

## TaDiRAH (Argentina - Brasil): uma iniciativa das humanidades digitais no Sul Global na pavimentação de uma ciência aberta

*Elizabeth Maria Freire de Jesus<sup>1</sup>, Márcia Cavalcanti<sup>2</sup>, Tainá Regly<sup>3</sup>, Ricardo M. Pimenta<sup>4</sup>, Renan M. Castro<sup>5</sup>*

<sup>1</sup> UFRJ, Rio de Janeiro, Brasil *beth@nce.ufrj.br*,

<sup>2</sup> USU, Rio de Janeiro, Brasil *marciacavalcanti@gmail.com*

<sup>3</sup> IBICT/UFRJ, Rio de Janeiro, Brasil *taina.regly@gmail.com*

<sup>4</sup> IBICT, Rio de Janeiro, Brasil *ricardopimenta@ibict.br*

<sup>5</sup> FGV, Rio de Janeiro, Brasil *renan.castro@fgv.br*

### Introdução

A Ciência Aberta engloba múltiplas dimensões de abertura, tal como observado na Taxonomia da Ciência Aberta apresentada pelo Facilitate Open Science Training for European Research[1], onde é possível perceber o quão amplas são as possibilidades de estratégias e ações em diferentes escopos e níveis que confluem para uma efetiva democratização do conhecimento que é coletiva e socialmente produzido.

Este trabalho apresenta uma primeira ação da parceria estabelecida do Centro Argentino de Información Científica y Tecnológica (CAICYT), por meio do seu Laboratório de Humanidades Digitales del Centro Argentino de Información Científica y Tecnológica del CONICET (HD CAICYT LAB) com o Laboratório em Rede de Humanidades Digitais (Larhud) da Coordenadoria de Ensino e Pesquisa, Ciência e Tecnologia da Informação (COEPE) do Instituto de Informação em Ciência e Tecnologia (IBICT), para o desenvolvimento da versão em português de uma taxonomia para pesquisas e métodos digitais aplicados à grande área das Humanidades: Taxonomia sobre Atividades Digitais de Pesquisa nas Humanidades-TaDiRAH[2].

### Proposta

A insuficiência de recursos materiais, com destaque aos softwares e hardwares, assim como aos financiamentos, já nos colocam em condição de reprodutores das formas, técnicas e estéticas provenientes do Norte global. O cenário é demasiado delicado e requer atenção. Requer, sobretudo, ação. Principalmente quando o espaço próprio para a realização destas pesquisas é atravessado pelos softwares e hardwares controlados pelas grandes corporações Big Tech que irão interferir diretamente nas formas de produção e divulgação do conhecimento construído. Em meio a este cenário, iniciativas que busquem elaborar acesso à informação e auxílio à organização do conhecimento por meio de recursos/ferramentas abertas e, sobretudo, idealizadas e criadas no Sul Global, são verdadeiras “trincheiras” contra a reificação debordiana das estruturas e instrumentos direcionados à educação, à ciência e tecnologia. O objetivo da versão em português da TaDiRAH foi de tornar familiar um vocabulário pertinente às atividades em Humanidades Digitais (HD) e possibilitar que as futuras pesquisas e produtos construídos no âmbito nacional possam ser melhor recuperados e preservados a partir de uma taxonomia comum e multilíngue.

## Metodologia

A tradução da taxonomia TaDiRAH representou minimamente um duplo desafio: o primeiro refere-se à tradução interlíngua, do espanhol para o português, sendo que em diversas situações houve a necessidade de se recorrer também à versão em inglês, a fim de verificarmos e/ou confirmarmos a equivalência de sentido atribuído ao termo em questão. Um segundo desafio diz respeito à comunicabilidade da taxonomia, considerando-se o domínio deste instrumento que abarca métodos, objetos e técnicas computacionais em sua relação com o público-alvo ou usuários a que se destina, que são pesquisadores das HD e que, potencialmente, poderão se beneficiar do suporte conceitual abarcado pelo âmbito da taxonomia. Para tal, sabemos que a capacidade de comunicação e divulgação desse instrumento entre pares, na comunidade científica em geral, é fundamental.

No processo de tradução da TaDiRAH foram tomadas precauções para evitar problemas semânticos, porém foi mantida a relação semântica entre os termos presentes na língua original. Ainda assim, o procedimento foi realizado com prudência para que a taxonomia refletisse a realidade das HD no Brasil. Dos 129 termos que deveriam ser trabalhados, Currícula, Personas, POS-tagging e Recolección ficaram sem tradução por não possuírem termos correspondentes no português e por seu uso não fazer sentido para os usuários brasileiros em sua língua materna. Para outros termos como Búsqueda e Investigación tomou-se a decisão de traduzir Búsqueda como “Busca” e Investigación como “Pesquisa”.

