

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D1W31F>

Title: Impacto social da UFSCar: análise a partir dos egressos

Creator: Marisa Cubas Lozano - **ORCID:** [0000-0002-0888-9145](https://orcid.org/0000-0002-0888-9145)

Affiliation: Universidade Federal de São Carlos (ufscar.br)

Principal Investigator: Marisa Cubas Lozano

Data Manager: Marisa Cubas Lozano

Project Administrator: Marisa Cubas Lozano

Funder: Universidade Federal de São Carlos (ufscar.br)

Template: Digital Curation Centre

Project abstract:

A fim de entender como a UFSCar impacta na sociedade foram levantados dados públicos de egressos no Currículo Lattes. Os dados do Currículo Lattes foram recuperados de forma automatizada a partir da ferramenta LattesMachina e tratados com o software proprietário VantagePoint e script da Linguagem de Programação Python.

Start date: 12-01-2020

End date: 07-30-2023

Last modified: 01-18-2024

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Impacto social da UFSCar: análise a partir dos egressos

Foram coletados dados dos Currículos Lattes dos egressos da Universidade Federal de São Carlos.

Para o tratamento dos dados foram usados o software pago Vantage Point e um script em linguagem de programação Python (<https://github.com/Leluke/python-xml-parser/tree/master>).

Para análise dos dados foram solicitados dados sobre os cursos de graduação da UFSCar e o seu número de egressos.

Foram realizadas extrações de forma automatizada dos Currículos Lattes de egressos da UFSCar a partir da ferramenta LattesMachina* em 08 abr. 2021 e em 05 dez. 2022. Foram recuperados dados como formação acadêmica, atividades profissionais e acadêmicas (como participação em cursos e eventos), e produção intelectual, artística e técnica, em formato .xml.

Para tratamento dos dados recuperados em arquivos .xml, foi usado um script em linguagem de programação Python (<https://github.com/Leluke/python-xml-parser/tree/master>), que gerou três planilhas no formato .csv.

Para análise dos dados, foram solicitados via rede de serviços da UFSCar (<https://servicos.ufscar.br/>) o número de egressos por curso de graduação da UFSCar desde a sua criação e uma relação com todos os cursos de graduação que a UFSCar já teve.

*

MATIAS, Mesailde Souza de Oliveira. **Base referencial para o povoamento de repositórios institucionais: coleta automatizada de metadados da Plataforma Lattes**. 2015. Dissertação (Mestrado em Gestão de Organizações e Sistemas Públicos) - Universidade Federal de São Carlos, São Carlos, SP, 2015. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/6932>. Acesso em: 16 ago. 2021.

MATIAS, Mesailde Souza De Oliveira; AMARAL, Roniberto Morato Do; MATIAS, Paulo. Proxy customizado para acesso ao web service da Plataforma Lattes. *In*: WORKSHOP DE TECNOLOGIA DA INFORMAÇÃO E COMUNICAÇÃO DAS INSTITUIÇÕES FEDERAIS DE ENSINO SUPERIOR, 11., 2017, Salvador. **Anais [...]**. Salvador: UFBA, 2017. Disponível em: <http://www.xiwticifes.ufba.br/modulos/submissao/Upload-353/86143.pdf>. Acesso em: 17 ago. 2021.

Os dados já são disponibilizados com os metadados.

Para leitura dos dados pelo software Vantage Point é necessário uso de um filtro no formato .config que será depositado junto com os dados (arquivo 20220511config-importacao-xml-synclattes.config).

Os dados utilizados estão disponibilizados de forma pública na Plataforma Lattes.

Os dados estão disponíveis de forma pública na Plataforma Lattes (<https://lattes.cnpq.br/>) e aqueles que foram coletados para essa pesquisa serão disponibilizados em repositório digital sob uma licença Creative Commons.

Os dados ficarão armazenados no computador pessoal da pesquisadora, um servidor de uso exclusivo do Grupo de Pesquisa NIT/Materiais UFSCar e no Repositório de Dados Zenodo (<https://doi.org/10.48321/D1W31F>).

Como são dados públicos, estão disponíveis em um repositório público.

