado e seu vínculo com uma empresa, se aquele item trata de uma admissão (valor 1) ou demissão (valor -1).

Essa contabilização, além de compor os dados para demonstrar a evolução da empregabilidade mês a mês, chamada “saldo”, também é utilizada no cálculo do “estoque” de trabalhadores (trabalhadores com vínculos ativos) no mês. O “estoque” é computado desde a última Rais processada, que traz o estoque na posição do último mês de referência somando-se o “saldo” mês a mês.

Uma peculiaridade a respeito do Caged é que, em um arquivo de mês específico, pode trazer registros de admissões ou demissões de meses anteriores (ajustes). Assim, não basta processar o Caged de forma incremental, alterando valores somente do mês de referência. A cada arquivo Caged, admissões e demissões são consideradas retroativamente desde o último mês da última Rais processada.

Apesar de o arquivo do Caged trazer alguns campos relativos à remuneração, tais valores não seguem os mesmos métodos utilizado na Rais. Logo, evoluir a remuneração obtida na Rais com a aplicação do Caged não é viável. Portanto, a remuneração média apresentada nos painéis é a presente na Rais do ano em referência ou na última Rais disponível.

A remuneração média obtida na Rais é dada pela soma dos valores da remuneração nominal de dezembro (coluna vl\_rem\_dezembro\_nom) onde tal valor esteja na faixa entre 0,3 e 150 salários mínimos, dividido pela quantidade de vínculos identificados no mesmo filtro. Esse filtro é uma convenção para descartar valores inválidos que eventualmente são encontrados na Rais, oriundos de problemas de preenchimento dos formulários enviados ao Ministério do Trabalho pelas empresas.

## **ELABORAÇÃO DOS PAINÉIS**

Considerando a estrutura de armazenamento de dados agregados, conforme modelo de dados exposto na figura 1, os painéis foram dispostos em estruturas específicas para análises direcionadas.

## **PERFIL DO TRABALHADOR**

O painel Perfil do Trabalhador apresenta informações para análises considerando as posições da CBO em relação aos cenários de emprego, desemprego. Esse indicador possibilita realizar um mapeamento do perfil dos trabalhadores, considerando dados como gênero, remuneração média, quantidade de vínculos ativos, entre outros, podendo ser filtrados por ano, CBO, região geográfica, UF ou município. Ademais, disponibiliza informações com o intuito de viabilizar a realização de análises sobre as necessidades de investimentos direcionados à qualificação profissional dos cidadãos, conforme mostra a figura 4.

Um exemplo de análise é a identificação de dados sobre a construção civil no Brasil, considerando os dados entre 2016 e 2018, retirados do painel Perfil do Trabalhador, onde é possível identificar a redução de vínculos ativos no país em 6,4 % nesse período, e que a quantidade de mulheres alocadas nesse nicho de mercado também reduziu cerca de 5,5 %. Essa análise considera o cenário nacional, mas também pode ser realizada considerando somente uma região, uma UF ou um município específico. Ainda no contexto da construção civil, é possível identificar que a remuneração média para essa atividade possui forte variação entre as regiões: por exemplo, a Região Sul tem remuneração de aproximadamente 14% superior à remuneração da mesma ocupação na Região Nordeste.

Figura 4 – Painel Perfil do Trabalhador / Perfil



Fonte: <http://indicadores.escoladotrabalhador.gov.br/> - Acesso em: Jan. 2019.

Os painéis iniciam com uma série de filtros que influenciam diretamente nos números mostrados nos gráficos. Os filtros funcionam como um *drill down* sobre os dados gerais. Pode-se também combinar vários filtros, dando possibilidade de elaboração de vasta quantidade de cenários de pesquisa.

Em seguida são indicados alguns números de maior relevância, como o estoque total de trabalhadores, a distribuição por gênero e a remuneração média.

São também mostrados alguns gráficos com a distribuição de estoque de trabalhadores e remuneração média por grau de instrução, faixa etária e tipo de deficiência.

No próprio painel do trabalhador há uma aba intermediária que dá acesso aos dados de “caracterização do trabalhador” e “admissões/desligamentos”, conforme as figuras 5 e 6 respectivamente.