Ao levar em consideração o contexto dos usuários, o vocabulário controlado multilíngue toma como prioridade a busca pelo termo mais adequado para representar um dado conceito levando em consideração a especificidade e estrutura da língua escolhida. Caso o vocabulário apresente inconsistências relacionadas à sua hierarquia, às relações entre os termos ou até mesmo a presença indevida de um termo, a disseminação da informação de maneira igualitária será prejudicada e sua confiabilidade poderá ser posta em questão[3].

No processo de tradução não foi possível acrescentar termos considerados relevantes para a realidade das HD no Brasil, estabelecer novas relações entre os termos traduzidos e inserir definições aos termos que não possuíam esse tipo de conteúdo. Porém, existe uma vontade comum presente nas instituições responsáveis pelas diversas versões da TaDiRAH de promover uma atualização da taxonomia. Nesse cenário, representantes brasileiros seriam eleitos para colaborar e construir de maneira simultânea um vocabulário multilíngue que levasse em consideração a sua realidade e seguisse os preceitos ideais abordados por [4].

## Considerações finais

O desenvolvimento da taxonomia TaDiRAH e seu reconhecimento pela comunidade dedicada às Humanidades Digitais busca gerar possibilidades de interoperabilidade entre o vocabulário controlado e as demais plataformas e artefatos advindos de novas pesquisas de fundo humanístico debruçadas sobre recursos computacionais. Formar-se-á, em tal contexto de práticas e tecnologias, uma “cadeia metodológica” em prol de atividades investigativas. Estas ferramentas heurísticas são fundamentais para a construção e para a descoberta de novos conhecimentos. Obviamente, esta “boa” construção não está desprovida de pensamento crítico. É

ele, inclusive, que nos faz chegar até aqui. Comunicar a construção e reprodução de novas estruturas para o emprego nas Humanidades está atrelada ao entendimento de sua contribuição para o desenvolvimento do emprego dos recursos digitais nas pesquisas da área, bem como para sua permanente recuperação da informação.

#### Referências

- [1] Facilitate Open Science Training for European Research [Online]. disponível: <https://www.fosteropenscience.eu/foster#taxonomy>
- [2] TaDiRAH - Taxonomía sobre Actividades de investigación digital en humanidades [Online]. disponível: <https://www.vocabularyserver.com/tadirah/es/index.php>
- [3] R. Biscalchin; W. Moreira, "Construção de vocabulários multilíngues: perspectivas culturais," in *Perspectivas em Ciência da Informação*, Belo Horizonte, MG, 2020, pp. 47-67
- [4] M. Hudon, "Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge and concepts." in *Knowledge Organization*, Würzburg, BY, pp. 84-91, 1997.

## APRESENTAÇÃO

Este trabalho apresenta uma primeira ação da parceria estabelecida entre o **Laboratório de Humanidades Digitales del Centro Argentino de Información Científica y Tecnológica del CONICET (HD CAYCIT LAB)** e o **Laboratório em Rede de Humanidades Digitais (Larhud) do Instituto de Informação em Ciência e Tecnologia (IBICT)**, para o desenvolvimento da versão em português de uma taxonomia para pesquisas e métodos digitais aplicados à grande área das Humanidades: Taxonomia sobre Atividades Digitais de Pesquisa nas Humanidades-TaDiRAH.

## PROBLEMA

A insuficiência de recursos materiais, com destaque aos softwares e hardwares, assim como aos financiamentos, já nos colocam em condição de reprodutores das formas, técnicas e estéticas provenientes do Norte. Ou seja, enquanto o espaço próprio para a realização destas pesquisas é atravessado pelos softwares e hardwares controlados pelas grandes corporações *Big Tech*, é mister que surjam iniciativas que busquem viabilizar o acesso à informação e também a organização do conhecimento por meio de recursos/ferramentas abertas e, sobretudo, criadas no Sul Global, atendendo às singularidades características de nossos contextos tecnopolíticos e sociotécnicos.

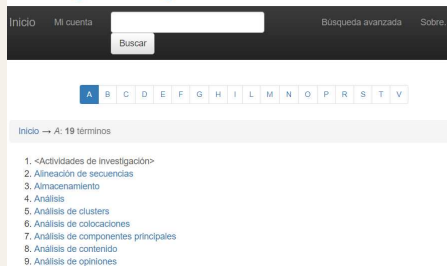
## OBJETIVO

O objetivo da versão em português da TaDiRAH foi de tornar familiar à língua portuguesa um vocabulário pertinente às atividades realizadas no âmbito das Humanidades Digitais (HD) e possibilitar que as futuras pesquisas e produtos construídos no cenário brasileiro possam ser melhor recuperados e preservados a partir de uma taxonomia comum e multilíngue. Outro objetivo é estimular o desenvolvimento de mais estruturas informacionais que auxiliem as HD no Brasil.

## OBJETO

Tradução de uma taxonomia do campo das HD da língua espanhola para o português.