Os dados coletados e aqueles gerados após o tratamento serão preservados a longo prazo:

- arquivos .xml dos Currículos Lattes dos egressos (xml-egressos-bci.rar),
- planilhas .csv gerada a partir do script do Python,
- filtro para leitura dos arquivos .xml no *software* Vantage Point (20220511config-importacao-xml-synclattes.config)

Os dados coletados e os tratados serão disponibilizados em repositórios digitais de acesso público.

Os dados serão disponibilizados em repositórios digitais de acesso público e divulgados em publicações científicas, como artigos e trabalhos apresentados em eventos.

Os dados não têm restrição de acesso.

Durante a pesquisa, a pesquisadora e o orientador da pesquisa serão os únicos responsáveis pelo gerenciamento dos dados.

Foi necessário acesso ao LattesMachina, limitado a pesquisadores vinculados ao NIT/Materiais UFSCar devido o convênio realizado com o CNPq, para obtenção dos dados.

O acesso ao servidor para armazenamento, onde o software Vantage Point está instalado, foi possibilitado pela participação da pesquisadora no Grupo de Pesquisa NIT/Materiais UFSCar.

O script usado para tratamento dos dados recuperados dos Currículos Lattes foi desenvolvido pelo analista de sistema Luccas Mendes Rodrigues (<https://github.com/Leluke>).

Planned Research Outputs

Text - "Dados dos egressos BCI - Lattes"

São 534 arquivos .xml referente aos Currículos Lattes dos egressos do curso de Biblioteconomia e Ciência da Informação da UFSCar

Text - "Filtro para leitura do dados "

Filtro criado para facilitar a interpretação dos .xml obtidos dos Currículos Lattes pelo software Vantage Point.

Text - "dados sobre atuação profissional - BCI"

Documento .csv gerado a partir do script em Linguagem Python com dados sobre a atuação profissional e pós-graduação extraídos dos arquivos .xml extraídos dos Currículos Lattes dos egressos da UFSCar.

Text - "dados sobre atuação profissional - todos"

Arquivo .csv contendo informações sobre atuação profissional dos egressos da UFSCar obtido a partir dos arquivos .xml de seus Currículos Lattes.

Text - "dados sobre pós-graduação - todos"

Arquivo .csv contendo informações sobre a realização de pós-graduação dos egressos da UFSCar obtido a partir dos arquivos .xml de seus Currículos Lattes.

Text - "Egressos da UFSCar/curso"

Arquivo .xlsx com o número de egressos por curso de graduação da UFSCar.

Text - "Cursos de graduação da UFSCar"

Relação em .xlsx de todos os cursos da UFSCar desde a sua fundação.

Text - "Dados dos egressos UFSCar - Lattes"

Os Currículos Lattes de 18.701 egressos em arquivos .xml.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Dados dos egressos BCI - Lattes	Text	2022-05-20	Open	Zenodo	25 MB	Creative Commons Zero v1.0 Universal	None specified	No	Yes
Filtro para leitura do dados	Text	2022-05-20	Open	Zenodo	1 MB	Creative Commons Zero v1.0 Universal	None specified	No	No
dados sobre atuação profissional - BCI	Text	2022-05-20	Open	Zenodo	1 MB	Creative Commons Zero v1.0 Universal	None specified	No	Yes
dados sobre atuação profissional - todos	Text	2023-06-22	Open	Zenodo	1 MB	Creative Commons Zero v1.0 Universal	None specified	No	Yes
dados sobre pós-graduação - todos	Text	2023-06-22	Open	Zenodo	5 MB	Creative Commons Zero v1.0 Universal	None specified	No	Yes
Egressos da UFSCar/curso	Text	2023-06-22	Open	Zenodo	1 MB	Creative Commons Zero v1.0 Universal	None specified	No	No
Cursos de graduação da UFSCar	Text	2023-06-27	Open	Zenodo	1 MB	Creative Commons Zero v1.0 Universal	None specified	No	No
Dados dos egressos UFSCar - Lattes	Text	2023-08-21	Open	Zenodo	250 MB	Creative Commons Zero v1.0 Universal	None specified	No	Yes