Figura 5 – Aba caracterização do trabalhador

**Distribuição do emprego formal por Ocupação - CBO**

Código	Classificação Brasileira de Ocupações - CBO	Vínculos ativos	Homens	Mulheres	Remuneração Média
411010	Assistente administrativo	2.159.885	821.173	1.338.712	3.057
411005	Auxiliar de escritório	2.037.437	749.747	1.287.690	1.912
521110	Vendedor de comércio varejista	1.907.141	790.451	1.116.690	1.828
514320	Faxineiro	1.479.425	419.865	1.059.560	1.270
784205	Alimentador de linha de produção	1.019.408	699.170	320.238	1.699
782510	Motoneira de caminhão (notas regionais e internacionais)	883.709	875.906	7.803	2.409
421125	Operador de caixa	834.697	122.050	712.647	1.440
331205	Professor de nível médio no ensino fundamental	716.108	131.993	584.115	3.436
322205	Técnico de enfermagem	616.781	91.147	525.634	2.319
517330	Vigilante	595.411	538.270	57.141	2.267

**Distribuição Por Setor - Classificação Nacional de Atividades Econômicas (CNAE)**

Classe/Subclasse	Vínculos ativos	Homens	Mulheres	Remuneração Média
Administração Pública em Geral	8.051.023	2.968.615	5.082.408	3.717
Comércio Varejista de Mercadorias em Geral, com Predominância de Produtos Alimentícios - Hiper	1.283.190	648.351	634.839	1.846