TaDiRAH - Taxonomía sobre Actividades de investigación digital en humanidades



## METODOLOGIA: DESAFIOS

- O alongamento de uma língua para fazê-la se ajustar na estrutura conceitual da estrangeira a ponto de se tornar irreconhecível a seus próprios falantes;
- A transferência da estrutura conceitual de uma cultura para outra, sendo ela apropriada ou não;
- A tradução de termos da língua de origem para expressões sem sentido no idioma de destino.

## METODOLOGIA: TRADUÇÃO

Seguimos o método de *peer-review* onde os termos foram traduzidos inicialmente por um responsável, na sequência foram revisados, por outro membro que poderia editar e adicionar mais informações.

LARHUD - Laboratorio en Rede de Humanidades Digitales

Proyecto	Objetivo	Estado	Fecha	Responsable
1. Actividades de investigación	Traducción para el portugués de la taxonomía de actividades de investigación digital en humanidades (TaDiRAH)	En curso	2021	Elizabeth M. F. de Jesus
2. Almacenamiento de secuencias	Traducción para el portugués de la taxonomía de almacenamiento de secuencias	En curso	2021	Márcia Cavalcanti
3. Almacenamiento	Traducción para el portugués de la taxonomía de almacenamiento	En curso	2021	Tainá Regly
4. Análisis	Traducción para el portugués de la taxonomía de análisis	En curso	2021	Ricardo M. Pimenta
5. Análisis de clusters	Traducción para el portugués de la taxonomía de análisis de clusters	En curso	2021	Renan M. Castro
6. Análisis de colocaciones	Traducción para el portugués de la taxonomía de análisis de colocaciones	En curso	2021	Elizabeth M. F. de Jesus
7. Análisis de componentes principales	Traducción para el portugués de la taxonomía de análisis de componentes principales	En curso	2021	Márcia Cavalcanti
8. Análisis de contenido	Traducción para el portugués de la taxonomía de análisis de contenido	En curso	2021	Tainá Regly
9. Análisis de opiniones	Traducción para el portugués de la taxonomía de análisis de opiniones	En curso	2021	Ricardo M. Pimenta
10. Almacenamiento de secuencias	Traducción para el portugués de la taxonomía de almacenamiento de secuencias	En curso	2021	Renan M. Castro

## METODOLOGIA: PLATAFORMA

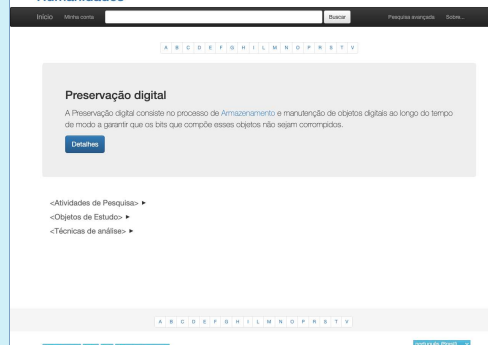
Após a tradução dos termos e sua devida revisão, os significados foram inseridos na Plataforma Tematres que hospeda a taxonomia traduzida.



## TADIRAH EM PORTUGUÊS

Dos 129 termos que deveriam ser trabalhados, *Currícula*, *Personas*, *POS-tagging* e *Recolección* ficaram sem tradução por não possuírem termos correspondentes no português e por seu uso não fazer sentido para os usuários brasileiros em sua língua materna.

TADIRAH - Taxonomia sobre Atividades Digitais de pesquisa nas Humanidades



Não foi possível acrescentar termos considerados relevantes para a realidade das HD no Brasil, estabelecer novas relações entre os termos traduzidos e inserir definições aos termos que não possuíam esse tipo de conteúdo.

## CONSIDERAÇÕES FINAIS

O desenvolvimento da taxonomia TaDiRAH e seu reconhecimento pela comunidade dedicada às HD busca gerar possibilidades de interoperabilidade entre o vocabulário controlado e as demais plataformas e artefatos advindos de novas pesquisas de fundo humanístico debruçadas sobre recursos computacionais. Comunicar a construção e reprodução de novas estruturas para o emprego nas Humanidades está atrelada ao entendimento de sua contribuição para o desenvolvimento do emprego dos recursos digitais nas pesquisas da área, bem como para permanente recuperação da informação.

## CONTATO

- Elizabeth M. Freire de Jesus [beth@nce.ufrj.br](mailto:beth@nce.ufrj.br)
- Márcia Cavalcanti [marciacavalcanti@gmail.com](mailto:marciacavalcanti@gmail.com)
- Tainá Regly [taina.regly@gmail.com](mailto:taina.regly@gmail.com)
- Ricardo M. Pimenta [ricardopimenta@ibict.br](mailto:ricardopimenta@ibict.br)
- Renan M. Castro [renan.castro@fgv.br](mailto:renan.castro@fgv.br)