Restaurantes e Outros Estabelecimentos de Serviços de Alimentação e Bebidas	1.271.917	598.659	673.258	1.513
Atividades de Atendimento Hospitalar	1.221.472	312.210	909.262	3.990
Transporte Rodoviário de Carga	907.993	794.182	113.811	2.247
Construção de Edifícios	700.781	628.992	71.789	2.068
Comércio Varejista de Artigos do Vestuário e Acessórios	620.506	131.899	488.607	1.802
Limpeza em Prédios e em Domicílios	616.698	252.796	363.902	1.352
Comércio Varejista de Ferragens, Madeiras e Materiais de Construção	583.658	448.390	145.268	1.725
Condomínios Prediais	572.987	457.249	115.738	2.049

**Setor de atividade**

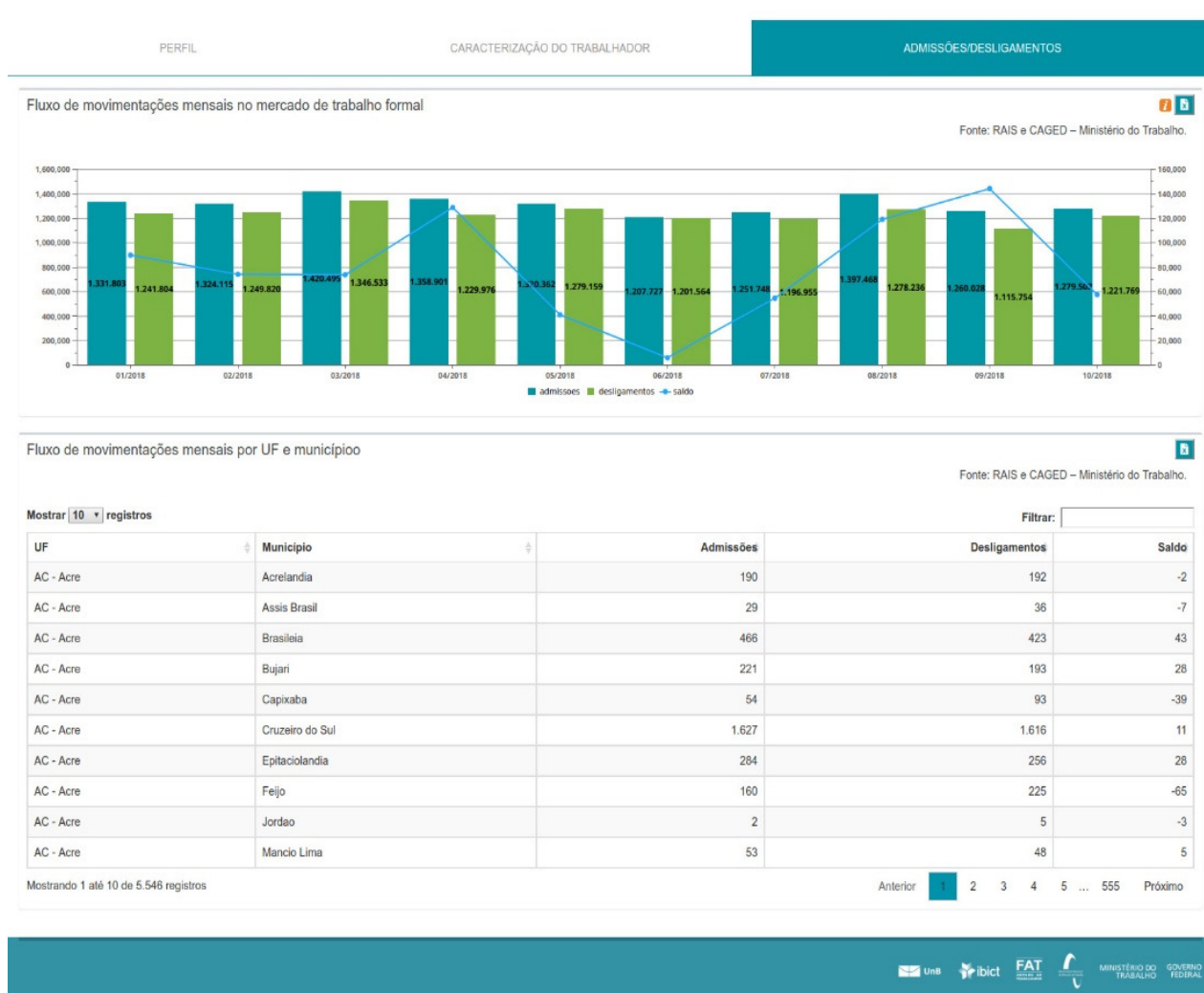
Setor IBGE	Vínculos ativos	Homens	Mulheres	Remuneração Média
Serviços	17.243.826	8.818.707	8.425.119	2.937
Comércio	9.220.410	5.219.949	4.000.461	2.004
Administração pública direta e autárquica	9.209.345	3.817.632	5.391.713	4.225
Indústria de transformação	7.250.543	4.991.845	2.258.698	2.887
Construção civil	1.921.055	1.733.139	187.916	2.347
Agricultura, silvicultura, criação de animais, extrativismo vegetal	1.575.559	1.322.036	253.523	1.826
Serviços industriais de utilidade pública	435.865	354.473	81.392	4.954
Extrativa mineral	215.586	190.335	25.251	6.216

**Unidade da Federação**

UF	Vínculos ativos	Homens	Mulheres	Remuneração Média
SP - São Paulo	13.364.536	7.438.995	5.927.630	3.286
MG - Minas Gerais	4.829.132	2.731.231	2.097.901	2.533
RJ - Rio de Janeiro	4.050.275	2.308.982	1.741.293	3.406
PR - Paraná	3.089.758	1.698.753	1.391.005	2.907
RS - Rio Grande do Sul	2.933.873	1.574.754	1.359.119	2.965
BA - Bahia	2.261.926	1.275.431	986.495	2.542
SC - Santa Catarina	2.260.592	1.237.297	1.023.295	2.836
PE - Pernambuco	1.599.077	923.691	675.386	2.452
GO - Goiás	1.558.754	875.852	683.902	2.816
CE - Ceará	1.491.160	825.032	666.128	2.248

Fonte: <http://indicadores.escoladotrabalhador.gov.br/> - Acesso em: Jan. 2019.

Figura 6 – Aba admissões/desligamentos



Fonte: <http://indicadores.escoladotrabalhador.gov.br/> - Acesso em: Jan. 2019.

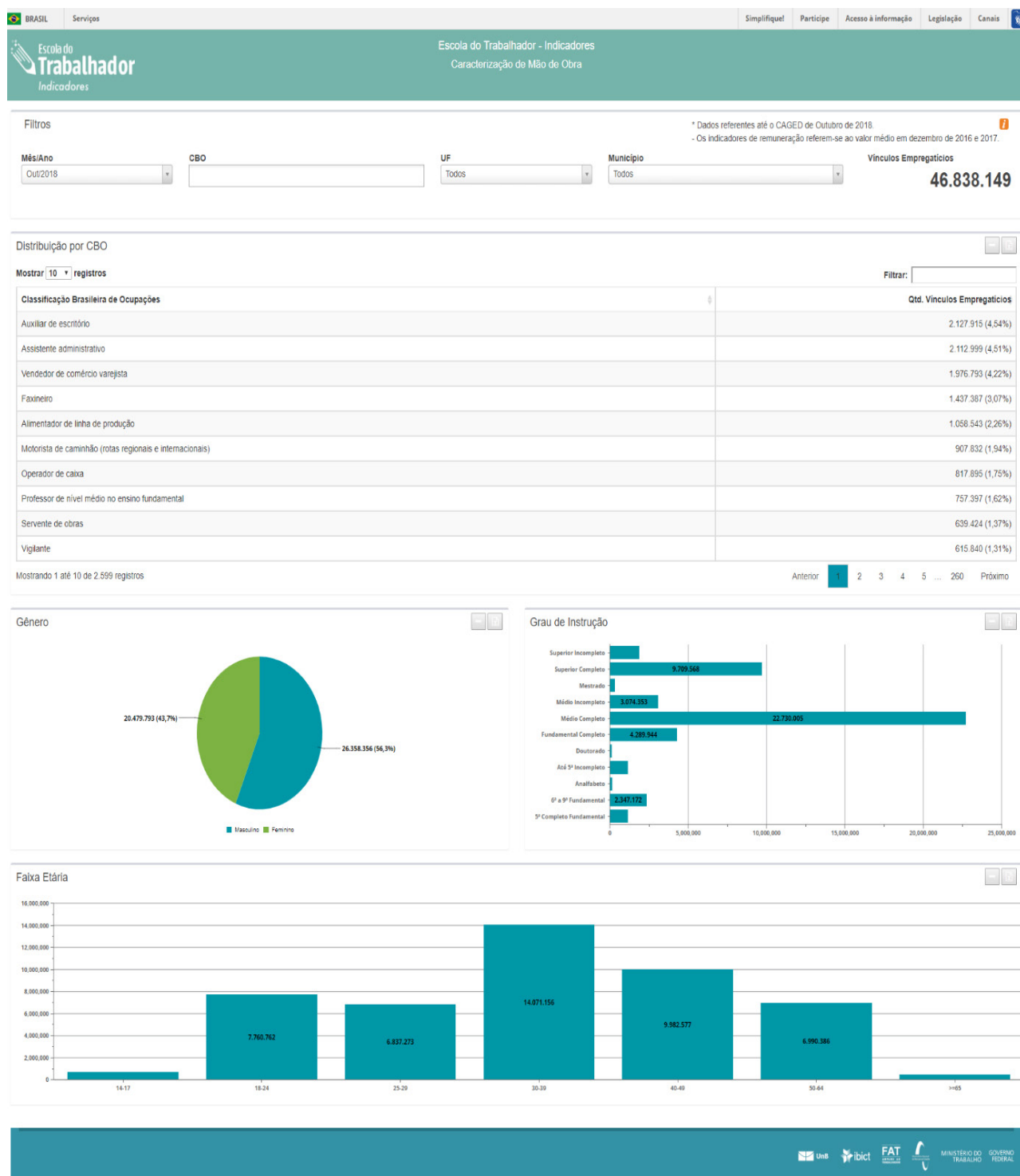
A aba de “caracterização do trabalhador”, conforme a figura 5, lista as maiores quantidades de vínculos por CBO, Cnae, setor de atividade e UF. Enquanto a aba “admissões/desligamentos”, conforme a figura 6, traz um demonstrativo de admissões e desligamentos ocorridos mês a mês, disposto no formato de gráfico de barras.

Com as informações sobre a quantidade de admissões e desligamentos indicados na figura 6, é possível identificar se o saldo entre admitidos e desligados está positivo ou negativo, ou seja, como está o comportamento do mercado de trabalho em determinada região, UF ou até mesmo município,

através de uma visão genérica por localização ou também considerando mais uma visão sobre determinada ocupação neste cenário.

Seguindo a mesma linha do painel Perfil do Trabalhador, foram desenvolvidos outros dois painéis com recortes de públicos prioritários. O painel de Aprendizagem Profissional, somente com os trabalhadores classificados como aprendizes, e o Perfil do Trabalhador Mulher, que abrange somente o gênero feminino no mercado de trabalho. Trata-se de uma oportunidade de melhor avaliação do cenário atual de trabalho para a mulher e para os aprendizes, bem como o acompanhamento da evolução de ocupação nos postos trabalhos.

Figura 7 – Painel caracterização da mão de obra



Fonte: <http://indicadores.escoladotrabalhador.gov.br/> - Acesso em: Jan. 2019.

## CARACTERIZAÇÃO DA MÃO DE OBRA

O painel de caracterização de mão de obra disponibiliza informações específicas sobre os trabalhadores e suas ocupações. Considera a CBO como referência para exibição de informações atuais sobre o total de vínculos empregatícios, de acordo com suas características.

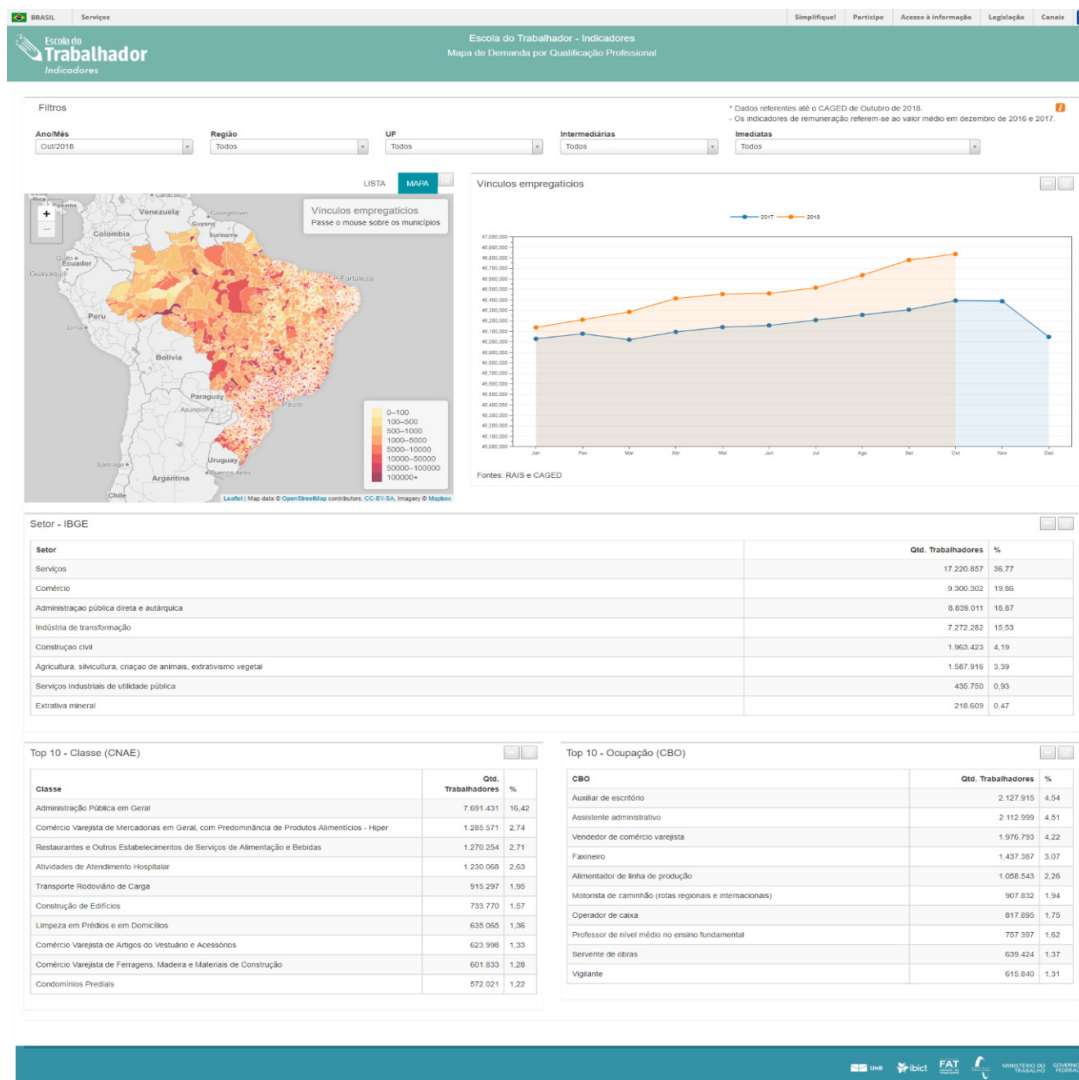
O painel de caracterização de mão de obra disponibiliza informações mensais sobre a quantidade de trabalhadores, que podem ser caracterizados pela Classificação Brasileira de Ocupações (CBO) e a distribuição por gênero, grau de instrução e escolaridade.

Os dados podem ser filtrados por áreas específicas da CBO, por Unidade Federativa (UF) e por município. O painel também apresenta um indicador de quantidade de vínculos empregatícios, que é calculado a partir do cruzamento dos dados das bases da Rais e do Caged.

## MAPA DE DEMANDA POR QUALIFICAÇÃO PROFISSIONAL

Este painel possibilita uma visualização sobre a concentração de trabalhadores de forma georreferenciada, através de um mapa de calor e a visualização em relação às atividades das empresas e dos trabalhadores.

Figura 8 – Painel mapa de demanda por qualificação profissional



Fonte: <http://indicadores.escoladotrabalhador.gov.br/> - Acesso em: Jan. 2019.

O indicador mapa de demanda por qualificação profissional faz o cruzamento das bases de dados do Rais e do Caged para gerar um extrato mensal sobre a empregabilidade no cenário nacional. Esse indicador possibilita a análise do histórico do total de trabalhadores por mês. Nas análises, são considerados elementos como vínculos, Setor do IBGE, Cnae e CBO.

## CONCLUSÕES

Os indicadores disponibilizados por meio dos painéis possibilitam o mapeamento e entendimento do mercado de trabalho, chegando em nível municipal e com possibilidades de auxiliar na formação de políticas públicas voltadas para a qualificação profissional com o uso de tecnologias. Com as informações da Rais e Caged agregadas, é possível visualizar as peculiaridades em relação ao mercado de trabalho brasileiro considerando as atividades dos trabalhadores (CBO) e também as atividades das empresas (Cnae), com diversos tipos de visualizações e possibilidades de construção de cenários de análise para casos específicos. Os painéis também possibilitam o acompanhamento de profissões e setores da economia que sejam alvo de políticas públicas direcionadas para a qualificação profissional. Com os dados disponibilizados, é possível verificar se determinada ação obteve efetividade na criação de novos cargos ou na melhoria da remuneração.

---

## REFERÊNCIAS

- ARAÚJO, L. R.; SOUZA, J. F.de. Aumentando a transparência do governo por meio da transformação de dados governamentais abertos em dados ligados. *Revista Eletrônica de Sistemas de Informação*, v. 10, n. 1, 2011.
- ARAÚJO JÚNIOR, R. H. de; SOUSA, R. T. B. de. Estudo do ecossistema de Big Data para conciliação das demandas de acesso, por meio da representação e organização da informação. *Ciência da Informação*, v. 45, n. 3, 2016.

FREY, C. B.; OSBORNE, M. A. The future of employment: how susceptible are jobs to computerisation? *Technological forecasting and social change*, v. 114, p. 254-280, 2017.

INMON, W. H. *Building the data warehouse*. New Jersey: John wiley & sons, 2005.

KIMBALL, R. *The Data Warehouse Toolkit*. [S.l.]: Editora Campus, 2002.

MAJCHRZAK, T. A.; JANSEN, T.; KUCHEN, H. Efficiency evaluation of open source ETL tools. *In: ACM SYMPOSIUM ON APPLIED COMPUTING*, 26., 2011, Taichung, Taiwan. *Proceedings* [...]. Taichung, Taiwan: Tunghai University, 2011. p. 287-294.

MINISTÉRIO DO TRABALHO. *Relação Anual de Informações Sociais - RAIS*. Brasília: [s.n.], [2018?]. Disponível em: <http://www.rais.gov.br>. Acesso em: 29 dez. 2018.

MINISTÉRIO DO TRABALHO. *Cadastro Geral de Empregados e Desempregados (CAGED)*. LEI Nº 4923/65. Brasília: TEM, 2018. Disponível em: <https://caged.maisemprego.mte.gov.br/portalcaged/>. Acesso em: 29 dez. 2018.

RAHM, E.; DO, H. H. Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull.*, v. 23, n. 4, p. 3-13, 2000.

TURBAN, E. *et al. Business Intelligence: um enfoque gerencial para a inteligência do negócio*. Bookman Editora, 2009.

VASSILIADIS, P.; SIMITSIS, A.; SKIADOPOULOS, S. Conceptual modeling for ETL processes. *In: ACM INTERNATIONAL WORKSHOP ON DATA WAREHOUSING AND OLAP*, 5., 2002, United States. *Proceedings* [...]. United States, 2002.

---

## AGRADECIMENTOS

Os valores apresentados nos painéis foram aferidos e validados pelo Observatório Nacional do Trabalho, na pessoa da Sra. Mariana Almeida. O Observatório também contribuiu com várias sugestões essenciais para o aprimoramento dos painéis e construção dos indicadores. O presente trabalho contou com suporte do Termo de Execução Descentralizada TED 001/2017 MTb/SPPE - FUB.